Metrics and Topology for Nonlinear and Hybrid Systems

Mihály Petreczky and René Vidal

Center for Imaging Science, Johns Hopkins University, Baltimore MD 21218, USA
`{mihaly,rvidal}@cis.jhu.edu`

Abstract. This paper presents an approach to defining distances between nonlinear and hybrid dynamical systems based on formal power series theory. The main idea is that the input-output behavior of a wide range of dynamical systems can be encoded by rational formal power series. Hence, a natural distance between dynamical systems is the distance between the formal power series encoding their input-output behavior. The paper proposes several computable distances for rational formal power series and discusses the application of such distances to various classes of nonlinear and hybrid systems. In particular, the paper presents a detailed discussion of distances for stochastic jump-linear systems.

1 Introduction

In this paper, we present several possible definitions of a computable distance for rational formal power series and their representations. The main motivation for studying distances between rational formal power series is that the input-output behavior of various classes of dynamical systems can be encoded by rational formal power series. For example, linear and bilinear systems, switched linear and bilinear systems, finite state hidden Markov models, and linear and bilinear hybrid systems [1–5]. Therefore, one can define a distance between two dynamical systems as the distance between the formal power series that encode the input-output behavior of the systems. By construction, the proposed distances will be invariant under any transformation that preserves the input-output behavior of the systems. In addition, restricting attention to *rational* formal power series will enable us to compute the distances by using the rational representations of the formal power series. In general, such rational representations can easily be computed from the dynamical system. Another advantage of the proposed approach is that it connects well with identification and realization theory, because several identification methods are based on realization theory and hence on finding an appropriate rational representation of a family of formal power series.

Endowing the space of dynamical systems with a topology and a metric not only is an interesting theoretical exercise, but also has several interesting applications. A classical application is in system identification, more precisely, in finding a continuous parameterization and suitable canonical forms of dynamical systems [6–9]. Another important application comes from the field of computer vision, where one is interested in automatically recognizing different types of motions in a video sequence. That is, given a sequence of images depicting moving objects and people at different time instances, we would like to determine automatically the object class, the person identify, and the type of motion we see in the video sequence. For instance, we would like to determine

whether the video sequence depicts a running person or a galloping deer. One of the traditional mathematical tools for recognition and classification is machine learning. However, many of the classical machine learning techniques require a metric on the observation space. Since our observations are video sequences depicting multiple motions, it is rather natural to model such videos as the output of one or more dynamical systems, where each dynamical system describes a particular motion. Our observations are then outputs of dynamical systems, or, after an identification procedure, dynamical systems themselves. Therefore, in order to apply machine learning algorithms for recognizing motions in video sequences, one needs to define a suitable metric and topology on the space of dynamical systems. The study of topological and metric properties of dynamical systems from this point of view is a relatively recent development, see [10–13].

The outline of the paper is as follows. Section 2 presents the background material on the theory of rational formal power series. Section 3 presents the definition of several possible distances for rational formal power series and their rational representations. Section 4 discusses the relationship between formal power series and various classes of dynamical systems. In particular, it presents a detailed description of this relationship as well as a distance for stochastic jump-linear systems. Section 5 discusses the relationship between the results of the current paper and earlier results in the literature, as well as issues concerning the practical computability of the defined distances.

2   Rational Power Series

This section presents several results on formal power series that will be used throughout the rest of the paper. The material in Subsections 2.1 and 2.2 can be found in [1]. The results in Subsection 2.3 are, to the best of our knowledge, new. For more details on the classical theory of rational formal power series, the reader is referred to [14, 3, 15].

2.1   Definition and Basic Theory

Let X be a finite set. We will refer to X as the *alphabet*. The elements of X will be called *letters*, and every finite sequence of letters will be called a *word* or *string* over X. Denote by X^* the set of all finite words from elements in X. An element $w \in X^*$ of length $|w| = k \geq 0$ is a finite sequence $w = w_1 w_2 \cdots w_k$ with $w_1, \ldots, w_k \in X$. The empty word is denoted by ϵ and its length is zero, i.e. $|\epsilon| = 0$. The concatenation of two words $v = v_1 \cdots v_k$ and $w = w_1 \cdots w_m \in X^*$ is the word $vw = v_1 \cdots v_k w_1 \cdots w_m$.

For any two sets J and A, an *indexed subset* of A with the *index set* J is simply a map $Z : J \to A$, denoted by $Z = \{a_j \in A \mid j \in J\}$, where $a_j = Z(j)$ for all $j \in J$. Notice that we do not require the elements a_j to be all different.

A *formal power series* S with coefficients in $\mathbb{R}^p$ is a map $S : X^* \to \mathbb{R}^p$. We will call the values $S(w) \in \mathbb{R}^p$, $w \in X^*$, the *coefficients* of S. We denote by $\mathbb{R}^p \ll X^* \gg$ the set of all formal power series with coefficients in $\mathbb{R}^p$. Consider the indexed set of formal power series $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ with an arbitrary (not necessarily 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 is called *rational* if there exists an integer $n \in \mathbb{N}$, a matrix $C \in \mathbb{R}^{p \times n}$, a collection of matrices $A_\sigma \in \mathbb{R}^{n \times n}$ where $\sigma \in X$

runs through all elements of X, 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 $\sigma_1, \ldots, \sigma_k \in X$, $k \geq 0$,

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

The 4-tuple $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in X}, B, C)$ is called a *representation* of Ψ, and the number $n = \dim R$ is called the *dimension* of the representation R. If $S \in \mathbb{R}^p \ll X^* \gg$ is a single power series, then S will be called *rational* if the singleton set $\{S\}$ is rational, and by a representation of S we will mean a representation of $\{S\}$. 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, $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in X}, B, C)$ and $\widetilde{R} = (\mathbb{R}^n, \{\widetilde{A}_\sigma\}_{\sigma \in X}, \widetilde{B}, \widetilde{C})$, 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 $\sigma \in X$, $T\widetilde{B}_j = B_j$ for all $j \in J$, and $\widetilde{C} = CT$.

Let $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in X}, B, C)$ be a representation of Ψ. In the sequel, we will use the following short-hand notation $A_w \doteq A_{w_k} A_{w_{k-1}} \cdots A_{w_1}$ for $w = w_1 \cdots w_k \in X^*$. A_ϵ will be identified with the identity map. The representation R is called *observable* if $O_R = \{0\}$ and *reachable* if $\dim R = \dim W_R$, where W_R and O_R are the following subspaces of $\mathbb{R}^n$

$$W_R = \mathrm{Span}\{A_w B_j \mid w \in X^*, |w| \leq n - 1, j \in J\} \quad \text{and} \quad O_R = \bigcap_{w \in X^*, |w| \leq n-1} \ker C A_w. \tag{2}$$

Observability and reachability of representations can be checked numerically. One can formulate an algorithm for transforming any representation to a minimal representation of the same family of formal power series (see [1] and the references therein for details).

Let $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ be a family of formal power series, and define the Hankel-matrix H_Ψ of Ψ as the matrix $H_\Psi \in \mathbb{R}^{(X^* \times I) \times (X^* \times J)}$, where $I = \{1, 2, \ldots, p\}$ and $(H_\Psi)_{(u,i)(v,j)} = (S_j(vu))_i$. That is, the rows of H_Ψ are indexed by pairs (u, i) where u is a word over X and i is and integer in the range $1, \ldots, p$. The columns of H_Ψ are indexed by pairs (v, j) where v is a word over X and j is an element of the index set J. 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 [3, 15, 1].

Theorem 1 (Realization of formal power series) *Let $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ 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.
(iv) If the rank of the Hankel matrix H_Ψ is finite, i.e. $n = \mathrm{rank}\, H_\Psi < +\infty$, then one can construct a representation $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in X}, B, C)$ of Ψ using the columns of H_Ψ (see [1] for details).

2.2 Realization Algorithm

In this subsection, we present an algorithm for computing a minimal representation of a family of formal power series Ψ from finite data, more precisely, from a finite left-upper

block of the infinite Hankel matrix H_Ψ. The theorem guaranteeing the correctness of the algorithm, Theorem 2, will also enable us to define a distance between families of rational formal power series.

Let $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg | \ j \in J\}$ be a family of formal power series indexed by a *finite* set J. Let $H_{\Psi,N,M} \in \mathbb{R}^{I_M \times J_N}$ be a finite upper-left block of the infinite Hankel matrix H_Ψ obtained by taking all columns of H_Ψ indexed by words over X of length at most N, and all the rows of H_Ψ indexed by words of length at most M. More specifically, $H_{\Psi,N,M} \in \mathbb{R}^{I_M \times J_N}$ is the matrix whose rows are indexed by elements of the set $I_M = \{(u,i) \mid u \in X^*, |u| \leq M, i = 1,\ldots,p\}$, whose columns are indexed by elements of the set $J_N = \{(v,j) \mid j \in J, v \in X^*, |v| \leq N\}$, and whose entries are defined by $(H_{\Psi,N,M})_{(u,i),(v,j)} = (S_j(vu))_i$.

The following algorithm computes a representation of Ψ from $H_{\Psi,N+1,N}$.

Algorithm 1 [16] $(\mathbb{R}^r, \{A_\sigma\}_{\sigma \in X}, B, C) = $ *ComputeRepresentation*$(H_{\Psi,N+1,N})$

1: Let $r = \operatorname{rank} H_{\Psi,N,N}$ and choose $j_1,\ldots,j_r \in J$, $i_1,\ldots,i_r \in \{1,\ldots,p\}$, $v_1,\ldots,v_r, u_1,\ldots,u_r \in X^*$ such that for all $l = 1,\ldots,r$, $|v_l| \leq N$ and $|u_l| \leq N$, and the minor $T = ((S_{j_k}(v_k u_l))_{i_l})_{l,k=1,\ldots,r} \in \mathbb{R}^{r \times r}$ of $H_{\Psi,N,N}$ is of rank r.

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

$$A_\sigma T = Z_\sigma \text{ where } Z_\sigma = ((S_{j_k}(v_k \sigma u_l))_{i_l})_{l,k=1,\ldots,r}.$$

Let $B = \{B_j \mid j \in J\}$, where for each index $j \in J$, the vector $B_j \in \mathbb{R}^r$ is given by

$$B_j = T^{-1}((S_j(u_1)_{i_1}), (S_j(u_2))_{i_2}, \ldots (S_j(u_r))_{i_r})^T.$$

Let $C \in \mathbb{R}^{p \times r}$ be given by $C = \begin{bmatrix} C_1 & \cdots & C_r \end{bmatrix}$, where $C_l = S_{j_l}(v_l)$, for $l = 1,\ldots,r$.

Theorem 2 ([1, 3, 16]) *If* $\operatorname{rank} H_{\Psi,N,N} = \operatorname{rank} H_\Psi$, *then the representation* $\widetilde{R}_N$ *of* Ψ *returned by ComputeRepresentation is minimal. Furthermore, if* $\operatorname{rank} H_\Psi \leq N$, *or, equivalently, there exists a representation* R *of* Ψ, *such that* $\dim R \leq N$, *then* $\operatorname{rank} H_\Psi = \operatorname{rank} H_{\Psi,N,N}$, *hence* $\widetilde{R}_N$ *is a minimal representation of* Ψ.

From a computational point of view, algorithm *ComputeRepresentation* may not be the best way to compute a representation of Ψ. However, we have chosen to present it, because it makes theoretical reasoning easier. The algorithm is essentially a reformulation of the construction presented in [16]. An alternative algorithm, which uses the factorization of the finite Hankel-matrix $H_{\Psi,N,N+1}$ can be found in [1].

2.3 A Notion of Stability for Formal Power Series

Since our ultimate goal is to compare formal power series, we might want to restrict our attention to formal power series that are stable in some sense. 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.

Consider a formal power series $S \in \mathbb{R}^p \ll X^* \gg$, and denote by $|| \cdot ||_2$ the Euclidean norm in $\mathbb{R}^p$. Consider the following sequence,

$$L_n = \sum_{k=0}^n \sum_{\sigma_1 \in X} \cdots \sum_{\sigma_k \in X} ||S(\sigma_1 \sigma_2 \cdots \sigma_k)||_2^2. \qquad (3)$$

The series S will be called *square summable*, if the limit $\lim_{n \to +\infty} L_n$ exists and is finite. We will call the family $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg | j \in J\}$ *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 $R = (\mathbb{R}^n, \{A_\sigma\}_{\sigma \in X}, B, C)$ be an arbitrary representation of $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg | j \in J\}$. Assume that $X = \{\sigma_1, \ldots, \sigma_d\}$, where d is the number of elements of X, and consider the matrix

$$\widetilde{A}_R = \begin{bmatrix} A_{\sigma_1} \otimes A_{\sigma_1} & A_{\sigma_2} \otimes A_{\sigma_2} & \cdots & A_{\sigma_d} \otimes A_{\sigma_d} \\ A_{\sigma_1} \otimes A_{\sigma_1} & A_{\sigma_2} \otimes A_{\sigma_2} & \cdots & A_{\sigma_d} \otimes A_{\sigma_d} \\ \vdots & \vdots & \vdots & \vdots \\ A_{\sigma_1} \otimes A_{\sigma_1} & A_{\sigma_2} \otimes A_{\sigma_2} & \cdots & A_{\sigma_d} \otimes A_{\sigma_d} \end{bmatrix} \in \mathbb{R}^{n^2 d \times n^2 d}, \qquad (4)$$

where $\otimes$ denotes the Kronecker product. We will call R *stable*, if the matrix $\widetilde{A}_R$ is stable, i.e. if all its eigenvalues λ lie inside the unit disk ($|\lambda| < 1$). We have the following.

Theorem 3 *A rational family of formal power series is square summable if and only if all minimal representations are stable.*

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

3   Distances for Rational Power Series

The goal of this section is to present a notion of distance for families of rational formal power series, or equivalently, a distance between their minimal representations. The choice of a distance is by no means unique, in fact, we will suggest several different distances. The common feature of all these distances is that they all can be computed either from a minimal representation of the family, or from a big enough but finite set of values of the formal power series constituting the families.

Through the section, we will fix the space of coefficients $\mathbb{R}^p$ and the alphabet X. Also, we will fix a *finite* index set J and consider the space of all rational families of formal power series indexed by J, i.e. $\mathcal{P}_J = \{\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg | j \in J\} \mid \Psi \text{ is rational}\}$. Define the subset $\mathcal{P}_{J,N} = \{\Psi \in \mathcal{P}_J \mid \text{rank } H_\Psi \leq N\}$ of all rational families of formal power series whose Hankel-matrix is of rank at most N. Then it is easy to see that $\mathcal{P}_{J,N} \subseteq \mathcal{P}_{J,K}$ for all $N \leq K$, and $\mathcal{P}_J = \bigcup_{N=0}^{+\infty} \mathcal{P}_{J,N}$.

3.1 Distances Based on Truncation

We will first consider distances based on truncation, that is distances that compare finitely many values of formal power series.

Fix a natural number N, and denote by $m = card(J)$ the cardinality of J. Since J is finite, without loss of generality, we can assume that $J = \{1, \ldots, m\}$. Assume that the alphabet X is of the form $X = \{z_1, \ldots, z_d\}$ where d is the number of elements of X. For each $N \geq 1$, let $F_{N,J} : \mathbb{R}^{mpN} \times \mathbb{R}^{mpN} \to \mathbb{R}$ be a distance on $\mathbb{R}^{mpN}$ and denote by $F = \{F_{N,J} \mid N \in \mathbb{N}\}$ the family of distances.

We now define a pseudo-metric on $\mathcal{P}_J$ using the family of distances F. The main idea is the following. If $S \in \mathbb{R}^p \ll X^* \gg$ is a formal power series, then it can be viewed as a map $S : X^* \to \mathbb{R}^p$ on words over X. There are $M(N) = \sum_{j=0}^{2N+1} d^j$ words of length at most $2N + 1$ over the alphabet X, if X has d elements. Hence, we can view the restriction of the map S to the set of all words of length at most $2N + 1$ as a vector in $\mathbb{R}^{M(N)p}$. We can then define the distance $d_{F,N,J}(\Psi_1, \Psi_2)$ between two families of formal power series indexed by J, Ψ_1 and Ψ_2, as the distance $F_{M(N),J}(\phi_1, \phi_2)$ between the vectors ϕ_1 and ϕ_2 in $\mathbb{R}^{mpM(N)}$ representing the restriction of the elements of Ψ_1 and Ψ_2, respectively, to the set of words of length at most $2N + 1$. More formally,

1. Define an enumeration of all the words over the alphabet X as the bijective map $\psi : X^* \to \mathbb{N}$ defined as follows. For the empty word ϵ, let $\psi(\epsilon) = 0$ and for each letter z_i, $i = 1, \ldots, d$, let $\psi(z_i) = i$. Then, for each word of the form $w = vz_j$, $j = 1, \ldots, d$, $v \in X^*$ define $\psi(w)$ recursively as $\psi(w) = d \cdot \psi(v) + j$.

2. Denote by $X^{\leq 2N+1} = \{w \in X^* \mid |w| \leq 2N + 1\}$ the set of all words on X of length at most $2N + 1$. Notice that the restriction of ψ to the set $X^{\leq 2N+1}$ yields a bijective map with the range $[0, M(N) - 1]$.

3. For each $S \in \mathbb{R}^p \ll X^* \gg$ define $\pi_N(S)$ as the vector $(Z_0^T, Z_1^T, \ldots, Z_{M(N)-1}^T)^T$ in $\mathbb{R}^{pM(N)}$, where $Z_i = S(\psi^{-1}(i)) \in \mathbb{R}^p$. Since the integer i goes through all the values $[0, M(N) - 1]$, $\psi^{-1}(i)$ goes through all possible words of length at most $2N + 1$. Hence $\pi_N(S)$ is just the vector of all values $S(w)$ where $|w| \leq 2N + 1$.

4. For each rational family of formal power series $\Psi = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ define the vector $\pi_{J,N}(\Psi) = (\pi_N(S_1)^T, \ldots, \pi_N(S_m)^T)^T \in \mathbb{R}^{mpM(N)}$. That is, $\pi_{J,N}(\Psi)$ is obtained by stacking up the vectors $\pi_N(S_1), \ldots, \pi_N(S_m)$ representing the values of $S_1, \ldots, S_m$ on words of length at most $2N + 1$.

5. For each $\Psi_1, \Psi_2 \in \mathcal{P}_J$, define the functions $d_{F,N,J} : \mathcal{P}_J \times \mathcal{P}_J \to \mathbb{R}$ by

$$d_{F,N,J}(\Psi_1, \Psi_2) = F_{M(N),J}(\pi_{J,N}(\Psi_1), \pi_{J,N}(\Psi_2)) \tag{5}$$

We then have the following result.

Lemma 1 (Properties of $d_{F,N,J}$) $d_{F,N,J}$ *is a pseudo-distance in $\mathcal{P}_J$. That is, for each* $\Psi_1, \Psi_2, \Psi_3 \in \mathcal{P}_J$

$$d_{F,N,J}(\Psi_1, \Psi_2) = d_{F,N,J}(\Psi_2, \Psi_1) \geq 0,$$
$$d_{F,N,J}(\Psi_1, \Psi_2) \leq d_{F,N,J}(\Psi_1, \Psi_3) + d_{F,N,J}(\Psi_2, \Psi_3) \quad and \tag{6}$$
$$\Psi_1 = \Psi_2 \implies d_{F,N,J}(\Psi_1, \Psi_2) = 0.$$

The following theorem formulates an important property of $d_{F,N,J}$ and it relies on the partial realization result of Theorem 2.

Theorem 4 (Distance for Rational Formal Power Series) *The restriction of $d_{F,N,J}$ to $\mathcal{P}_{N,J}$ is a distance. That is, in addition to the properties listed in Lemma 1, the following holds.*

$$\forall\, \Psi_1,\, \Psi_2 \in \mathcal{P}_{N,J}: \quad \Psi_1 = \Psi_2 \iff d_{F,N,J}(\Psi_1, \Psi_2) = 0 \tag{7}$$

Proof. If Ψ_1 and Ψ_2 belong to $\mathcal{P}_{N,J}$, then by Theorem 2 rank $H_{\Psi_i,N,N}$ = rank H_{Ψ_i} holds for $i = 1, 2$. It is easy to see that $d_{F,N,J}(\Psi_1, \Psi_2) = 0$ if and only if $\pi_{N,J}(\Psi_1) = \pi_{N,J}(\Psi_2)$, i.e. the values of the elements of Ψ_1 and Ψ_2 coincide for all the words of length at most $2N+1$. Hence, $H_{\Psi_1,N+1,N} = H_{\Psi_2,N+1,N}$. Therefore, by Theorem 2, the representation $\widetilde{R}_N$ produced by **Algorithm 1** with the input $H_{\Psi_1,N+1,N}$ is a minimal representation of both Ψ_1 and Ψ_2, which implies that $\Psi_1 = \Psi_2$.

3.2 The Hilbert Space of Square Summable Families of Formal Power Series

In what follows we will define a scalar product on the space of square summable rational families of formal power series. With this scalar product the space of square summable rational families becomes a Hilbert space, and the corresponding distance will take all values of the formal power series into account.

Consider the set $\mathcal{P}_{s,J} = \{\Psi \in \mathcal{P}_J \mid \Psi$ is square summable $\}$ of *square summable rational families of formal power series*. Assume that J is finite. It is clear that $\mathcal{P}_{s,J}$ is a vector space, if we define addition and multiplication by a scalar as follows. Let $\Psi_1 = \{S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ and $\Psi_2 = \{T_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$ be two square summable rational families of formal power series. Then for each $\alpha, \beta \in \mathbb{R}$, let $\alpha\Psi_1 + \beta\Psi_2 = \{\alpha T_j + \beta S_j \in \mathbb{R}^p \ll X^* \gg \mid j \in J\}$, where $\alpha S_j + \beta T_j$ is just the usual point-wise linear combination on formal power series ([14]), i.e. for all $w \in X^*$, $(\alpha S_j + \beta T_j)(w) = \alpha S_j(w) + \beta T_j(w)$. Now, for Ψ_1 and Ψ_2 defined as before, define the bilinear map $< \cdot, \cdot >_J \colon \mathcal{P}_{s,J} \times \mathcal{P}_{s,J} \to \mathbb{R}$ as

$$< \Psi_1, \Psi_2 >_J \doteq < \cdot, \cdot >_J (\Psi_1, \Psi_2) = \sum_{j \in J} \sum_{w \in X^*} S_j(w)^T T_j(w) \tag{8}$$

Since J is finite and S_j, T_j are both square summable, the infinite sum in (8) is well defined and finite. The following lemma characterizes some properties of $< \cdot, \cdot >_J$.

Lemma 2 *The map $< \cdot, \cdot >_J$ is a scalar product and the space $\mathcal{P}_{s,J}$ with the scalar product $< \cdot, \cdot >_J$ is a Hilbert space.*

As a consequence, we can view $\mathcal{P}_{s,J}$ as a normed space with the norm $||.||_J$ induced by $< \cdot, \cdot >_J$, i.e.

$$||\Psi||_J \doteq \sqrt{\sum_{j \in J} \sum_{w \in X^*} ||S_j(w)||_2^2} = \sqrt{< \Psi, \Psi >_J}. \tag{9}$$

The following theorem gives a formula for computing $< \Psi_1, \Psi_2 >_J$ for all $\Psi_1, \Psi_2 \in \mathcal{P}_{s,J}$, provided that the representations of Ψ_1 and Ψ_2 are available.

Theorem 5 *For $i = 1, 2$, assume that $R_i = (\mathbb{R}^{n_i}, \{A_{i,\sigma}\}_{\sigma \in X}, C_i, B_i)$ is a stable representation of $\Psi_i \in \mathcal{P}_{s,J}$ and that $B_i = \{B_{i,j} \in \mathbb{R}^{n_i} \mid j \in J\}$. Then there exists a unique solution $P \in \mathbb{R}^{n_1 \times n_2}$ to the Sylvester equation*

$$P = \sum_{\sigma \in X} A_{1,\sigma}^T P A_{2,\sigma} + C_1^T C_2 \tag{10}$$

and the scalar product $< \Psi_1, \Psi_2 >_J$ can be written explicitly

$$< \Psi_1, \Psi_2 >_J = \sum_{j \in J} B_{1,j}^T P B_{2,j}. \tag{11}$$

Notice from Theorem 3 that if R_1 and R_2 are minimal representations of Ψ_1 and Ψ_2 in $\mathcal{P}_{s,J}$, respectively, then the condition of Theorem 5 holds. Hence we can use any minimal representation to compute $< \Psi_1, \Psi_2 >_J$. From this we may compute the distance between Ψ_1 and Ψ_2 as $||\Psi_1 - \Psi_2||_J^2 = < \Psi_1, \Psi_1 >_J - 2 < \Psi_1, \Psi_2 >_J + < \Psi_2, \Psi_2 >_J$.

4 Rational Power Series and Input-Output Behavior of Dynamical Systems

The main motivation for introducing the framework of rational formal power series is that it provides a common algebraic framework for realization theory and system identification of a wide-variety of input/output systems. The classes of systems whose behaviors can be described in terms of rational formal power series include linear systems [17, 9], bilinear systems, [15, 3, 2, 18], multidimensional systems [4], finite state hidden Markov models [5], continuous-time linear and bilinear switched systems [1] and continuous-time linear and bilinear hybrid systems [1].

Hence, if we pick two dynamical systems Σ_1 and Σ_2 from any of the classes mentioned above we can compare them as follows. We can construct the families of formal power series Ψ_1 and Ψ_2 corresponding to the input-output behaviors of Σ_1 and Σ_2, respectively. Then, we can compare Ψ_1 and Ψ_2 using one of the distances defined in Section 3. Note that the choice of the families Ψ_i, $i = 1, 2$ is unique if Σ_i belongs to one of the classes of systems described in the previous paragraph. Alternatively, we can construct the rational representations R_{Σ_1} and R_{Σ_2} of the families Ψ_1 and Ψ_2 respectively. In general, the representations R_{Σ_1} and R_{Σ_2} can be easily computed from the parameters of Σ_1 and Σ_2. Then, we can use R_{Σ_1} and R_{Σ_2} to compute the distance between Ψ_1 and Ψ_2. This approach is particularly appealing if Ψ_1 and Ψ_2 are square summable and one wants to use the norm (9). Notice, that even if Ψ_1 and Ψ_2 are square summable, R_{Σ_1} or R_{Σ_2} might fail to be stable. In this case we have to minimize R_{Σ_1} and R_{Σ_2} first, and use the resulting stable minimal representations (see Theorem 3 and Theorem 5) for computing the distance. Algorithms for minimizing representations can be found in [1]. Notice also that, in general, there need not be any connection between square summability of Ψ_i, $i = 1, 2$ and stability of the dynamical systems Σ_i, $i = 1, 2$.

In what follows, we will demonstrate the use of rational formal power series for stochastic discrete-time jump-linear systems. This class of hybrid systems has a wide variety of applications including computer vision. To the best of our knowledge, the relationship between stochastic jump-linear systems and formal power series presented here is new, though there are some similarities between our approach and that in [18].

4.1 Stochastic Jump-Linear Systems

The terminology and notation used in this section is based on the conventions adopted in literature, see [19, 9]. A *stochastic jump-linear systems* [19] is a discrete-time stochastic system described by the equations

$$\Sigma : \begin{cases} \mathbf{x}(k+1) = A_{\boldsymbol{\theta}(k)}\mathbf{x}(k) + B_{\boldsymbol{\theta}(k)}\mathbf{v}(k) \\ \quad \mathbf{y}(k) = C_{\boldsymbol{\theta}(k)}\mathbf{x}(k) \text{ and } \mathbf{o}(k) = \lambda(\boldsymbol{\theta}(k)) \end{cases}. \tag{12}$$

Here, $\mathbf{x}$, $\boldsymbol{\theta}$, $\mathbf{y}$, $\mathbf{o}$ and $\mathbf{v}$ are stochastic processes of the following form. The process $\mathbf{x}$ is called the *continuous state process* and takes values in the *continuous-state space* $\mathbb{R}^n$. The process $\boldsymbol{\theta}$ is called the *discrete state process* and takes values in the *set of discrete states* $Q = \{1, 2, \ldots, d\}$. The process $\mathbf{y}$ is the *continuous output process* and takes values in the *set of continuous outputs* $\mathbb{R}^p$. The process $\mathbf{o}$ is the *discrete output process* and takes values in the *set of discrete outputs* $O = \{1, 2, \ldots, l\}$. Finally, the process $\mathbf{v}$ is the *continuous noise* and takes values in $\mathbb{R}^m$. The matrices A_q, B_q, C_q, $q \in Q$, are of the form $A_q \in \mathbb{R}^{n \times n}$, $B_q \in \mathbb{R}^{n \times m}$, and $C_q \in \mathbb{R}^{p \times n}$. The map $\lambda : Q \to O$ is called the *readout map* and it assigns a discrete output to each discrete state. We will assume that $E[\mathbf{v}(k)\mathbf{v}(l)] = \delta_{k,l}I$ and $E[\mathbf{v}(k)] = 0$, for all $k, l \in \mathbb{N}$, that is $\mathbf{v}$ is a zero mean process and $\mathbf{v}(k)$, $k \in \mathbb{N}$ are uncorrelated. Furthermore, we will assume that for each $k, l \in \mathbb{N}$, $\mathbf{x}(0), \mathbf{v}(k), \boldsymbol{\theta}(l)$ are mutually independent random variables. We will also assume that the state-transition of the Markov process $\boldsymbol{\theta}$ is governed by the transition probabilities $p_{q_1,q_2}, q_1, q_2 \in Q$, where p_{q_1,q_2} is the probability that $\boldsymbol{\theta}$ changes its value from q_2 to q_1, i.e. $p_{q_1,q_2} = \text{Prob}(\boldsymbol{\theta}_{k+1} = q_1 \mid \boldsymbol{\theta}_k = q_2)$. In addition, we will assume that the initial probability distribution of $\boldsymbol{\theta}$ is given by the vector $\pi = (\pi_1, \ldots, \pi_d)^T \in \mathbb{R}^d$, where $\pi_q = \text{Prob}(\boldsymbol{\theta}(0) = q)$ denotes the probability that the process $\boldsymbol{\theta}$ is in state q at time 0. The evolution of system (12) is as follows. At each time instant k, the continuous state $\mathbf{x}$ and the continuous output $\mathbf{y}$ change according to the discrete-time stochastic linear system $(A_{\boldsymbol{\theta}(k)}, B_{\boldsymbol{\theta}(k)}, C_{\boldsymbol{\theta}(k)})$. The discrete state process $\boldsymbol{\theta}$, together with the discrete output process $\mathbf{o}$, form a finite state hidden Markov model [5].

In the next subsection, we study the concept of realization for stochastic jump-linear systems. To that end, we will assume that the stochastic processes $\mathbf{x}$ and $\mathbf{y}$ are wide-sense stationary and zero mean, which is guaranteed under the following assumptions.

Assumption 1 *The Markov process $\boldsymbol{\theta}$ is stationary and ergodic, hence for all $q \in Q$, $\sum_{s \in Q} p_{q,s}\pi_s = \pi_q$.*

Assumption 2 *There exists $n \times n$ matrices P_q, $q \in Q$, such that for each $q \in Q$*

$$P_q = \sum_{s \in Q} p_{q,s} A_s P_s A_s^T + B_s B_s^T p_{q,s}\pi_s, \tag{13}$$

$E[\mathbf{x}(0)] = 0$, *and* $E[\mathbf{x}(0)\mathbf{x}(0)^T \chi(\boldsymbol{\theta}(0) = q)] = P_q$, *where* χ *denotes the indicator function, i.e.* $\chi(A) = 1$ *if the event A is true, and* $\chi(A) = 0$ *otherwise.*

These assumptions are not particularly strong. For instance, under suitable conditions [19], there is a unique collection of positive semi-definite matrices P_q such that (13) holds.

4.2 Realization of Stochastic Jump-Linear Systems

Recall the notion of *weak realization* for linear stochastic systems [9]. In this subsection, we will formulate a similar concept for stochastic jump-linear systems.

Consider a stationary process $\widetilde{\mathbf{o}}$ taking values in the finite output space O, and a wide-sense stationary, zero-mean stochastic process $\widetilde{\mathbf{y}}$ taking values in the continuous output space $\mathbb{R}^p$. let O^+ be the set of all nonempty words in O, i.e. $O^+ = O^* \setminus \{\epsilon\}$. For all $o_0, o_1, \ldots, o_k \in O$, $k \geq 0$, define the maps $\mathcal{P}_{\widetilde{o}} : O^+ \to \mathbb{R}$ and $\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}} : O^+ \to \mathbb{R}^{p \times p}$

$$\mathcal{P}_{\widetilde{o}}(o_0 o_1 \cdots o_k) = \mathrm{Prob}(\widetilde{\mathbf{o}}(i) = o_i, i = 0, \ldots, k)$$
$$\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}(o_0 o_1 \cdots o_k) = E[\widetilde{\mathbf{y}}(k)\widetilde{\mathbf{y}}(0)^T \chi(\widetilde{\mathbf{o}}(i) = o_i, i = 0, \ldots, k)]. \tag{14}$$

Notice that the map $\mathcal{P}_{\widetilde{o}}$ gives the probability distribution of the stochastic process $\widetilde{\mathbf{o}}$, while the map $\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ gives the covariance of $\widetilde{\mathbf{y}}(k)$ and $\widetilde{\mathbf{y}}(0)$, provided that the proces $\widetilde{\mathbf{o}}$ takes values $o_0, \ldots, o_k$ in the first $k+1$ time instances. That is, $\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ collects information on the second-order moments of $\widetilde{\mathbf{y}}$.[1]

Consider now a jump-linear system Σ of the form (12) and recall the definition of the processes $\mathbf{y}$ and $\mathbf{o}$. If **Assumption 1** and **Assumption 2** hold, then $\mathbf{y}$ is wide-sense stationary and zero-mean and $\mathbf{o}$ is stationary. Hence, $\mathcal{C}_{\mathbf{o},\mathbf{y}}$ and $\mathcal{P}_{\mathbf{o}}$ are well-defined. In fact, they depend only on the matrices A_q, B_q, C_q, $q \in Q$, the discrete state-transition probabilities p_{q_1,q_2}, $q_1, q_2 \in Q$, the probability distribution of the initial discrete-state π, and the readout map λ. To emphasize that $\mathcal{C}_{\mathbf{o},\mathbf{y}}$ and $\mathcal{P}_{\mathbf{o}}$ depend only on the parameters of Σ, we will denote $\mathcal{C}_{\mathbf{o},\mathbf{y}}$ by $\mathcal{C}_\Sigma$ and $\mathcal{P}_{\mathbf{o}}$ by $\mathcal{P}_\Sigma$. These maps are important, because they contain information about the probability distribution of the output processes generated by Σ. In fact, the following is true.

Proposition 1 *If $\mathbf{x}(0)$ and $\mathbf{v}$ are Gaussian, $Q = O$, and $\lambda = id$, i.e. the discrete state is fully observed, then the map $\mathcal{C}_\Sigma$ uniquely determines the distribution of $\mathbf{y}$.*

The assumption that $Q = O$ is critical here. Intuitively, the more information about the discrete state is preserved by the discrete output, i.e. the closer $\mathbf{o}$ is to $\boldsymbol{\theta}$, the better estimate of the probability distribution of $\mathbf{y}$ is provided by $\mathcal{C}_\Sigma$.

With the notation above, we are now ready to define a notion of weak realization for stochastic jump-linear systems. Let $\widetilde{\mathbf{y}}$ be a wide-sense stationary, zero mean $\mathbb{R}^p$-valued process, and let $\widetilde{\mathbf{o}}$ be a stationary O-valued process. A stochastic-jump linear system Σ is said to be a *weak stochastic realization of* $(\widetilde{\mathbf{y}}, \widetilde{\mathbf{o}})$, if $\mathcal{P}_{\widetilde{o}} = \mathcal{P}_\Sigma$ and $\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}} = \mathcal{C}_\Sigma$.

Clearly, the fact that Σ is a weak-realization of the processes $(\widetilde{\mathbf{y}}, \widetilde{\mathbf{o}})$ imposes some constraints on the probability distribution of the processes $\widetilde{\mathbf{o}}$ and $\widetilde{\mathbf{y}}$. We will now show that $(\widetilde{\mathbf{y}}, \widetilde{\mathbf{o}})$ has a weak realization by a stochastic jump-linear system, only if certain families of formal power series are rational. We will construct two families of formal power series $\Psi_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ and $S_{\widetilde{\mathbf{o}}}$ based on the maps $\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ and $\mathcal{P}_{\widetilde{o}}$, respectively, as follows.

Let $X = O = \{1, 2, \ldots, l\}$ and $J_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}} = \{1, \ldots, p\} \times O$ be, respectively, the alphabet and the index set over which the formal power series will be defined. For each integer $i = 1, \ldots, p$, letter $o \in O$, and word $w \in O^*$, let $\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}},(i,o)}(w) \in \mathbb{R}^{lp}$ be the ith column of the matrix

$$\left[\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}^T(ow1), \mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}^T(ow2), \cdots, \mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}^T(owl)\right]^T \in \mathbb{R}^{lp \times p}. \tag{15}$$

---

[1] Notice the similarity between $\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ and the generalized covariances in [18].

We define the family of formal power series $\Psi_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ associated with $\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ as

$$\Psi_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}} = \{\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}},(i,o)} \in \mathbb{R}^{pl} \ll O^* \gg | \ (i,o) \in J_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}\} \tag{16}$$

The construction of $S_{\widetilde{\mathbf{o}}}$ is much simpler. We can simply identify the map $\mathcal{P}_{\widetilde{\mathbf{o}}}$ with a formal power series $S_{\widetilde{\mathbf{o}}} \in \mathbb{R} \ll O^* \gg$ by defining $S_{\widetilde{\mathbf{o}}}(\epsilon) = 1$ for the empty word and $S_{\widetilde{\mathbf{o}}}(w) = \mathcal{P}_{\widetilde{\mathbf{o}}}(w)$ for all $w \in O^+$. By abuse of notation we will denote $S_{\widetilde{\mathbf{o}}}$ simply by $\mathcal{P}_{\widetilde{\mathbf{o}}}$. We will call $(\Psi_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}, \mathcal{P}_{\widetilde{\mathbf{o}}})$ *the pair of formal power series associated with* $(\widetilde{\mathbf{y}}, \widetilde{\mathbf{o}})$.

The next step is to define a *pair of representations* $(R_{\Sigma,C}, R_{\Sigma,D})$ *associated with a jump-linear system Σ of the form (12)*.

We define the representation $R_{\Sigma,C}$ as follows. For each $o \in O$ and $q, q_1, q_2 \in Q$, let $C_q^o = C_q \chi(\lambda(q) = o) \in \mathbb{R}^{p \times n}$, $A_{q_1,q_2}^o = p_{q_1,q_2} \chi(\lambda(q_2) = o) A_{q_2} \in \mathbb{R}^{n \times n}$, and $B_q^o = P_q(C_q^o)^T \in \mathbb{R}^{n \times p}$, where $P_q \in \mathbb{R}^{n \times n}$ is the matrix defined in (13). Using this notation, define the matrices $\widetilde{A}_o \in \mathbb{R}^{nd \times nd}$, $\widetilde{C} \in \mathbb{R}^{lp \times nd}$ and $\widetilde{B}_o \in \mathbb{R}^{nd \times p}$, $o \in O$, as

$$\widetilde{A}_o = \begin{bmatrix} A_{1,1}^o & A_{1,2}^o & \cdots & A_{1,d}^o \\ A_{2,1}^o & A_{2,2}^o & \cdots & A_{2,d}^o \\ \vdots & \vdots & \vdots & \vdots \\ A_{d,1}^o & A_{d,2}^o & \cdots & A_{d,d}^o \end{bmatrix}, \quad \widetilde{C} = \begin{bmatrix} C_1^1 & C_2^1 & \cdots & C_d^1 \\ C_1^2 & C_2^2 & \cdots & C_d^2 \\ \vdots & \vdots & \vdots & \vdots \\ C_1^l & C_2^l & \cdots & C_d^l \end{bmatrix}, \text{ and } \widetilde{B}_o = \begin{bmatrix} B_1^o \\ B_2^o \\ \vdots \\ B_d^o \end{bmatrix}.$$

Then, for each $o \in O$, and $i = 1, \ldots, p$ let $\widetilde{B}_{(i,o)} \in \mathbb{R}^{nd}$ be the ith column of $\widetilde{B}_o$ and define the set $\widetilde{B} = \{\widetilde{B}_{(i,o)} \in \mathbb{R}^{nd} \mid (i,o) \in J_{o,y}\}$ indexed by $J_{o,y} = \{1, \ldots, p\} \times O$. We define the representation $R_{\Sigma,C}$ as

$$R_{\Sigma,C} = (\mathbb{R}^{nd}, \{\widetilde{A}_o\}_{o \in O}, \widetilde{B}, \widetilde{C}). \tag{17}$$

As per the representation $R_{\Sigma,D}$, we define it as

$$R_{\Sigma,D} = (\mathbb{R}^d, \{M_o\}_{o \in O}, \{\pi\}, e), \tag{18}$$

where $e = (1, 1, \ldots, 1) \in \mathbb{R}^{1 \times d}$, and for each $o \in O$ and $q_1, q_2 \in Q$, the (q_1, q_2) entry of the matrix $M_o \in \mathbb{R}^{d \times d}$ is defined from the transition probabilities of the process $\boldsymbol{\theta}$ as $p_{q_1,q_2} \chi(\lambda(q_2) = o)$. Notice the similarity between the definition of $R_{\Sigma,D}$ and the definition of a quasi-realization for the finite state hidden Markov model formed by $(\boldsymbol{\theta}, \mathbf{o})$ given in [5].

We these definitions, we have the following result.

Theorem 6 (Weak Realization) *A jump-linear system Σ of the form (12) is a weak realization of $(\widetilde{\mathbf{y}}, \widetilde{\mathbf{o}})$ if and only if $R_{\Sigma,C}$ is a representation of $\Psi_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ and $R_{\Sigma,D}$ is a representation of $\mathcal{P}_{\widetilde{\mathbf{o}}}$. Hence, $(\widetilde{\mathbf{y}}, \widetilde{\mathbf{o}})$ admits a weak stochastic realization by a stochastic jump-linear system, only if $\Psi_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ is a rational family of formal power series and $\mathcal{P}_{\widetilde{\mathbf{o}}}$ is a rational formal power series.*

An important implication of the theorem above is the following. If we know that the processes $(\widetilde{\mathbf{y}}, \widetilde{\mathbf{o}})$ admit a weak stochastic realization by a stochastic jump-linear system, then we can find representations of $\Psi_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ and $\mathcal{P}_{\widetilde{\mathbf{o}}}$ from finite data. More precisely, if $\operatorname{rank} H_{\Psi_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}} \leq N$ and $\operatorname{rank} H_{\mathcal{P}_{\widetilde{\mathbf{o}}}} \leq N$, then from $\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}(o_0 \cdots o_k)$, $\mathcal{P}_{\widetilde{\mathbf{o}}}(o_0 \cdots o_k)$

$k \leq 2N + 1$, $o_0, \ldots, o_k \in O$, we can construct the Hankel matrices $H_{\Psi_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}, N+1, N}$ and $H_{\mathcal{P}_{\widetilde{\mathbf{o}}}, N+1, N}$ and compute a representation $R_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ of $\Psi_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}$ and $R_{\widetilde{\mathbf{o}}}$ of $\mathcal{P}_{\widetilde{\mathbf{o}}}$ respectively. Note that if we know that $(\widetilde{\mathbf{y}}, \widetilde{\mathbf{o}})$ has a weak stochastic realization by a jump-linear system Σ with a continuous state-space of dimension n and a discrete state-space of cardinality d, then we can take $N \geq nd > 0$. Finally, recall that the problem of estimating $\mathcal{C}_{\widetilde{\mathbf{o}},\widetilde{\mathbf{y}}}(o_0 \cdots o_k)$ and $\mathcal{P}_{\widetilde{\mathbf{o}}}(o_0 \cdots o_k)$ is a classical statistical problem. In particular, if the joint process $(\widetilde{\mathbf{y}}, \widetilde{\mathbf{o}})$ is ergodic, then these quantities can easily be estimated from a long enough sequence of measurements.

4.3 Distances between Stochastic Jump-Linear Systems

Imagine we would like to compare the probability distributions of the output processes $(\widetilde{\mathbf{o}}_1, \widetilde{\mathbf{y}}_1)$, and $(\widetilde{\mathbf{o}}_2, \widetilde{\mathbf{y}}_2)$ of two stochastic jump-linear systems Σ_1 and Σ_2, respectively. We can do that by using one of the distances defined in Section 3 to compare their associated pairs of families of formal power series: $\Psi_{\widetilde{\mathbf{o}}_1, \widetilde{\mathbf{y}}_1}$ with $\Psi_{\widetilde{\mathbf{o}}_2, \widetilde{\mathbf{y}}_2}$ and $\mathcal{P}_{\widetilde{\mathbf{o}}_1}$ with $\mathcal{P}_{\widetilde{\mathbf{o}}_2}$. When Σ_1 and Σ_2 are known, we can construct the representations $R_{\Sigma_i, C}$ and $R_{\Sigma_i, D}$, $i = 1, 2$. Then, we can use $R_{\Sigma_i, C}$, $i = 1, 2$ to compute the distance between $\Psi_{\widetilde{\mathbf{o}}_1, \widetilde{\mathbf{y}}_1}$ and $\Psi_{\widetilde{\mathbf{o}}_2, \widetilde{\mathbf{y}}_2}$. Likewise, we can use $R_{\Sigma_i, D}$, $i = 1, 2$ to compute the distance between $\mathcal{P}_{\widetilde{\mathbf{o}}_1}$ and $\mathcal{P}_{\widetilde{\mathbf{o}}_2}$. The advantage of using distances on formal power series is even more apparent if Σ_1 and Σ_2 are unknown, because the identification of stochastic jump-linear systems is poorly developed.[2] Instead, one could use the estimates of finitely many values of $\mathcal{C}_{\widetilde{\mathbf{o}}_i, \widetilde{\mathbf{y}}_i}$ and $\mathcal{P}_{\widetilde{\mathbf{o}}_i}$, $i = 1, 2$ to compute the minimal representations $R_{C,i}$ of $\Psi_{\widetilde{\mathbf{o}}_i, \widetilde{\mathbf{y}}_i}$, $i = 1, 2$ and $R_{D,i}$ of $\mathcal{P}_{\widetilde{\mathbf{o}}_i}$, $i = 1, 2$, and use the computed representations to compare the behavior of the two systems. The procedure for computing such representations from their Hankel matrices is known [1, 15, 16] and it is likely to be computationally less costly than identifying the original jump-linear systems.

5 Discussion and Conclusion

In this paper several definitions of distances for rational formal power series and rational representations were presented. It was argued that the results can be used to define metrics and topology on the space of a wide variety of dynamical systems. The key argument is that for many classes of dynamical systems there is a correspondence between the input-output behaviors of the systems and rational formal power series. In particular, this is the case for a number of hybrid systems and some nonlinear systems.

To the best of our knowledge, the problem of distances between hybrid systems had not been addressed so far. In the case of nonlinear systems, there are some results on the topological and geometric structure of the space of bilinear systems, see for example [16], where the algebraic variety structure of that space was described. In contrast, there is a fair amount of literature on distances between linear systems and on the topological and geometric structure of the space of linear systems. Note the relationship between input-output maps and output processes of linear systems and families of formal power

---

[2] Even in the linear case, the full identification procedure for linear stochastic systems is computationally costly.

series over the one letter alphabet $X = \{z\}$. Because of this correspondence, any distance on rational families of formal power series will give us a distance between linear systems. Spaces of equivalence classes of minimal linear systems were already studied before, for both the stochastic and deterministic settings. In fact, it was shown in [7] that, for each N, the set of all equivalence classes $M_N^{m,p}$ of minimal linear systems of dimension N with m inputs and p outputs forms both an analytic manifold and an algebraic variety and admits a natural topology. Here two minimal linear systems belong to the same equivalence class if they are algebraically similar. Denote by $M_N^{m,p,a}$ the set of equivalence classes of stable minimal linear systems. Then it was shown in [8, 6] that $M_N^{m,p,a}$ and $M_N^{m,p}$ are diffeomorphic as analytic manifolds and the topology of $M_N^{m,p,a}$ as an analytic manifold can be obtained by the metric induced by the H_2 norm. It is easy to see that the distance induced by the H_2 norm is a particular case of the distance induced by the norm (9), if we identify equivalence classes of minimal linear systems with equivalence classes of minimal rational representations (two minimal representations are equivalent, if they are isomorphic). More recent papers on distances between stochastic linear systems can be found in [20, 21, 12]. In particular, [12] introduces the trace distance between linear systems and gives a formula to compute it. Surprisingly, the distance induced by the norm (9) is closely related to the trace distance.

In practical situations, the families of formal power series are likely to encode the external behaviors of some dynamical systems. In such cases, the available information is either a rational representation of each family of formal power series, or a finite collection of values of the formal power series.

If we are given a rational representation R_Ψ of each family Ψ, then any of the distances described in Section 3 can easily be computed. Notice that if Ψ is square summable, then we can minimize R_Ψ and the obtained minimal representation of $R_{m,\Psi}$ will be stable, due to Theorem 3. Hence, using Theorem 5 we can compute the distance induced by $< \cdot, \cdot >_J$ by solving the corresponding Sylvester equation from Theorem 5. Hence, except for the computational complexity, there are few restrictions on using any of the distances. Therefore, one may choose different distances depending on the particular application and the computational costs. Note that the issue of computational complexity is still open for the distances we presented. However, the computational cost of computing distances of type $d_{F,N,J}$ between formal power series is exponential in N, if the underlying alphabet X has more than one element.

If only the finite collection of values is available then the task of choosing the right distance is more complex. Assume that we know the values of the elements of the families for all words of length at most N. Then there are two cases to be considered. If $N \geq 2M + 1$ and $M = \operatorname{rank} H_\Psi$ for all families Ψ involved, then we may apply the algorithm described in Theorem 2 to compute a minimal representation R_Ψ of Ψ. If Ψ is a square summable family, then the resulting representation R_Ψ is stable. Even in this ideal case when N is big enough several issues require attention. First of all, computing R_Ψ might be computationally expensive. If we want to compute one of the distances $d_{F,M,J}$, $M \leq N$, then we might do better by using the data directly, rather than computing the representations first and then computing the distance from the representations. However, computing the representations R_Ψ might be a good idea if we want to use the distances induced by the norm (9). Moreover, if the data are noisy, we

do not know whether our algorithm will still produce a stable representation, which is a prerequisite for the existence of the solution of the Sylvester equation. If we cannot ensure that N is big enough, then the algorithm from Theorem 2 might fail to produce a representation of Ψ.

References

1. Petreczky, M.: Realization Theory of Hybrid Systems. PhD thesis, Vrije Universiteit, Amsterdam (2006) available online: `http://www.cis.jhu.edu/~mihaly`.
2. Isidori, A.: Nonlinear Control Systems. Springer Verlag (1989)
3. Sontag, E.D.: Realization theory of discrete-time nonlinear systems: Part I – the bounded case. IEEE Transaction on Circuits and Systems **26**(4) (1979)
4. Ball, J.A., Groenewald, G., Malakorn, T.: Structured noncommutative multidimensional linear systems. In: Proceedings Sixteenth International Symposium on Mathematical Theory of Networks and Systems. (2004)
5. Anderson, B.D.O.: The realization problem for hidden markov models. Math. Control Signals Systems **12** (1999) 80–120
6. Peeters, R.: System Identification Based on Riemannian Geometry: Theory and Algorithms. PhD thesis, Free University, Amsterdam (1994)
7. Hazewinkel, M.: Moduli and canonical forms for linear dynamical systems II: The topological case. Mathematical Systems Theory **10** (1977) 363–385
8. Hanzon, B.: On the differentiable manifold of fixed order stable linear systems. Systems and Control Letters **13** (1989) 345 – 352
9. Caines, P.: Linear Stochastic Systems. John Wiley and Sons, New-York (1988)
10. Bissacco, A., Chiuso, A., Ma, Y., Soatto, S.: Recognition of human gaits. In: IEEE Conference on Computer Vision and Pattern Recognition. Volume 2. (2001) 52–58
11. Doretto, G., Chiuso, A., Wu, Y., Soatto, S.: Dynamic textures. International Journal of Computer Vision **51**(2) (2003) 91–109
12. Vishwanathan, S., Smola, A., Vidal, R.: Binet-Cauchy kernels on dynamical systems and its application to the analysis of dynamic scenes. International Journal of Computer Vision (2006)
13. Bissacco, A., Chiuso, A., Soatto, S.: Classification and recognition of dynamical models. Technical Report CSD-TR 060020, UCLA (2006)
14. Berstel, J., Reutenauer, C.: Rational series and their languages. EATCS Monographs on Theoretical Computer Science, Springer-Verlag (1984)
15. Sontag, E.D.: Polynomial Response Maps. Volume 13 of Lecture Notes in Control and Information Sciences. Springer Verlag (1979)
16. Sontag, D.E.: A remark on bilinear systems and moduli spaces of instantons. Systems and Control Letters **9**(5) (1987) 361–367
17. Callier, M.F., Desoer, A.C.: Linear System Theory. Springer-Verlag (1991)
18. Desai, U.: Realization of bilinear stochastic systems. IEEE Transactions on Automatic Control **AC-31**(2) (1986)
19. Costa, O., Fragoso, M., Marques, R.: Discrete-Time Markov Jump Linear Systems. Springer-Verlag, London (2005)
20. Martin, A.: A metric for ARMA processes. IEEE Trans. on Signal Processing **48**(4) (2000) 1164–1170
21. Cock, K.D., Moor, B.D.: Subspace angles and distances between ARMA models. System and Control Letters **46**(4) (2002) 265–270