Realization Theory of Stochastic Jump-Markov Linear Systems

Mihály Petreczky
Eindhoven University of Technology, The Netherlands
M.Petreczky@tue.nl

René Vidal
Johns Hopkins University, Baltimore MD, USA
rvidal@cis.jhu.edu

***Abstract*— We present a stochastic realization theory for sto-chastic jump-Markov linear systems (JMLSs). We derive necessary and sufficient conditions for existence of a realization, along with a characterization of minimality in terms of reachability and observability. We also sketch a realization algorithm and argue that minimality can be checked algorithmically. The main tool for solving the stochastic realization problem for JMLSs is the formulation and solution of a stochastic realization problem for a general class of bilinear systems with nonwhite-noise inputs using the theory of formal power series.**

I. Introduction

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 minimal) state-space representation of the observed behavior.

For linear systems and deterministic bilinear systems, the realization problem is relatively well understood thanks to the works of Kalman, Brockett, Fliess, Isidori, Sontag and Sussmann in the sixties and seventies. However, arguably the only paper on realization of stochastic bilinear systems is [5], which requires the input to be white noise. There are a number of papers on identification of bilinear systems with inputs that are not white noise, see e.g., [3], [6], [20]. However, these papers require a number of conditions on the underlying system 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 (see [21] and references therein) and for general smooth systems [8], [19]. However, the algorithmic aspects of this theory are not well developed. There is a substantial amount of work on realization theory of polynomial systems [17], and rational systems [22] both in continuous and discrete time. However, the issue of minimality for polynomial systems is not well understood.

For deterministic hybrid systems, one of the first works on realization is [7], though a formal theory is not presented. Later work deals with switched linear systems [14], switched bilinear systems [10], linear/bilinear hybrid systems without guards and partially observed discrete states [9], [11], and nonlinear analytic hybrid systems without guards [15]. [12] presents necessary and sufficient conditions for existence of a realization of piecewise-affine autonomous hybrid systems with guards but it does not address minimality. To the best of our knowledge, the only paper on realization theory of stochastic hybrid systems is [16], where only necessary conditions for existence of a realization are presented.

In this paper we present a complete stochastic realization theory of discrete-time stochastic jump-Markov linear systems (JMLSs). JMLSs have a vast literature and numerous applications (see for example [4] and the references therein). For simplicity, we consider only JMLSs with fully observed discrete state. In addition, we 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 obtain a more general model, which we call *generalized stochastic jump-Markov linear systems*. It turns out that the class of classical JMLSs 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 necessary and sufficient conditions for existence of a realization as well as a neat characterization of minimality. We also formulate a realization algorithm and argue that minimality can be checked algorithmically.

The main tool for solving the realization problem for JMLSs is the formulation and solution to the following generalized bilinear realization problem. 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 will show that one can construct a bilinear state-space representation of the output process in the forward innovation form. The results on realization theory of JMLSs are then obtained by viewing the discrete state process as an input process. To the best of our knowledge, our solutions to both the generalized bilinear realization problem and the JMLS realization problem are new.

II. Rational Power Series

This section presents several results on formal power series [1], [18], [17]. These results will be used in §III for solving a generalized bilinear realization problem. In turn, the solution to this bilinear realization problem will yield a solution to the realization problem for JMLSs, as we will show in §IV.

A. *Definition and Basic Theory*

Let Σ be a finite set called the *alphabet*. The elements of Σ are called *letters*, and every finite sequence of letters is 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 $\sigma_1, \dots, \sigma_k \in \Sigma$. 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 $v = \nu_1 \cdots \nu_k$ and $w = \sigma_1 \cdots \sigma_m \in \Sigma^*$ is the word $vw = \nu_1 \cdots \nu_k \sigma_1 \cdots \sigma_m$. For any two sets J and A, an *indexed subset* of A with the *index set* J is a map $Z : J \to A$, denoted by $Z = \{a_j \in A \mid j \in J\}$, where $a_j = Z(j)$ for all $j \in J$. The elements a_j need not be different.

A *formal power series* S with *coefficients* in $\mathbb{R}^p$ is a map $S : \Sigma^* \to \mathbb{R}^p$. The values $S(w) \in \mathbb{R}^p$, $w \in \Sigma^*$, are called the coefficients of S. We denote by $\mathbb{R}^p \ll \Sigma^* \gg$ the set of all formal power series with coefficients in $\mathbb{R}^p$. A *family of formal power series* is an indexed set $\Psi = \{S_j \in \mathbb{R}^p \ll \Sigma^* \gg \mid j \in J\}$ with an arbitrary (not necessarily finite) index set J. A family of formal power series Ψ is called *rational* if there is 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 $B = \{B_j \in \mathbb{R}^n \mid j \in J\}$, such that for each $j \in J$ and for all sequences $\sigma_1, \dots, \sigma_k \in \Sigma$, $k \geq 0$,

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

The 4-tuple $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in \Sigma}, B, C)$ is called a *representation* of Ψ and the number $n = \dim R$ is called the *dimension* of R. A representation R_{min} of Ψ is 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 $\widetilde{R} = (\mathbb{R}^n, \{\widetilde{A}_\sigma\}_{\sigma \in \Sigma}, \widetilde{B}, \widetilde{C})$, are 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 the sequel, we will use the following short-hand notation $A_w \doteq A_{\sigma_k} A_{\sigma_{k-1}} \cdots A_{\sigma_1}$ for $w = \sigma_1 \cdots \sigma_k \in \Sigma^*$ and $\sigma_1, \dots, \sigma_k \in \Sigma$, $k \geq 0$. The map A_ϵ will be identified with the identity map. Denote by $W(n)$ the number of all words over Σ of length at most $n-1$. Define the *reachability matrix* of R by $W_R = [A_w B_j \mid w \in \Sigma^*, |w| \leq n-1, j \in J] \in \mathbb{R}^{n \times W(n) \cdot |J|}$ and the *observability matrix* of R by $O_R = [(CA_w)^T \mid w \in \Sigma^*, |w| \leq n-1]^T \in \mathbb{R}^{W(n)p \times n}$. We call the representation R *reachable* if $\dim R = \text{rank } W_R$, and *observable* if $\ker O_R = \{0\}$.

Let $\Psi = \{S_j \in \mathbb{R}^p \ll \Sigma^* \gg \mid j \in J\}$ be a family of formal power series. We define the Hankel-matrix of Ψ as the matrix $H_\Psi \in \mathbb{R}^{(\Sigma^* \times I) \times (\Sigma^* \times J)}$ whose entries are given by $(H_\Psi)_{(u,i)(v,j)} = (S_j(vu))_i$, where $I = \{1, 2, \dots, p\}$. That is, 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 $S_j(vu) \in \mathbb{R}^p$. The following result on realization of formal power series can be found in [18], [17], [13].

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

(i) Ψ is rational $\iff$ $\text{rank } H_\Psi < +\infty$.
(ii) R is a minimal representation of Ψ $\iff$ R is reachable and observable $\iff$ $\dim R = \text{rank } H_\Psi$.
(iii) All minimal representations of Ψ are isomorphic.
(iv) If $n = \text{rank } H_\Psi < +\infty$, then one can construct a representation of Ψ using the columns of H_Ψ (see [13] for details).

Notice that H_Ψ is an infinite matrix and hence the construction in part (iv) of Theorem 1 is not directly computable. However, it is possible to compute a minimal representation of Ψ from finitely many data using a generalization of the well-known Kalman-Ho partial realization algorithm for linear systems. One defines a 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 $\text{rank } H_{\Psi,N,N} = \text{rank } H_\Psi$ holds, then there exists an algorithm for computing a minimal representation R_N of Ψ by factorizing the matrix $H_{\Psi,N+1,N}$. The condition $\text{rank } H_{\Psi,N,N} = \text{rank } H_\Psi$ holds, if, for example, N is chosen to be bigger than the dimension of some representation of Ψ. More details on the computation of a minimal representation from a Hankel-matrix can be found in [13] and the references therein.

B. A Notion of Stability for Formal Power Series

To derive results on stochastic realization theory, we will need a notion of stability of a representation. To that end, consider a formal power series $S \in \mathbb{R}^p \ll \Sigma^* \gg$, and denote by $||\cdot||_2$ the Euclidean norm in $\mathbb{R}^p$. Consider the sequence, $L_n = \sum_{k=0}^{n} \sum_{\sigma_1 \in \Sigma} \cdots \sum_{\sigma_k \in \Sigma} ||S(\sigma_1 \sigma_2 \cdots \sigma_k)||_2^2$. The series S is called *square summable*, if the limit $\lim_{n \to +\infty} L_n$ exists and it is finite. We call the family $\Psi = \{S_j \in \mathbb{R}^p \ll \Sigma^* \gg \mid j \in J\}$ *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 $\Psi = \{S_j \in \mathbb{R}^p \ll \Sigma^* \gg \mid j \in J\}$. Assume that $\Sigma = \{\sigma_1, \dots, \sigma_d\}$, where d is the number of elements of Σ, and consider the matrix $\widetilde{A} = \sum_{i=1}^{d} A_{\sigma_i} \otimes A_{\sigma_i}$, 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 then have the following result.

Theorem 2 ([16]): A rational family of formal power series is square summable if and only if all minimal representations are stable.

III. Realization of Generalized Bilinear Systems

This section formulates and solves a stochastic realization problem for bilinear systems with nonwhite noise inputs using the results in §II. Particular cases of this generalized bilinear realization problem include realization of classical linear and bilinear systems. Also, by allowing finite-state Markov processes as inputs, we will obtain a solution to the realization problem for JMLSs, as we will show in §IV.

A. Generalized Bilinear Stochastic Realization Problem

Let the *output process* $\mathbf{y} \in \mathbb{R}^p$ be a wide-sense stationary and zero mean discrete-time (i.e. the time axis is $\mathbb{Z}$) stochastic process. Let the *input process* be a collection $\{\mathbf{u}_\sigma \in \mathbb{R}\}_{\sigma \in \Sigma}$ of discrete-time stochastic processes indexed by the elements of a finite alphabet Σ. For each nonempty word $w = \sigma_1\sigma_2 \cdots \sigma_k \in \Sigma^+$, $k \geq 1$, $\sigma_1, \ldots, \sigma_k \in \Sigma$, define

$$\mathbf{z}_w(t) = \mathbf{y}(t-k)\mathbf{u}_{\sigma_1}(t-k) \cdots \mathbf{u}_{\sigma_k}(t-1). \quad (2)$$

We call the random variables $\mathbf{z}_w(t), w \in \Sigma^+$ the *predictor variables*. We assume that the output and predictor variables $(\mathbf{y}(t), \{\mathbf{z}_w(t) \mid w \in \Sigma^+\})$ are jointly wide-sense stationary, i.e. for all $t, k \in \mathbb{Z}$, and for all $w, v \in \Sigma^+$ we have

$$E[\mathbf{y}(t+k)\mathbf{z}_w^T(t+k)] = E[\mathbf{y}(t)\mathbf{z}_w^T(t)], \quad \text{and} \quad (3)$$

$$E[\mathbf{z}_w(t+k)\mathbf{z}_v^T(t+k)] = E[\mathbf{z}_w(t)\mathbf{z}_v^T(t)]. \quad (4)$$

Notice that for any $p > 0$ the space $\mathcal{H}_p$ of zero-mean square-integrable random variables with values in $\mathbb{R}^p$ is a Hilbert-space with the scalar product $< \mathbf{x}, \mathbf{z} >= E[\mathbf{x}^T\mathbf{y}]$, see [2]. Recall the notions of closure and orthogonal projection for Hilbert-spaces. If Z is an arbitrary subset of $\mathcal{H}_p$ and x is an element of $\mathcal{H}_p$, then $E_l[x \mid Z]$ denotes the orthogonal projection of x onto the closure of the linear space spanned by the elements of Z. Notice that both the output $\mathbf{y}(t)$ and the predictors $\mathbf{z}_w(t)$ at time t belong to $\mathcal{H}_p$. Denote by $\mathcal{H}(t)$ the closure in $\mathcal{H}_p$ of the linear span of the predictors $\{\mathbf{z}_w(t) \mid w \in \Sigma^+\}$ at time t. We will call $\mathcal{H}(t)$ the *predictor space at time* t.

We are now ready to introduce our generalized bilinear stochastic realization problem.

Problem 1 (Generalized Bilinear Stochastic Realization): Given an output process $\mathbf{y}$ and an input process $\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ indexed by a given finite alphabet Σ, find a forward innovation state-space realization of $\mathbf{y}$ of the form

$$\begin{aligned} \mathbf{x}(t+1) &= \sum_{\sigma\in\Sigma}(A_\sigma\mathbf{x}(t) + K_\sigma\mathbf{e}(t))\mathbf{u}_\sigma(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) + \mathbf{e}(t), \end{aligned} \quad (5)$$

where the equalities are assumed to hold in the square-mean sense. In equation (5), the system matrices are of the form $A_\sigma \in \mathbb{R}^{n\times n}$, $K_\sigma \in \mathbb{R}^{n\times p}$, and $C \in \mathbb{R}^{p\times n}$ for all $\sigma \in \Sigma$, and $\mathbf{x}(t)$ is a random process taking values in $\mathbb{R}^n$ such that $C\mathbf{x}(t)$ is the orthogonal projection of the output $\mathbf{y}(t)$ onto the predictor space $\mathcal{H}(t)$, i.e. $C\mathbf{x}(t) = E_l[\mathbf{y}(t) \mid \mathcal{H}(t)]$ and $\mathbf{e}(t)$ is the forward innovation process

$$\mathbf{e}(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t) \mid \mathcal{H}(t)]. \quad (6)$$

Remark 1 (Realization of Linear Systems): If $\Sigma = \{z\}$ and $\mathbf{u}_z = 1$, then Problem 1 reduces to the classical linear realization problem and (5) becomes a linear state-space model in the forward innovation form.

Remark 2 (Realization of Bilinear Systems): If $\mathbf{u}_{z_1} = 1$, $\mathbf{u}_{z_2}$ is white noise, and $\Sigma = \{z_1, z_2\}$, then Problem 1 reduces to the classical bilinear realization problem and (5) becomes a bilinear state-space model in the forward innovation form.

B. Generalized Bilinear Stochastic Realization Theory

To solve the generalized bilinear stochastic realization problem, we will make a number of assumptions on the covariances between the output and predictor variables

$$\Lambda_w = E[\mathbf{y}(t)\mathbf{z}_w^T(t)] \in \mathbb{R}^{p\times p}, \quad \text{and} \quad (7)$$

$$T_{v,w} = E[\mathbf{z}_v(t)\mathbf{z}_w^T(t)] \in \mathbb{R}^{p\times p}. \quad (8)$$

Assumption 1 (Admissible words): Let L be a given set of non-empty words over Σ, i.e. $L \subseteq \Sigma^+$. We call L the *set of admissible words*. Every symbol $\sigma \in \Sigma$ is an element of L. Furthermore, if for some word $w \in \Sigma^+$ and letter $\sigma \in \Sigma$ the word $w\sigma \in L$ or $\sigma w \in L$, then $w \in L$. Also, if w is not admissible, i.e. $w \in \Sigma^+ \setminus L$, then $\mathbf{z}_w = 0$ and $\Lambda_w = 0$.

This assumption allows us to deal with the case where not every sequence of inputs is admissible. In particular, this will be the case for JMLSs, where the discrete state process will play the role of an input. We will discuss this case in §IV.

Assumption 2 (Square-summable formal power series): For each $j \in I = \{1, \ldots, p\}$ and $\sigma \in \Sigma$, define the formal power series $S_{(j,\sigma)} \in \mathbb{R}^p \ll\Sigma^*\gg$ as

$$S_{(j,\sigma)}(w) = (\Lambda_{\sigma w})_{.,j}, \quad (9)$$

where $(\Lambda_{\sigma w})_{.,j}$ denotes the jth column of the $p \times p$ covariance matrix $\Lambda_{\sigma w}$. Define the family of formal power series Ψ with the index set $J = I \times \Sigma$ as

$$\Psi = \{S_{(j,\sigma)} \mid j \in I, \sigma \in \Sigma\}. \quad (10)$$

We assume that Ψ is square summable.

Assumption 3 (Positive definiteness of finite covariance): For each $N > 0$, let $T^N = (T_{w,v})_{w,v\in L,|w|,|v|\leq N}$ be a finite covariance matrix formed by all matrices $T_{w,v}$ indexed by admissible words $w, v \in L$ of length at most N. For each $N > 0$, the matrix T^N is strictly positive definite, that is, for all $S \neq 0$, where $S_v \in \mathbb{R}^p$, we have $\sum_{w,v\in L,|w|,|v|\leq N} S_v T_{v,w} S_w > 0$.

This is mainly a technical condition, which simplifies the proofs. It is analogous to the assumption of the strict positive definiteness of the Toeplitz-matrix for the linear case.

Assumption 4 (Full rank innovation process): For each $\sigma \in \Sigma$ the covariance $E[\mathbf{e}(t)\mathbf{e}^T(t)\mathbf{u}_\sigma^2(t)]$ is of rank p.

This is also a technical assumption, which is used to obtain a nice expression for K_σ. For linear systems, it boils down to the classical requirement that $\mathbf{y}$ be a *full rank process* [2].

Assumption 5: There are nonzero reals $\{p_\sigma\}_{\sigma\in\Sigma}$ such that for all admissible words $w, v \in L$ satisfying $w\sigma, v\sigma' \in L$, and symbols $\sigma, \sigma' \in \Sigma$, we have

$$T_{w\sigma,v\sigma'} = \begin{cases} p_\sigma T_{w,v} & \sigma = \sigma' \\ 0 & \sigma \neq \sigma' \end{cases} \text{ and } T_{w\sigma,\sigma'} = \begin{cases} p_\sigma \Lambda_w^T & \sigma = \sigma' \\ 0 & \sigma \neq \sigma' \end{cases}.$$

In addition, if $w\sigma \in L$ then for all $v\sigma \notin L$, $T_{v,w} = 0$, and conversely, if $v\sigma \in L$, then for all $w\sigma \notin L$, $T_{v,w} = 0$.

This assumption is crucial for finding a time-invariant matrix K_σ. For linear systems, it follows from the wide-sense stationarity of the outputs. For bilinear systems, it follows from the assumption that the input is white noise.

Assumption 6: For all $t \in \mathbb{Z}$, $k \geq 0$, and $v \in \Sigma^+$, $\mathbf{y}(t-k)$ and $\mathbf{z}_v(t-k)$ belong to the closure (in the mean-square sense) of the linear space spanned by $\{\mathbf{z}_w(t), w \in \Sigma^+\}$.

This assumption is needed to ensure that the innovation processes are uncorrelated.

We the assumptions above, we have the following result.

Theorem 3: (Stochastic realization of bilinear systems with non-white inputs): Assume that the processes $\mathbf{y}$ and

$\{\mathbf{u}_\sigma\}_{\sigma\in\Sigma}$ satisfy Assumptions 1-6. Then, $\mathbf{y}$ has a realization by a generalized bilinear system of the form (5) if and only if Ψ is rational. Furthermore, the generalized bilinear stochastic realization problem has a solution of the form

$$\begin{aligned}\mathbf{x}(t+1) &= \sum_{\sigma\in\Sigma}(\frac{1}{p_\sigma}A_\sigma\mathbf{x}(t) + K_\sigma\mathbf{e}(t))\mathbf{u}_\sigma(t) \\ \mathbf{y}(t) &= C\mathbf{x}(t) + \mathbf{e}(t)\end{aligned} \tag{11}$$

where $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma\in\Sigma}, \{B_{(j,\sigma)}\}_{(j,\sigma)\in I\times\Sigma}, C)$ is a minimal representation of Ψ, $\mathbf{e}(t)$ is the (uncorrelated and zero-mean) innovation process defined in (6), and for each $\sigma \in \Sigma$

$$K_\sigma = (B_\sigma - \frac{1}{p_\sigma}A_\sigma P_\sigma C^T)(T_{\sigma,\sigma} - CP_\sigma C^T)^{-1} \tag{12}$$

where

$$P_\sigma = E[\mathbf{x}(t)\mathbf{x}^T(t)\mathbf{u}_\sigma(t)\mathbf{u}_\sigma(t)] \in \mathbb{R}^{n\times n} \quad \text{and} \tag{13}$$

$$B_\sigma = \begin{bmatrix} B_{(1,\sigma)}, & B_{(2,\sigma)}, & \dots, & B_{(p,\sigma)} \end{bmatrix} \in \mathbb{R}^{n\times p}. \tag{14}$$

Proof: [Sketch] Recall the definition of the observability matrix $O_R \in \mathbb{R}^{W(n)p\times n}$ of R. Define the $\mathbb{R}^{W(n)p}$-valued random variable $\mathbf{Y}_n(t) = [\mathbf{z}^f_w(t) \mid w \in \Sigma^*, |w| \le n-1]^T$, where $\mathbf{z}^f_\epsilon(t) = \mathbf{y}^T(t)$ and for all $w = \sigma_1\cdots\sigma_k \in \Sigma^+$, $\sigma_1,\dots,\sigma_k \in \Sigma$, $k > 0$, $\mathbf{z}^f_w(t) = \mathbf{y}^T(t+k)\mathbf{u}_{\sigma_k}(t+k-1)\mathbf{u}_{\sigma_{k-1}}(t+k-2)\cdots\mathbf{u}_{\sigma_1}(t)$. The variable $\mathbf{Y}_n(t)$ can be thought of as the products of future outputs and inputs. Since R is observable, the matrix O_R has a left inverse, which we will denote by O_R^{-1}. Then, we define the state $\mathbf{x}(t)$ as the linear projection of $O_R^{-1}(Y_n(t))$ onto the predictor space $\mathcal{H}(t)$. It follows that $C\mathbf{x}(t) = E_l[\mathbf{y}(t) \mid \mathcal{H}(t)]$. We are left with showing that $\mathbf{x}(t)$ satisfies the recursion in (11). To that end, let $\mathcal{H}_N(t)$ be the linear space spanned by $\mathbf{z}_w(t)$, where $w \in \Sigma^+$ and $|w| \le N$. It follows that $\mathbf{x}(t) = \lim_{N\to+\infty}\mathbf{x}_N(t)$, where $\mathbf{x}_N(t) = E_l[O_R^{-1}(Y_n(t)) \mid \mathcal{H}_N(t)]$. In turn, using Assumption 5 we can show that there exists a collection of matrices $\alpha_\sigma^{N+1} \in \mathbb{R}^{n\times p}$, such that $\mathbf{x}_{N+1}(t+1) = \sum_{\sigma\in\Sigma}(\frac{1}{p_\sigma}A_\sigma\mathbf{x}_N(t) + \alpha_\sigma^{N+1}\mathbf{e}_N(t))\mathbf{u}_\sigma(t)$, where $\mathbf{e}_N(t) = \mathbf{y}(t) - E_l[\mathbf{y}(t) \mid \mathcal{H}_N(t)]$. Thus, $\lim_{N\to+\infty}\mathbf{e}_N(t) = \mathbf{e}(t)$. Also, we can show that $\lim_{N\to+\infty}\alpha_\sigma^{N+1} = K_\sigma$ exists and satisfies (12). By taking the limits in the expression for $\mathbf{x}_{N+1}$ we obtain (11). The rest of the theorem can be shown using Assumptions 1–6. ∎

Theorem 3 implies the realization construction for linear [2] and bilinear stochastic systems [5]. Notice also that if $(\mathbf{y}, \{\mathbf{u}_w \mid w \in \Sigma^+\})$ are jointly ergodic, then we can replace stochastic processes with time series, and from the proof of Theorem 3 we get the following realization algorithm.

Algorithm 1: (Generalized Bilinear System Realization):

1) Let y and u_σ be the time-series corresponding to $\mathbf{y}$ and $\mathbf{u}_\sigma$, respectively, for $t = -N,\dots,0,\dots,N$. Compute the time-series z_w, z^f_w, and Y_n corresponding to the processes $\mathbf{z}_w$, $\mathbf{z}^f_w$, and $\mathbf{Y}_n$, respectively, by replacing all occurrences of $\mathbf{y}$ and $\mathbf{u}_\sigma$ with y and u_σ, respectively. For each stochastic process $\mathbf{s} \in \mathbb{R}^M$, let $P_N(\mathbf{s})(t) = (s(t),\dots,s(t+N-1)) \in \mathbb{R}^{M\times N}$, where s is the time-series associated with the process $\mathbf{s}$.
2) Approximate Λ_w by $\frac{1}{N}P_N(\mathbf{y})(0)(P_N(\mathbf{z}_w)(0))^T$ for all $w \in \Sigma^+$, $|w| \le N$. Construct a finite Hankel matrix from the approximations of Λ_w and use the algorithm described in Section II to obtain a representation R of Ψ in (9)–(10).
3) Compute the left inverse O_R^{-1} of the observability matrix of R. Let $x_N(0) \in \mathbb{R}^{n\times N}$ be the orthogonal projection of the rows of $O_R^{-1}(P_N(\mathbf{Y}_n)(0))$ onto the rows of the matrix $[(P_N(\mathbf{z}_w)(0))^T \mid w \in \Sigma^+, |w| \le N]^T$.
4) For $i = 1,\dots,N$, let $x_{N,i}$ be the ith column of x_N. Approximate P_σ by $\frac{1}{N}\sum_{i=1}^N x_{N,i}(0)x_{N,i}^T(0)u_\sigma^2(i-1)$, $T_{\sigma,\sigma}$ by $\frac{1}{N}P_N(\mathbf{z}_\sigma)(0)P_N(\mathbf{z}_\sigma^T)(0)$ and use (12) to compute K_σ.
5) Substitute K_σ and the matrices associated with the representation R into 11. As N goes to infinity, the matrices of the thus obtained realization converge entry-wise to the matrices of a generalized bilinear realization of $\widetilde{\mathbf{y}}$.

IV. Realization Theory of Generalized JMLSs

The goal of this section is to present a realization theory for JMLSs [4]. However, for the purposes of realization theory we will look at stochastic hybrid systems of a slightly more general form than the ones defined in [4].

A. Generalized Jump-Markov Linear Systems

A *generalized jump-Markov linear system* (GJMLS), H, is a system of the form

$$H : \begin{cases} \mathbf{x}(t+1) = M_{\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)}\mathbf{x}(t) + B_{\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)}\mathbf{v}(t) \\ \quad\mathbf{y}(t) = C_{\boldsymbol{\theta}(t)}\mathbf{x}(t) + D_{\boldsymbol{\theta}(t)}\mathbf{v}(t) \end{cases}. \tag{15}$$

Here $\boldsymbol{\theta}$, $\mathbf{x}$, $\mathbf{y}$ and $\mathbf{v}$ are stochastic processes defined on the 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 finite-state Markov process, with state-transition probabilities $p_{j,i} = \mathrm{Prob}(\boldsymbol{\theta}(t+1) = i \mid \boldsymbol{\theta}(t) = j) > 0$ for all $i, j \in Q$. Moreover, the probability distribution of the discrete state $\boldsymbol{\theta}(t)$ is given by $\pi = (\pi_1,\dots,\pi_d)^T$, where $\pi_i = \mathrm{Prob}(\boldsymbol{\theta}(t) = i)$. 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 system.

Assumption 7: Let $\mathcal{D}_t = \{\boldsymbol{\theta}(t-k)\}_{k\ge 0}$ be the collection of past discrete states, and denote by $E[\mathbf{z}|\mathcal{D}_t]$ the conditional expectation of $\mathbf{z}$ given $\mathcal{D}_t$. We assume that for all $t \in \mathbb{Z}$,
– $\mathbf{v}(t)$ is conditionally zero mean given $\mathcal{D}_t$, i.e. $E[\mathbf{v}(t) \mid \mathcal{D}_t] = 0$, and for all $l > 0$, $\mathbf{v}(t)$ and $\mathbf{v}(t-l)$ are conditionally uncorrelated given $\mathcal{D}_t$, i.e. $E[\mathbf{v}(t)\mathbf{v}(t-l)^T \mid \mathcal{D}_t] = 0$,
– the collections $\{\mathbf{x}(t-l), \mathbf{v}(t-l), l \ge 0\}$ and $\{\boldsymbol{\theta}(t+l), l > 0\}$ are conditionally independent given $\mathcal{D}_t$, and
– for all $l \ge 0$, $\mathbf{x}(t)$ and $\mathbf{v}(t+l)$ are conditionally uncorrelated given $\mathcal{D}_{t+l}$, i.e. $\forall l \ge 0$, $E[\mathbf{x}(t)\mathbf{v}(t+l) \mid \mathcal{D}_{t+l}] = 0$.

Assumption 8: Let $n = n_1 + n_2 + \cdots + n_d$. The matrix

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

is *stable*, i.e. for any eigenvalue λ of $\widetilde{M}$, we have $|\lambda| < 1$.

Assumption 9: The Markov process $\boldsymbol{\theta}$ is stationary and ergodic. Therefore, for all $q \in Q$, $\sum_{s\in Q}\pi_s p_{s,q} = \pi_q$.

Lemma 1: Let χ be the indicator function, i.e. $\chi(A) = 1$ if the event A is true, and $\chi(A) = 0$ otherwise. If Assumptions 7–8 hold, then there exists a unique collection of matrices $\{P_q \in \mathbb{R}^{n_q \times n_q}, q \in Q\}$, such that

$$P_q = \sum_{s \in Q} p_{s,q} M_{s,q} P_s M_{s,q}^T + B_{s,q} Q_{s,q} B_{s,q}^T, \qquad (16)$$

where $Q_{s,q} = E[\mathbf{v}(t)\mathbf{v}(t)^T \chi(\boldsymbol{\theta}(t+1) = q, \boldsymbol{\theta}(t) = s]$.
Lemma 1 is based on the well-known criteria for mean-square stability of JMLSs [4]. To make the continuous state and output processes $\mathbf{x}(t)$ and $\mathbf{y}(t)$ wide-sense stationary, we also need the following.

Assumption 10: Under Assumptions 7–8, let $\{P_q\}_{q \in Q}$, be the unique collection of matrices satisfying (16). Recall also the definition of $\mathcal{D}_t$ from Assumption 7. For all $t \in \mathbb{Z}$, $\mathbf{x}(t)$ is conditionally zero mean given $\mathcal{D}_t$, i.e. $E[\mathbf{x}(t) \mid \mathcal{D}_t] = 0$, and for all $q \in Q$, $E[\mathbf{x}(t)\mathbf{x}(t)^T \chi(\boldsymbol{\theta}(t) = q)] = P_q$.

Remark 3: Note that the classical definition of a discrete-time JMLS [4] differs from (15). 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. Nevertheless, the classical definition and (15) are equivalent in the following sense. On one hand, it is clear that a classical JMLS also satisfies our definition. Conversely, a GJMLS of the form (15) can be rewritten as a classical JMLS with the same noise and output processes, but with the discrete state process $\boldsymbol{\theta}$ replaced by $\widetilde{\boldsymbol{\theta}}(t) = (\boldsymbol{\theta}(t), \boldsymbol{\theta}(t+1))$ and the continuous state process and the system matrices replaced by a continuous state process and system matrices living in the continuous space $\mathbb{R}^{n_1+n_2+\cdots+n_d}$. The reason why we choose to work with GJMLSs of the form (15) instead of classical JMLSs is that, as we will show later, systems of the form (15) admit a nice realization theory. However, it is not clear if one can also obtain such results for classical JMLSs.

B. Existence of a Realization by a GJMLS

Let $\widetilde{\mathbf{y}}$ be a zero-mean wide-sense stationary process taking values in $\mathbb{R}^p$. Let $\boldsymbol{\theta}$ be a Markov-process taking values in $Q = \{1, \ldots, d\}$. Let H be a GJMLS of the form (15), with discrete state process $\boldsymbol{\theta}$ and output process $\mathbf{y}$, satisfying Assumptions 7-10. In the sequel we will keep $\boldsymbol{\theta}$ fixed and whenever we speak of a GJMLS realization of $\widetilde{\mathbf{y}}$, we will always mean a GJMLS of $\widetilde{\mathbf{y}}$ with discrete state process $\boldsymbol{\theta}$.

Definition 1 (Realization by GJMLSs): The GJMLS H is said to be a *realization* of $\widetilde{\mathbf{y}}$ if the continuous output process $\mathbf{y}$ of H is equals to $\widetilde{\mathbf{y}}$ in the square-mean sense, that is, for any time instant $t \in \mathbb{Z}$, $E[(\widetilde{\mathbf{y}}(t) - \mathbf{y}(t))^T (\widetilde{\mathbf{y}}(t) - \mathbf{y}(t))] = 0$.

This section presents necessary and sufficient conditions for existence of a realization of the output process $\widetilde{\mathbf{y}}$ by a GJMLS with discrete state process $\boldsymbol{\theta}$. The construction proceeds by associating a generalized bilinear system $\mathcal{B}$ to the processes $\widetilde{\mathbf{y}}$ and $\boldsymbol{\theta}$ and building a formal power series associated with the covariance sequence of $\mathcal{B}$.

Let the alphabet Σ of $\mathcal{B}$ 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 $\mathcal{B}$ be defined as

$$\mathbf{u}_{(q_1,q_2)}(t) = \chi(\boldsymbol{\theta}(t+1) = q_2, \boldsymbol{\theta}(t) = q_1). \qquad (17)$$

For each nonempty word $w = \sigma_1 \cdots \sigma_k \in \Sigma^+$, $\sigma_1, \ldots, \sigma_k \in \Sigma$, define the predictor variables as in (2), except that the output $\mathbf{y}$ is replaced by the given process $\widetilde{\mathbf{y}}$, i.e. $\mathbf{z}_w(t) = \widetilde{\mathbf{y}}(t-k)\mathbf{u}_{\sigma_1}(t-k) \cdots \mathbf{u}_{\sigma_k}(t-1)$. Notice that if w is not of the form $w = (q_0, q_1)(q_1, q_2) \cdots (q_{k-1}, q_k)$, for $k \geq 0$ and $q_0, \ldots, q_k \in Q$, then $\mathbf{z}_w(t) = 0$. This prompts us to define the set of admissible sequences L (see Assumption 1) as

$$L = \{(q_0, q_1)(q_1, q_2) \cdots (q_{k-1}, q_k) \mid k > 0, q_1, \ldots, q_k \in Q\}. \qquad (18)$$

Notice that if $w = (q_0, q_1)(q_1, q_2) \cdots (q_{k-1}, q_k) \in L$, then the covariance $\Lambda_w = E[\widetilde{\mathbf{y}}(t)\mathbf{z}_w^T(t)]$ can be written as $\Lambda_w = E[\widetilde{\mathbf{y}}(t+k)\widetilde{\mathbf{y}}^T(t)\chi(\boldsymbol{\theta}(t+i) = q_i, i = 0, \ldots, k)]$. As in (9), we can associate the covariance sequence Λ_w with a family of formal power series $\{S_{(i,\sigma)} \in \mathbb{R}^p \ll \Sigma^* \gg \mid \sigma \in Q \times Q, i = 1, \ldots, p\}$, where $S_{i,\sigma}(w)$ is the ith column of $\Lambda_{\sigma w}$. We denote this family by $\Psi_{\widetilde{\mathbf{y}}}$ to emphasize it depends on $\widetilde{\mathbf{y}}$.

In order to find necessary and sufficient conditions for existence of a GJMLS realization for $\widetilde{\mathbf{y}}$ we need to make a number of assumptions on $\widetilde{\mathbf{y}}$ and $\boldsymbol{\theta}$.

Assumption 11 (Conditional independence of $\widetilde{\mathbf{y}}$ and $\boldsymbol{\theta}$): For each $t \in \mathbb{Z}$, the collection of random variables $\{\widetilde{\mathbf{y}}(t-l), l \geq 0\}$ and $\{\boldsymbol{\theta}(t+l) \mid l > 0\}$ are conditionally independent given $\{\boldsymbol{\theta}(t-l) \mid l \geq 0\}$.

Assumption 12 (Stability of $\Psi_{\widetilde{\mathbf{y}}}$): The family of formal power series $\Psi_{\widetilde{\mathbf{y}}}$ is square summable.

Assumption 13 (Ergodicity and strong connectedness): The Markov process $\boldsymbol{\theta}$ is stationary, ergodic and for each $q_1, q_2 \in Q$ the transition probability $p_{q_1,q_2} > 0$ is nonzero.

Assumption 14 (Positive definiteness of finite covariance): For each $w, v \in L$, let $T_{v,w} = E[\mathbf{z}_v(t)\mathbf{z}_w^T(t)]$. We assume that the finite matrix $T^N = (T_{w,v})_{w,v \in L, |w|,|v| \leq N}$ formed from admissible words of length at most $N > 0$ is strictly positive definite, i.e. it satisfies Assumption 3 in Section III.

Assumption 15 (Full-rank predictor space): The innovation process $\mathbf{e}(t) = \widetilde{\mathbf{y}}(t) - E_l[\widetilde{\mathbf{y}}(t) \mid \{\mathbf{z}_w(t) \mid w \in \Sigma^+\}]$ is full-rank, i.e. it satisfies Assumption 4 in Section III.

The following lemmas characterize the relationships among Assumptions 11–13 and Assumptions 1–10.

Lemma 2: If $\widetilde{\mathbf{y}}$ has a realization by a GJMLS for which Assumptions 7–10 hold, then Assumptions 11–13 hold. Also, if $\widetilde{\mathbf{y}}$ and $\boldsymbol{\theta}$ satisfy Assumptions 11–15, then they satisfy Assumptions 1–6. In particular, Assumption 5 is satisfied with p_σ defined as the transition probability of the Markov process $\boldsymbol{\theta}$, that is, for $\sigma = (q_1, q_2)$, we let $p_{(q_1,q_2)} = p_{q_1,q_2}$.

We are now ready to formulate the main theorem.

Theorem 4 (Existence of a GJMLS Realization): Assume that $\widetilde{\mathbf{y}}$ and $\boldsymbol{\theta}$ satisfy Assumptions 11–15. Then $\widetilde{\mathbf{y}}$ has a realization by a GJMLS system if and only if $\Psi_{\widetilde{\mathbf{y}}}$ is rational.

Proof: [Sketch] We first show that if $\Psi_{\widetilde{\mathbf{y}}}$ is rational, then $\widetilde{\mathbf{y}}$ has a realization by a GJMLS. If Ψ is rational, then we can find a minimal representation $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in \Sigma}, \{B_{(i,\sigma)}\}_{i \in I, \sigma \in \Sigma}, C)$ of $\Psi_{\widetilde{\mathbf{y}}}$. Then, by applying Theorem 3, we can obtain a generalized bilinear realization of $\widetilde{\mathbf{y}}$ of the form (11). Based on this realization we can define a GJMLS realization H_R of $\widetilde{\mathbf{y}}$ of the form

$$H_R : \begin{cases} \hat{\mathbf{x}}(t+1) = M_{\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)}\hat{\mathbf{x}}(t) + K_{\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)}\mathbf{e}(t) \\ \widetilde{\mathbf{y}}(t) = C_{\boldsymbol{\theta}(t)}\hat{\mathbf{x}}(t) + \mathbf{e}(t), \end{cases} \qquad (19)$$

where
– *Continuous state-spaces.* For each $q \in Q$ define $\mathcal{X}_q \subseteq \mathbb{R}^n$ as the subspace spanned by $A_{(q_1,q)} A_w B_{(i,\sigma)}$ and $B_{(i,(q_1,q))}$ for all $q_1 \in Q$, $w \in \Sigma^+$, $\sigma \in \Sigma$, $i = 1, \ldots, p$. Then identify the elements of $\mathcal{X}_q$ with the elements of $\mathbb{R}^{n_q}$, where $n_q = \dim \mathcal{X}_q$.
– *State process.* Obtain the continuous state process $\hat{\mathbf{x}}(t)$ of the GJMLS from the continuous state $\mathbf{x}(t)$ of the generalized bilinear system (11) by viewing $\mathbf{x}(t)$ as an element of $\mathcal{X}_{\boldsymbol{\theta}(t)}$ and identifying it with the corresponding vector in $\mathbb{R}^{n_q}$ for $q = \boldsymbol{\theta}(t)$.
– *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 the matrix associated with the linear map

$$\mathcal{X}_{q_1} \ni x \mapsto \frac{1}{p_{q_1,q_2}} A_{q_1,q_2} x \in \mathcal{X}_{q_2}. \tag{20}$$

For all $q \in Q$, $C_q \in \mathbb{R}^{p \times n_q}$ is the matrix associated with the linear map $\mathcal{X}_q \ni x \mapsto Cx \in \mathbb{R}^p$, i.e. C_q is the restriction of C to $\mathcal{X}_q$.
– *Noise process.* The noise process is the innovation process

$$\begin{aligned}\mathbf{e}(t) = \widetilde{\mathbf{y}}(t) - E_l[\widetilde{\mathbf{y}}(t) \mid \{\widetilde{\mathbf{y}}(t-k)\chi(\boldsymbol{\theta}(t-k) = q_0) \cdots \\ \chi(\boldsymbol{\theta}(t) = q_k) \mid q_0, \ldots, q_k \in Q, k > 0\}].\end{aligned} \tag{21}$$

– *Noise gain.* The matrix $K_{q_1,q_2} \in \mathbb{R}^{n_{q_2} \times p}$ is defined as

$$\begin{aligned}K_{q_1,q_2} = &(B_{q_1,q_2} - M_{q_1,q_2} P_{q_1,q_2} C_{q_1}^T) \times \\ &\times (T_{(q_1,q_2),(q_1,q_2)} - C_{q_1} P_{q_1,q_2} C_{q_1}^T)^{-1}.\end{aligned} \tag{22}$$

In this expression, $T_{(q_1,q_2),(q_1,q_2)} \in \mathbb{R}^{p \times p}$ is the self covariance

$$T_{(q_1,q_2),(q_1,q_2)} = E[\widetilde{\mathbf{y}}(t)\widetilde{\mathbf{y}}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)]. \tag{23}$$

Moreover, the matrix $P_{q_1,q_2} \in \mathbb{R}^{n_{q_1} \times n_{q_1}}$ is the self covariance

$$P_{q_1,q_2} = E[\hat{\mathbf{x}}(t)\hat{\mathbf{x}}^T(t)\chi(\boldsymbol{\theta}(t) = q_1, \boldsymbol{\theta}(t+1) = q_2)]. \tag{24}$$

Finally, the matrix $B_{q_1,q_2} \in \mathbb{R}^{n_{q_2} \times p}$ is defined as

$$B_{q_1,q_2} = \begin{bmatrix} B_{(1,(q_1,q_2))}, & \cdots, & B_{(p,(q_1,q_2))} \end{bmatrix}, \tag{25}$$

where each vector $B_{(i,(q_1,q_2))}$ is an element of $\mathcal{X}_{q_2}$ and hence can be identified uniquely with a vector in $\mathbb{R}^{n_{q_2}}$.

The system H_R is a well-defined GJMLS and it satisfies Assumptions 7–10. We will call H_R the *GJMLS associated with the representation* R.

We now show that if $\widetilde{\mathbf{y}}$ has a GJMLS realization, then $\Psi_{\widetilde{\mathbf{y}}}$ is rational. To that end, assume that H is a GJMLS of the form (15) satisfying Assumptions 7–10. We will define a representation R_H, referred to as the *representation* R_H *associated with* H, such that R_H is a representation of $\Psi_{\widetilde{\mathbf{y}}}$. We define R_H as

$$R_H = (\mathbb{R}^n, \{A_{(q_1,q_2)}\}_{(q_1,q_2) \in \Sigma}, B, C), \tag{26}$$

where the parameters of R_H are given by
– *State-space.* Let $n = n_1 + n_2 + \cdots n_d$ and let the state-space of R_H be $\mathbb{R}^n$. Notice that $\mathbb{R}^n$ can be viewed as a direct sum of the individual state-spaces $\mathcal{X}_q = \mathbb{R}^{n_q}$, i.e. $\mathbb{R}^n = \bigoplus_{q \in Q} \mathcal{X}_q$, hence each $\mathcal{X}_q$ can be viewed as a subspace of $\mathbb{R}^n$.
– *Matrices* A_{q_1,q_2}. For each $q_1, q_2 \in Q$, let $A_{q_1,q_2} \in \mathbb{R}^{n \times n}$ be the matrix defined by the following property: if $x \in \mathcal{X}_{q_1}$, then $A_{q_1,q_2} x = p_{q_1,q_2} M_{q_1,q_2} x \in \mathcal{X}_{q_2} \subseteq \mathbb{R}^n$ and if $x \in \mathcal{X}_q, q \neq q_1$, then $A_{q_1,q_2} x = 0$.
– *Matrix* C. The $p \times n$ matrix C is defined by the following property; for all $x \in \mathcal{X}_q$, $Cx = C_q x$.
– *Initial states* B. Define the family $B = \{B_{(i,(q_1,q_2))} \mid q_1, q_2 \in Q, i = 1, \ldots, p\}$ as follows. For each $q_1, q_2 \in Q$, $i = 1, \ldots, p$, $B_{(i,(q_1,q_2))}$ is the ith column of the $n \times p$ matrix

$$\begin{bmatrix} \delta_{1,q_2} G_{q_1,q_2}^T & \delta_{1,q_2} G_{q_1,q_2}^T & \cdots & \delta_{d,q_2} G_{q_1,q_2}^T \end{bmatrix}^T.$$

In this equation $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$, and

$$G_{q_1,q_2} = p_{q_1,q_2} M_{q_1,q_2} P_{q_1} C_{q_1}^T + B_{q_1,q_2} W_{q_1,q_2} D_{q_1}^T$$

where $W_{q_1,q_2} = E[\mathbf{v}(t)\mathbf{v}^T(t)\chi(\boldsymbol{\theta}(t+1) = q_2, \boldsymbol{\theta}(t) = q_1)]$ and $P_{q_1} \in \mathbb{R}^{n_{q_1} \times n_{q_1}}$ is defined by (16). Notice that $G_{q_1,q_2} = E[\mathbf{x}(t)\widetilde{\mathbf{y}}^T(t-1)\chi(\boldsymbol{\theta}(t) = q_2, \boldsymbol{\theta}(t-1) = q_1)]$.

The proof that H_R is a GJMLS realization of $\widetilde{\mathbf{y}}$ then follows from Theorem 3. ∎

C. Minimality of a Realization by a GJMLS

As in the case of linear systems, it is possible that several GJMLSs realize a given process $\widetilde{\mathbf{y}}$. Hence, we are interested in finding a realization of $\widetilde{\mathbf{y}}$ that is minimal in some sense. We define the notion of a minimal realization as follows.

Definition 2 (Minimal Realization by 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

$$\dim H = n_1 + n_2 + \cdots + n_d, \tag{27}$$

where n_q is the dimension of the continuous state-space associated with discrete state q, i.e. $n_q = \dim \mathcal{X}_q$, for $q \in Q$. We call a realization H of $\widetilde{\mathbf{y}}$ *minimal* if $\dim H \leq \dim H'$ for all GJMLSs H' that are realizations of $\widetilde{\mathbf{y}}$.

In the case of linear systems, a realization is minimal if and only if it is reachable and observable [2]. In this subsection, we will formulate similar concepts for GJMLSs. 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.

To that end, let H be a given GJMLS of the form (15) that satisfies Assumptions 7–10. Let $N = \dim H$ be the dimension of H. For all $(q_1, q_2) \in Q \times Q = \Sigma$ let

$$G_{q_1,q_2} = E[\mathbf{x}(t)\mathbf{y}^T(t-1)\chi(\boldsymbol{\theta}(t) = q_2, \boldsymbol{\theta}(t-1) = q_1)] \tag{28}$$

be a matrix in $\mathbb{R}^{n_{q_2} \times p}$. Recall the definition of $L \subset (Q \times Q)^+$ from (18). For any admissible word $w = \sigma_1 \cdots \sigma_k \in L$, where $\sigma_i = (q_i, q_{i+1}) \in \Sigma$ for $i = 1, \ldots, k-1$, let

$$M_w = M_{q_{k-1},q_k} M_{q_{k-2},q_{k-1}} \cdots M_{q_1,q_2} \in \mathbb{R}^{n_{q_k} \times n_{q_1}}. \tag{29}$$

If $w \notin L$, and $|w| > 0$, then M_w denotes the zero matrix. If $w = \epsilon$, then M_ϵ denotes the identity matrix whose domain of definition depends on the context. For each $q \in Q$, let $L^q(N)$ be the set of all words in L of length at most N that end in some pair whose second component is q, i.e. $L^q(N)$ is the set of all words in $w \in L$ such that $|w| \leq N$ and $w = v(q_1, q)$ for some $q_1 \in Q$ and $v \in \Sigma^*$. Similarly, for each $q \in Q$, let $L_q(N)$ be the set of 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 words $w \in L$ such that $|w| \leq N$ and $w = (q, q_2)v$ for some $q_2 \in Q$ and $v \in \Sigma^*$.

We define reachability and observability as follows.

Definition 3 (Reachability of GJMLS): For each discrete state $q \in Q$, define the matrix $R_{H,q} \in \mathbb{R}^{n_q \times |L^q(N)| p}$ as

$$[M_v G_{q_1,q_2} \,|\, q_1 \in Q, q_2 \in Q, v \in \Sigma^*, (q_1, q_2)v \in L^q(N)].$$

We say that the GJMLS H is *reachable*, if for each discrete state $q \in Q$, $\text{rank}\ (R_{H,q}) = n_q$.

Definition 4 (Observability of GJMLS): For each discrete state $q \in Q$, define the matrix $O_{H,q} \in \mathbb{R}^{|L_q(N)| p \times n_q}$ as

$$[(C_{q_k} M_v)^T \mid \sigma \in Q, q_k \in Q, v \in \Sigma^*, v(\sigma, q_k) \in L_q(N)]^T.$$

We say that a GJMLS H is *observable*, if for each discrete state $q \in Q$, $\mathrm{rank}\ (O_{H,q}) = n_q$.

Recall from (26) the definition of the representation R_H associated with a GMJLS H. Recall also the definition of reachability of a representation along with the definition of the observability matrix O_{R_H} of R_H. Observability and reachability of H can be characterized in terms of the observability and reachability of its representation R_H.

Lemma 3: The GJMLS H is reachable if and only if R_H is reachable, and H is observable if and only if for each $q \in Q$, $\ker O_{R_H} \cap \mathcal{X}_q = \{0\}$.

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

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

Theorem 5 (Minimality of a realization by a GJMLS): Let $\widetilde{\mathbf{y}}$ be an output process satisfying Assumptions 11–15. A GJMLS H of the form (15) is a minimal realization of $\widetilde{\mathbf{y}}$ if and only if it is reachable and observable. In addition, if $\hat{H}$ is another GJMLS realization of $\widetilde{\mathbf{y}}$ given by

$$\begin{aligned} \hat{\mathbf{x}}(t+1) &= \hat{M}_{\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)}\hat{\mathbf{x}}(t) + \hat{B}_{\boldsymbol{\theta}(t),\boldsymbol{\theta}(t+1)}\hat{\mathbf{v}}(t) \\ \widetilde{\mathbf{y}}(t) &= \hat{C}_{\boldsymbol{\theta}(t)}\hat{\mathbf{x}}(t) + \hat{D}_{\boldsymbol{\theta}(t)}\hat{\mathbf{v}}(t), \end{aligned} \tag{30}$$

where the dimension of the continuous state-space of $\hat{H}$ corresponding to the discrete state q is $\hat{n}_q$, then the GJMLS $\hat{H}$ is minimal if and only if $n_q = \hat{n}_q$ for all $q \in Q$. Furthermore, there exists a collection of nonsingular matrices, $T_q \in \mathbb{R}^{n_q \times n_q}$, $q \in Q$, such that for all $q_1, q_2 \in Q$

$$T_{q_2}M_{q_1,q_2}T_{q_1}^{-1} = \hat{M}_{q_1,q_2}, C_{q_1}T_{q_1}^{-1} = \hat{C}_{q_1}, T_{q_2}G_{q_1,q_2} = \hat{G}_{q_1,q_2},$$

where

$$\begin{aligned} G_{q_1,q_2} &= E[\mathbf{x}(t)\widetilde{\mathbf{y}}^T(t-1)\chi(\boldsymbol{\theta}(t) = q_2, \boldsymbol{\theta}(t-1) = q_1)], \\ \hat{G}_{q_1,q_2} &= E[\hat{\mathbf{x}}(t)\widetilde{\mathbf{y}}^T(t-1)\chi(\boldsymbol{\theta}(t) = q_2, \boldsymbol{\theta}(t-1) = q_1)]. \end{aligned}$$

That is, $\hat{H}$ and H are algebraically similar in some sense.

Proof: [Sketch] First, we can show that if R is a minimal representation of $\Psi_{\widetilde{\mathbf{y}}}$, then the GJMLS H_R defined in the proof of Theorem 4 is a minimal realization of $\widetilde{\mathbf{y}}$. It then follows from Lemma 3 that H_R is reachable and observable. Moreover, if H is a realization of $\widetilde{\mathbf{y}}$ which is not reachable or observable, then we can show that $\dim H_R < \dim H$. Hence, minimality implies observability and reachability.

Assume now that $\hat{H}$ is a reachable and observable GJMLS realization of $\widetilde{\mathbf{y}}$. Consider the representation $R_{\hat{H}}$ and transform it to a minimal representation $\hat{R}$. Construct the GJMLS $H_{\hat{R}}$ and notice that $H_{\hat{R}}$ is minimal. From reachability and observability of H it follows that $\hat{H}$ and $H_{\hat{R}}$ are isomorphic and hence $\hat{H}$ is minimal. Hence, reachability and observability implies minimality. The rest of the theorem follows from the properties of minimal rational representations. ■

Remark 4: Notice that in (5) we do not require any relationship between B_{q_1,q_2} and $\hat{B}_{q_1,q_2}$. This is consistent with the situation for linear stochastic systems.

Remark 5 (GJMLS Realization Algorithm): It is clear that reachability and observability, and hence minimality, of a GJMLS can be checked numerically. It is also easy to see that Algorithm 1 can be adapted to obtain the realization H_R described in Theorem 4.

V. Discussion and Conclusion

We presented a realization theory for stochastic JMLSs. The theory relies on the solution of a generalized bilinear realization problem. This solution represents an extension of the known results on linear and bilinear stochastic realization. Open research avenues include extending our results to more general classes of hybrid systems. In particular, it would be interesting to develop realization theory for jump-Markov linear systems with partially observed discrete states. Necessary conditions for existence of a realization by a system of this class were already presented in [16]. Another interesting line of research is to use the presented theory for developing subspace identification algorithms for stochastic JMLSs. Note that the classical stochastic bilinear realization theory gave rise to a number of subspace identification algorithms, see [6], [20], [3]. It is very likely that the presented results will lead to very similar subspace identification algorithms.

Acknowledgements

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

References

[1] J. Berstel and C. Reutenauer. *Rational series and their languages*. Springer-Verlag, 1984.

[2] P.E. Caines. *Linear Stochastic Systems*. John Wiley and Sons, 1998.

[3] H. Chen and J.M. Maciejowski. Subspace identification of deterministic bilinear systems. In *CDC*, volume 3, pages 1797–1801, 2000.

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

[5] U.B. Desai. Realization of bilinear stochastic systems. *IEEE Transactions on Automatic Control*, 31(2):189–192, 1986.

[6] M. Favoreel. *Subspace methods for identification and control of linear and bilinear systems*. PhD thesis, K.U. Leuven, 1999.

[7] R.L. Grossman and R.G. Larson. An algebraic approach to hybrid systems. *Theoretical Computer Science*, 138:101–112, 1995.

[8] B. Jakubczyk. Existence and uniqueness of realizations of nonlinear systems. *SIAM Journal on Control and Optimization*, 18(4):455–471.

[9] Mihály Petreczky. Realization theory for bilinear hybrid systems. In *Conf. on Methods and Models in Automation and Robotics*, 2005.

[10] Mihály Petreczky. Realization theory for bilinear switched systems. In *IEEE Conference on Decision and Control*, pages 690–695, 2005.

[11] Mihály Petreczky. Hybrid formal power series and their application to realization theory of hybrid systems. In *MTNS*, 2006.

[12] Mihály Petreczky. Realization theory for discrete-time piecewise-affine hybrid systems. In *MTNS*, 2006.

[13] Mihály Petreczky. *Realization Theory of Hybrid Systems*. PhD thesis, Vrije Universiteit, Amsterdam, 2006.

[14] Mihály Petreczky. Realization theory for linear switched systems: Formal power series approach. *Systems and Control Letters*, 2007.

[15] M. Petreczky and J.B. Pomet. Realization theory of nonlinear hybrid systems. In *Workshop on Embedded and Hybrid Control*, 2006.

[16] Mihály Petreczky and René Vidal. Metrics and topology for nonlinear and hybrid systems. In *Hybrid Systems: Computation and Control*. Springer Verlag, 2007.

[17] E. Sontag. *Polynomial Response Maps*, volume 13 of *Lecture Notes in Control and Information Sciences*. Springer Verlag, 1979.

[18] E. Sontag. Realization theory of discrete-time nonlinear systems: Part I - the bounded case. *IEEE Trans. Circ. and Sys.*, 26(5):342–356, 1979.

[19] H. Sussmann. Existence and uniqueness of minimal realizations of nonlinear systems. *Mathematical Systems Theory*, 3:263–284, 1977.

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

[21] Y. Wang and E. Sontag. Realization and input/output relations: the analytic case. In *Conf. on Dec. and Control*, pages 1975–1980, 1989.

[22] Y. Wang and E. Sontag. Algebraic differential equations and rational control systems. *SIAM Journal on Control and Optimization*, (30):1126–1149, 1992.