Proceedings of the 2006 American Control Conference
Minneapolis, Minnesota, USA, June 14-16, 2006
WeB06.1

On the Markov Property for Nonlinear Discrete-Time Systems with Markovian Inputs

Arturo Tejada, Oscar R. González and W. Steven Gray

Abstract— **The behavior of a general hybrid system in discrete-time can be represented by a non-linear difference equation $x(k+1) = F_k(x(k), \boldsymbol{\theta}(k))$, where $\boldsymbol{\theta}(k)$ is assumed to be a finite-state Markov chain. An important step in the stability analysis of these systems is to establish the Markov property of $(x(k), \boldsymbol{\theta}(k))$. There are, however, no complete proofs of this property which are simple to understand. This paper aims to correct this problem by presenting a complete and explicit proof, which uses only fundamental measure-theoretical concepts.**

I. INTRODUCTION

The last decade has witnessed a steady increase in the number of applications that combine logical decisions with continuous-time or discrete-time dynamics. These so-called hybrid systems have been used in several applications (cf. [1], [2]). Markovian jump linear systems (MJLS) are, arguably, the most studied sub-class of hybrid systems. Their simple dynamics, in discrete-time, are given by

$$\boldsymbol{x}(k+1) = A_{\boldsymbol{\theta}(k)}\boldsymbol{x}(k),$$

where $\boldsymbol{\theta}(k)$ represents a finite-state Markov chain. The stability of these systems has been studied employing the well known fact that their state, $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$, constitutes a Markov chain. Specifically, the long-term behavior of a Markov chain can be studied by analyzing its transition kernel (see [3] for a thorough discussion of the continuous-time case). The Markov property of $(\boldsymbol{x}(\cdot), \boldsymbol{\theta}(\cdot))$ has been stated without proof by many authors both in continuous-time (cf. [4], [5]) and in discrete-time (cf. [6]). There are some results available for the more general class of systems

$$\boldsymbol{x}(k+1) = F_k(\boldsymbol{x}(k), \boldsymbol{\theta}(k)), \qquad (1)$$

where F_k are measurable functions. However, they either address the simpler case when $\boldsymbol{\theta}(k)$ is an i.i.d. sequence [7], [8], use heuristic arguments [9], or employ sophisticated results [10]. To the best of our knowledge, there is no complete proof of the Markov nature of $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$ for system (1), when $\boldsymbol{\theta}(k)$ is a Markov chain.

This paper aims to correct this problem by presenting a complete and explicit proof which uses only fundamental measure-theoretical concepts and follows the probabilistic approach, i.e., it interprets the process $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$ as a sequence of random vectors and exploits its properties. The proof is used subsequently to establish the Markov property of complex hybrid systems like the ones introduced in [11], [12], enabling their stability analysis.

The authors are affiliated with the Department of Electrical and Computer Engineering, Old Dominion University, Norfolk, Virginia 23529-0246, USA. atejadar@odu.edu, {gonzalez,gray}@ece.odu.edu

The paper is organized as follows. Section II states and proves the main result. The proof relies on two supporting theorems, which are proven in Sections III and IV. Section V applies the main result to three examples: a Markov jump linear system, a Jump Linear System (JLS) driven by a Finite State Machine (FSM) and a Hybrid Jump Linear System (HJLS). Finally, Section VI provides the conclusions.

II. MAIN RESULT

A. Preliminaries

In the sequel, Ω and Φ denote arbitrary sets. Their elements are denoted, respectively, by ω and ϕ, and their various subsets by $\{\omega : \cdots\}$ and $\{\phi : \cdots\}$. The proofs that follow make frequent use of measurable and Borel functions. Recall that a function $f : \Omega \to \Omega'$ defined between two measurable spaces, $(\Omega, \mathscr{F})$ and $(\Omega', \mathscr{F}')$, is called an $\mathscr{F}/\mathscr{F}'$-measurable function if for every $B \in \mathscr{F}'$, the set $\{\omega : f(\omega) \in B\} \in \mathscr{F}$. In particular, a $\mathscr{B}(\mathbb{R}^n)/\mathscr{B}(\mathbb{R}^k)$-measurable function is called a Borel function, where $\mathscr{B}(\mathbb{R}^n)$ denotes the Borel algebra over $\mathbb{R}^n$. The Borel algebra has the property $\mathscr{B}(\mathbb{R}^n) = \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}^{n-1})$, for any $n > 1$, where $\otimes$ represents the direct product. When the fields are obvious from the context, f will simply be called measurable (or Borel).

Random variables ($\mathscr{F}/\mathscr{B}(\mathbb{R})$-measurable functions), random vectors ($\mathscr{F}/\mathscr{B}(\mathbb{R}^n)$-measurable functions), and random elements ($\mathscr{F}/\mathscr{F}'$-measurable functions) are denoted by lower case bold letters $\boldsymbol{x}, \boldsymbol{y}$, and $\boldsymbol{z}$. Throughout the paper, it is assumed that every random element (including random variables and vectors) and stochastic process is defined over the same underlying probability space $(\Omega, \mathscr{F}, \mathrm{P})$.

Finally, let $\mathscr{G}$ be any sub σ-algebra of $\mathscr{F}$. A random element $\boldsymbol{x} : \Omega \to \Omega'$ is called $\mathscr{G}$-measurable if for every $B \in \mathscr{F}'$, the set $\{\omega : \boldsymbol{x}(\omega) \in B\} \in \mathscr{G}$. As usual, the smallest σ-algebra with respect to which $\boldsymbol{x}$ is measurable is denoted by $\sigma(\boldsymbol{x})$.

B. Main Result

This subsection states and proves the main theorem, which is as follows.

Theorem 2.1: Consider the system

$$\boldsymbol{x}(k+1) = F_k(\boldsymbol{x}(k), \boldsymbol{\theta}(k)), \quad \boldsymbol{x}(0) = \boldsymbol{x}_0, \qquad (2)$$

where $\boldsymbol{x}(k) \in \mathbb{R}^n$, $\boldsymbol{\theta}(k)$ is a discrete-time Markov chain in $\mathbb{R}^m$, $\boldsymbol{x}_0$ is an integrable random vector independent of $\boldsymbol{\theta}(k)$, and $F_k : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$ are measurable functions for every $k \geq 0$. The process $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$ is a Markov chain.

The proof of Theorem 2.1 relies on the following three supporting results. Their proofs are relegated to Sections III and IV to simplify the presentation of the proof of Theorem 2.1.

Theorem 2.2: Consider system (2) under the conditions of Theorem 2.1, and let $f : \mathbb{R}^m \to \mathbb{R}$ be any bounded Borel function. Then, it follows that

$$\mathbf{E}\{f(\boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} = \mathbf{E}\{f(\boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}.^1 \quad (3)$$

Theorem 2.3 (Theorem 1, p. 252 in [7] *):* Let $\mathcal{H}$ be a linear space of bounded real-valued functions defined on a set Φ. Assume that

(i) $\mathcal{H}$ contains the constant functions.
(ii) If $\{f_n\}$ is a sequence in $\mathcal{H}$ and $f_n \to f$ uniformly, then $f \in \mathcal{H}$.
(iii) If $\{f_n\}$ is a monotone sequence in $\mathcal{H}$ and $0 \le f_n \le M$ for all n, then $\lim_{n\to\infty} f_n \in \mathcal{H}$.
(iv) $\mathcal{H}$ has a subset $\mathcal{C}$ with the property: if $c_1, c_2 \in \mathcal{C}$, then their product, $c_1 c_2$, also belongs to $\mathcal{C}$.

Then $\mathcal{H}$ contains every bounded function $g : \Phi \to \mathbb{R}$ which is measurable with respect to the σ-algebra generated by the sets: $\{\phi : c(\phi) \in B\}$, $c \in \mathcal{C}$, $B \in \mathscr{B}(\mathbb{R})$.

Theorem 2.4: Consider system (2) under the conditions of Theorem 2.1, and let $\Phi = \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$. If $\mathcal{H}$ is the set of all bounded Borel functions $f : \Phi \to \mathbb{R}$ such that

$$\mathbf{E}\{f(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} = \mathbf{E}\{f(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}, \quad (4)$$

then $\mathcal{H}$ is a linear space that satisfies conditions (i)-(iv) of Theorem 2.3.

The fundamental step in the proof of Theorem 2.1 is to establish that the set $\mathcal{H}$ of all bounded Borel functions which satisfy (4) contains *every* bounded Borel function on Φ. This is proven by combining Theorems 2.3 and 2.4. Theorem 2.2 is used to verify property (iv) in Theorem 2.4.

Proof of Theorem 2.1: First recall from [13, page 564] that a random process $\boldsymbol{y}_k$ is a Markov chain if and only if for every bounded, real-valued Borel function, h, it follows that $\mathbf{E}\{h(\boldsymbol{y}_{n+1})|\boldsymbol{y}_n, \ldots, \boldsymbol{y}_0\} = \mathbf{E}\{h(\boldsymbol{y}_{n+1})|\boldsymbol{y}_n\}$. Thus, to show that $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$ is a Markov chain, it is sufficient to prove that for every bounded Borel function $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$

$$\mathbf{E}\{h(\boldsymbol{x}(k+1), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} = \mathbf{E}\{h(\boldsymbol{x}(k+1), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}$$

or, equivalently,

$$\mathbf{E}\{h(F_k(\boldsymbol{x}(k), \boldsymbol{\theta}(k)), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} = \mathbf{E}\{h(F_k(\boldsymbol{x}(k), \boldsymbol{\theta}(k)), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}.$$

Let $\mathcal{H}$ be the set of all the bounded Borel functions $f : \Phi \to \mathbb{R}$, $\Phi = \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$, that satisfy the expression

$$\mathbf{E}\{f(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} = \mathbf{E}\{f(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}. \quad (5)$$

[1] Equalities and inequalities between random variables/vectors as well as limits and convergence of sequences of random variables/vectors are taken in the almost everywhere sense.

It follows from Theorem 2.4 that $\mathcal{H}$ satisfies all the conditions of Theorem 2.3. Thus, $\mathcal{H}$ contains all the bounded functions $g : \Phi \to \mathbb{R}$ which are measurable with respect to $\mathscr{A}$, the σ-algebra generated by the sets $\{\phi : c(\phi) \in B\}$, $c \in \mathcal{C} \subseteq \mathcal{H}$, $B \in \mathscr{B}(\mathbb{R})$.

As will be shown in the proof of Theorem 2.4, the set $\mathcal{C}$ is composed of all bounded, separable functions $c : \Phi \to \mathbb{R}$ of the form $c(\phi) = \hat{c}_1(\gamma)\hat{c}_2(\lambda)$, where $\hat{c}_1 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $\hat{c}_2 : \mathbb{R}^m \to \mathbb{R}$ are bounded Borel functions, and $\phi = (\gamma, \lambda)$.

Now, let $A \in \mathscr{A}$ and observe that A must be of the form

$$A = \bigcup_{i,j} \{\phi : c_i(\phi) \in B_{i_j}\},$$

for some $c_i \in \mathcal{C}$ and $B_{i_j} \in \mathscr{B}(\mathbb{R})$. But the functions c_i are Borel, thus $\{\phi : c_i(\phi) \in B_{i_j}\} \in \mathscr{B}(\mathbb{R}^n) \otimes \mathscr{B}(\mathbb{R}^m) \otimes \mathscr{B}(\mathbb{R}^m)$, and consequently, $A \in \mathscr{B}(\mathbb{R}^n) \otimes \mathscr{B}(\mathbb{R}^m) \otimes \mathscr{B}(\mathbb{R}^m)$. Moreover, since this is true for every $A \in \mathscr{A}$, it follows that $\mathscr{A} \subseteq \mathscr{B}(\mathbb{R}^n) \otimes \mathscr{B}(\mathbb{R}^m) \otimes \mathscr{B}(\mathbb{R}^m)$. Conversely, note that for every $B \in \mathscr{B}(\mathbb{R}^n) \otimes \mathscr{B}(\mathbb{R}^m) \otimes \mathscr{B}(\mathbb{R}^m)$, $B = \{\phi : \mathbf{1}_{\{B\}}(\phi) \in [0.5, 1.5)\} \in \mathscr{A}$, since the indicator function of B, $\mathbf{1}_{\{B\}}(\cdot)$, belongs to $\mathcal{C}$. This implies that $\mathscr{A} \supseteq \mathscr{B}(\mathbb{R}^n) \otimes \mathscr{B}(\mathbb{R}^m) \otimes \mathscr{B}(\mathbb{R}^m)$, so $\mathscr{A} = \mathscr{B}(\mathbb{R}^n) \otimes \mathscr{B}(\mathbb{R}^m) \otimes \mathscr{B}(\mathbb{R}^m)$.

Hence, $\mathcal{H}$ contains every bounded Borel function of the form $g : \Phi \to \mathbb{R}$, and each of these functions satisfies (5). In particular, let $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be any bounded Borel function and define $g(\phi) = h(F_k(\gamma), \lambda)$, $\phi = (\gamma, \lambda)$. Clearly, $g \in \mathcal{H}$, so in (5)

$$\mathbf{E}\{g(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} = \mathbf{E}\{g(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\},$$

which implies that

$$\mathbf{E}\{h(F_k(\boldsymbol{x}(k), \boldsymbol{\theta}(k)), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} = \mathbf{E}\{h(F_k(\boldsymbol{x}(k), \boldsymbol{\theta}(k)), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\},$$

or

$$\mathbf{E}\{h(\boldsymbol{x}(k+1), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} = \mathbf{E}\{h(\boldsymbol{x}(k+1), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}.$$

Since this is true for every bounded Borel function $h : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, we conclude that $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$ is a Markov chain. ■

III. Proof of Theorem 2.2

The proof of Theorem 2.2 relies on three items to be proven about system (2):

Item (i): The σ-algebra generated by the joint process $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$ can be simplified as follows:

$$\sigma(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)) = \sigma(\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0)).$$

Item (ii): For any integrable random variable $\boldsymbol{\xi}$,

$$\mathbf{E}\{\boldsymbol{\xi}|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} = \mathbf{E}\{\boldsymbol{\xi}|\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0)\}.$$

Item (iii): $\mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\} = \mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{\theta}(k)\}$.

Item (i) is the core of the proof of Theorem 2.2. It establishes that all the probabilistic information is contained in

$\boldsymbol{x}(0)$ and $\boldsymbol{\theta}(i)$, $i = 0, \ldots, k$. That is, no new information can be obtained by examining the variables $\boldsymbol{x}(i)$, $i = 0, \ldots, k$, when $\boldsymbol{x}(0)$ and $\boldsymbol{\theta}(i)$, $i = 0, \ldots, k$, are available. Item (i) is proven in Subsection *A* (Theorem 3.4). Its proof makes use of the binary operator $\Cap$ (defined below) and standard results on the structure of σ-algebras induced by random variables. Items (ii) and (iii) are proven in Subsection *B* (Theorems 3.6 and 3.7), after briefly reviewing some properties of expected values and the basic definitions of independence. Finally, the proof of Theorem 2.2 is developed in Subsection *C*.

A. Some Properties of σ-Algebras

1) Structure:

Theorem 3.1: [14, Theorem 20.1] Let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a vector of random variables $\boldsymbol{x}_i$.

(i) The σ-algebra $\sigma(\boldsymbol{x}) = \sigma(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ consists exactly of the sets $\{\omega : \boldsymbol{x}(\omega) \in B\}$ for $B \in \mathscr{B}(\mathbb{R}^n)$.

(ii) The random variable $\boldsymbol{y}$ is $\sigma(\boldsymbol{x})$-measurable if and only if there exists a measurable function $f : \mathbb{R}^n \to \mathbb{R}$ such that $\boldsymbol{y}(\omega) = f(\boldsymbol{x}_1(\omega), \ldots, \boldsymbol{x}_n(\omega))$ for all $\omega \in \Omega$.

2) Combination:

Definition 3.1: Let $\mathcal{C}$ and $\mathcal{D}$ be two classes of subsets of Ω. The binary operation $\Cap$ combines the classes $\mathcal{C}$ and $\mathcal{D}$ to form a new class of subsets of Ω as follows:

$$\mathcal{C} \Cap \mathcal{D} \triangleq \{E \subset \Omega : E = C \cap D; C \in \mathcal{C}, D \in \mathcal{D}\}.$$

It follows readily from the definition that $\Cap$ is commutative and associative. It also follows that if $\mathscr{C}$ is an algebra (or σ-algebra) in Ω then $\mathscr{C} \Cap \mathscr{C} = \mathscr{C}$. The operation $\Cap$ can be used to combine σ-algebras induced by random variables as described below.

Lemma 3.1: For any two random variables $\boldsymbol{x}_1, \boldsymbol{x}_2$, $\sigma(\boldsymbol{x}_1) \Cap \sigma(\boldsymbol{x}_2) = \sigma(\boldsymbol{x}_1, \boldsymbol{x}_2)$.

Proof : Let $H \in \sigma(\boldsymbol{x}_1) \Cap \sigma(\boldsymbol{x}_2)$. Then $H = H_1 \cap H_2$, where $H_1 \in \sigma(\boldsymbol{x}_1)$ and $H_2 \in \sigma(\boldsymbol{x}_2)$. That is, $H_1 = \{\omega : \boldsymbol{x}_1(\omega) \in B_1\}$ and $H_2 = \{\omega : \boldsymbol{x}_2(\omega) \in B_2\}$ for some $B_1, B_2 \in \mathscr{B}(\mathbb{R})$. Thus, it follows that

$$H = \{\omega : (\boldsymbol{x}_1(\omega), \boldsymbol{x}_2(\omega)) \in B_1 \times B_2\} \in \sigma(\boldsymbol{x}_1, \boldsymbol{x}_2),$$

since $B_1 \times B_2 \in \mathscr{B}(\mathbb{R}^2)$. But H is arbitrary, so $\sigma(\boldsymbol{x}_1) \Cap \sigma(\boldsymbol{x}_2) \subseteq \sigma(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Conversely, let $H = \{\omega : (\boldsymbol{x}_1(\omega), \boldsymbol{x}_2(\omega)) \in B_1 \times B_2\}$ for some $B_1 \times B_2 \in \mathscr{B}(\mathbb{R}^2)$, and let $H_1 = \{\omega : \boldsymbol{x}_1(\omega) \in B_1\}$ and $H_2 = \{\omega : \boldsymbol{x}_2(\omega) \in B_2\}$. Clearly, $H = H_1 \cap H_2 \in \sigma(\boldsymbol{x}_1) \Cap \sigma(\boldsymbol{x}_2)$. Again, H is arbitrary, which implies that $\sigma(\boldsymbol{x}_1) \Cap \sigma(\boldsymbol{x}_2) \supseteq \sigma(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Thus, $\sigma(\boldsymbol{x}_1) \Cap \sigma(\boldsymbol{x}_2) = \sigma(\boldsymbol{x}_1, \boldsymbol{x}_2)$. ∎

Remark 3.1: Let $\boldsymbol{x}_1, \boldsymbol{x}_2$, and $\boldsymbol{x}_3$ be random variables and let $\boldsymbol{y}_1 = (\boldsymbol{y}_{11}, \ldots, \boldsymbol{y}_{1n})$ and $\boldsymbol{y}_2 = (\boldsymbol{y}_{21}, \ldots, \boldsymbol{y}_{2n})$ be vectors of random variables.

(i) Observe from Theorem 3.1, the definition of $\Cap$ and Lemma 3.1 that $\sigma(\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3) = \sigma(\boldsymbol{x}_1) \Cap \sigma(\boldsymbol{x}_2, \boldsymbol{x}_3) = \sigma(\boldsymbol{x}_1, \boldsymbol{x}_2) \Cap \sigma(\boldsymbol{x}_2, \boldsymbol{x}_3)$.

(ii) A similar derivation shows that $\sigma(\boldsymbol{y}_1, \boldsymbol{y}_2) = \sigma(\boldsymbol{y}_{11}, \ldots, \boldsymbol{y}_{1n}, \boldsymbol{y}_{21}, \ldots, \boldsymbol{y}_{2n}) = \sigma(\boldsymbol{y}_1) \Cap \sigma(\boldsymbol{y}_2)$.

Both arguments can be extended inductively to any finite number of random variables or random vectors. This remark plays an important role in the following collection of results.

3) Simplification of σ-algebras:

Theorem 3.2: Let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a vector of random variables, and let $f : \mathbb{R}^n \to \mathbb{R}$ be a measurable function. If $\boldsymbol{y} = f(\boldsymbol{x})$ then $\sigma(\boldsymbol{y}, \boldsymbol{x}) = \sigma(\boldsymbol{x})$.

Proof : First, recall from Theorem 3.1 that $\sigma(\boldsymbol{y}, \boldsymbol{x})$ is composed of the sets $\{\omega : (\boldsymbol{y}(\omega), \boldsymbol{x}(\omega)) \in B\}$, $B \in \mathscr{B}(\mathbb{R}^{n+1})$. Fix any $B \in \mathscr{B}(\mathbb{R}^{n+1})$ and observe that $B = B_1 \times B_2$, for some $B_1 \in \mathscr{B}(\mathbb{R})$ and $B_2 \in \mathscr{B}(\mathbb{R}^n)$, since $\mathscr{B}(\mathbb{R}^{n+1}) = \mathscr{B}(\mathbb{R}) \otimes \mathscr{B}(\mathbb{R}^n)$. Then, it follows that

$$\begin{aligned}
&\{\omega : (\boldsymbol{y}(\omega), \boldsymbol{x}(\omega)) \in B\} \\
&\qquad = \{\omega : \boldsymbol{y}(\omega) \in B_1\} \cap \{\omega : \boldsymbol{x}(\omega) \in B_2\} \\
&\qquad = \{\omega : \boldsymbol{x}(\omega) \in f^{-1}(B_1) \cap B_2\} \in \sigma(\boldsymbol{x}).
\end{aligned}$$

The last inclusion above follows from the measurability of f, which implies that $f^{-1}(B_1) \cap B_2 \in \mathscr{B}(\mathbb{R}^n)$. Hence, $\{\omega : (\boldsymbol{y}(\omega), \boldsymbol{x}(\omega)) \in B\} \in \sigma(\boldsymbol{x})$ and, since B is arbitrary, this implies that $\sigma(\boldsymbol{y}, \boldsymbol{x}) \subseteq \sigma(\boldsymbol{x})$.

Conversely, note that $\sigma(\boldsymbol{x})$ is composed of the sets $\{\omega : \boldsymbol{x}(\omega) \in B\}$, $B \in \mathscr{B}(\mathbb{R}^n)$. Also note that $\boldsymbol{y}$ is a random variable (Theorem 3.1) and that

$$\Omega = \{\omega : \boldsymbol{y}(\omega) \in \text{Range}(f)\} = \{\omega : \boldsymbol{y}(\omega) \in \mathbb{R}\}.$$

Thus, fix any $B \in \mathscr{B}(\mathbb{R}^n)$ and observe that

$$\begin{aligned}
\{\omega : \boldsymbol{x}(\omega) \in B\} &= \{\omega : \boldsymbol{x}(\omega) \in B\} \cap \{\omega : \boldsymbol{y}(\omega) \in \mathbb{R}\} \\
&= \{\omega : (\boldsymbol{y}(\omega), \boldsymbol{x}(\omega)) \in \mathbb{R} \times B\}.
\end{aligned}$$

Now, $\mathbb{R} \times B \in \mathscr{B}(\mathbb{R}^{n+1})$ so $\{\omega : (\boldsymbol{y}(\omega), \boldsymbol{x}(\omega)) \in \mathbb{R} \times B\} \in \sigma(\boldsymbol{y}, \boldsymbol{x})$. This and the arbitrariness of B yields $\sigma(\boldsymbol{x}) \subseteq \sigma(\boldsymbol{y}, \boldsymbol{x})$. Hence, $\sigma(\boldsymbol{x}) = \sigma(\boldsymbol{y}, \boldsymbol{x})$. ∎

This result can be extended to the case when $\boldsymbol{y}$ is a vector-valued function of $\boldsymbol{x}$ as follows.

Theorem 3.3: Let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be vector of random variables and $F : \mathbb{R}^n \to \mathbb{R}^m$ be a measurable function. If $\boldsymbol{y} = F(\boldsymbol{x})$ then $\sigma(\boldsymbol{y}, \boldsymbol{x}) = \sigma(\boldsymbol{x})$.

4) Proof of Item (i):

Theorem 3.4: Consider system (2) under the conditions of Theorem 2.1. It follows that

$$\sigma(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)) = \sigma(\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0)). \tag{6}$$

Proof: Note that $\boldsymbol{x}(k) = F_k(\boldsymbol{x}(k-1), \boldsymbol{\theta}(k-1))$, $k \geq 1$. It follows from Theorem 3.3 that $\sigma(\boldsymbol{x}(k), \boldsymbol{x}(k-1), \boldsymbol{\theta}(k-1)) = \sigma(\boldsymbol{x}(k-1), \boldsymbol{\theta}(k-1))$. Trivially then (6) holds for $k = 1$. Now, suppose it holds up to some fixed $k = i-1$. Then

$$\begin{aligned}
&\sigma(\boldsymbol{x}(i), \boldsymbol{\theta}(i), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)) \\
&\quad = \sigma(\boldsymbol{\theta}(i)) \Cap \sigma(\boldsymbol{x}(i), \boldsymbol{x}(i-1), \boldsymbol{\theta}(i-1), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)) \\
&\quad = \sigma(\boldsymbol{\theta}(i)) \Cap \sigma(\boldsymbol{x}(i), \boldsymbol{x}(i-1), \boldsymbol{\theta}(i-1)) \Cap \\
&\qquad \sigma(\boldsymbol{x}(i-2), \boldsymbol{\theta}(i-2), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)) \\
&\quad = \sigma(\boldsymbol{\theta}(i)) \Cap \sigma(\boldsymbol{x}(i-1), \boldsymbol{\theta}(i-1)) \Cap \\
&\qquad \sigma(\boldsymbol{x}(i-2), \boldsymbol{\theta}(i-2), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)) \\
&\quad = \sigma(\boldsymbol{\theta}(i)) \Cap \sigma(\boldsymbol{x}(i-1), \boldsymbol{\theta}(i-1), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)) \\
&\quad = \sigma(\boldsymbol{\theta}(i)) \Cap \sigma(\boldsymbol{\theta}(i-1), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0)) \\
&\quad = \sigma(\boldsymbol{\theta}(i), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)).
\end{aligned}$$

Thus (6) also holds for $k = i$, and therefore by induction for all $k \geq 1$. ∎

B. On Independence and Expected Values

1) Independence:

Definition 3.2: Let $\mathscr{C}_i$, $i = 1, \ldots, n$ be a set of sub σ-algebras of $\mathscr{F}$. Also, let $f_i : (\Omega, \mathscr{F}) \to (\Omega_i, \mathscr{F}_i)$, $i = 1, \ldots, n$ be a set random elements.

- (i) The σ-algebras $\mathscr{C}_1, \ldots, \mathscr{C}_n$ are independent if for every choice of sets $C_i \in \mathscr{C}_i$ it follows that $\mathrm{P}(B_1 \cap \cdots \cap B_n) = \mathrm{P}(B_1) \cdot \cdots \cdot \mathrm{P}(B_n)$ for all the possible 2^n combinations formed by taking $B_i = C_i$ or $B_i = \Omega$, $i = 1, 2, \ldots, n$.
- (ii) The random elements $f_1, \ldots, f_n$ are independent if $\sigma(f_1), \ldots, \sigma(f_n)$ are independent.
- (iii) A random variable $\boldsymbol{y}$ and the stochastic process $\mathscr{X} = \{\boldsymbol{x}_k\}$ are independent if for every finite integer $n \geq 1$ and every sequence of integers $0 \leq t_1 < \cdots < t_n < \infty$ it follows that $\sigma(\boldsymbol{y})$ and $\sigma(\boldsymbol{x}_{t_1}, \ldots, \boldsymbol{x}_{t_n})$ are independent.

Remark 3.2: Let $\mathscr{C}$ and $\mathscr{D}$ be independent sub σ-algebras of $\mathscr{F}$. Clearly, if $\boldsymbol{x}$ and $\boldsymbol{y}$ are, respectively, $\mathscr{C}$-measurable and $\mathscr{D}$-measurable random elements, then $\boldsymbol{x}$ and $\boldsymbol{y}$ are independent.

2) Expected Values:

Lemma 3.2: The following are standard results from the literature (cf. [13]).

- (i) Let $\boldsymbol{x}$ and $\boldsymbol{y}$ be independent random variables. Then $\mathbf{E}\{\boldsymbol{x}\boldsymbol{y}\} = \mathbf{E}\{\boldsymbol{x}\}\mathbf{E}\{\boldsymbol{y}\}$, where $\boldsymbol{x}\boldsymbol{y}$ is the product of $\boldsymbol{x}$ and $\boldsymbol{y}$.
- (ii) Let $\boldsymbol{x} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$ be a vector of random variables. If $\boldsymbol{x}$ is independent of a random variable $\boldsymbol{y}$, then so is each $\boldsymbol{x}_i$. Furthermore, $\mathbf{E}\{\boldsymbol{x}\boldsymbol{y}\} = \mathbf{E}\{\boldsymbol{x}\}\mathbf{E}\{\boldsymbol{y}\}$.
- (iii) Let $\mathscr{F}_1$ and $\mathscr{F}_2$ be σ-algebras such that $\mathscr{F}_1 \subseteq \mathscr{F}_2$. If $\boldsymbol{x}$ is an integrable random variable then
$$\mathbf{E}\{\mathbf{E}\{\boldsymbol{x}|\mathscr{F}_1\}|\mathscr{F}_2\} = \mathbf{E}\{\mathbf{E}\{\boldsymbol{x}|\mathscr{F}_2\}|\mathscr{F}_1\} = \mathbf{E}\{\boldsymbol{x}|\mathscr{F}_1\}.$$
- (iv) For any $\mathscr{F}$-measurable random element $\boldsymbol{x}$, $\mathbf{E}\{\boldsymbol{x}|\mathscr{F}\} = \boldsymbol{x}$.

The next theorem extends a basic result for random variables. It is offered without proof for space limitations.

Theorem 3.5: Let $\boldsymbol{x}$ be a vector of random variables and let $\boldsymbol{y}, \boldsymbol{z}$ be two random elements such that $\boldsymbol{z}$ is independent of $(\boldsymbol{x}, \boldsymbol{y})$, i.e., $\sigma(\boldsymbol{z})$ is independent from $\sigma(\boldsymbol{z})$, $\sigma(\boldsymbol{y})$, and $\sigma(\boldsymbol{x}, \boldsymbol{y})$. Then it follows that $\mathbf{E}\{\boldsymbol{x}|\boldsymbol{y}\boldsymbol{z}\} = \mathbf{E}\{\boldsymbol{x}|\boldsymbol{y}\}$.

Corollary 3.1: Let $\boldsymbol{\theta}(k)$ be a discrete-time Markov chain in $\mathbb{R}^m$ and $\boldsymbol{x}_0$ a second order random vector. If $\boldsymbol{x}_0$ is independent from $\boldsymbol{\theta}(k)$, then $\mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}_0\} = \mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0)\}$.

3) Proofs of Items (ii) and (iii):

Theorem 3.6: Consider system (2) under the conditions of Theorem 2.1. Then, for any integrable random variable $\boldsymbol{\xi}$,

$$\mathbf{E}\{\boldsymbol{\xi}|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} = \mathbf{E}\{\boldsymbol{\xi}|\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0)\}.$$

Proof: Observe that the integrability of $\boldsymbol{\xi}$ ensures $\mathbf{E}\{\boldsymbol{\xi}|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\}$ is well defined and

$$\int_B \mathbf{E}\{\boldsymbol{\xi}|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\}\, \mathrm{dP} = \int_B \boldsymbol{\xi}\, \mathrm{dP},$$

for every $B \in \sigma(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0))$. But Theorem 3.4 shows that $\sigma(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)) = \sigma(\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0))$. Thus, for every such B

$$\int_B \boldsymbol{\xi}\, \mathrm{dP} = \int_B \mathbf{E}\{\boldsymbol{\xi}|\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0)\}\, \mathrm{dP},$$

and the result follows. ∎

Theorem 3.7: Consider system (2) under the conditions of Theorem 2.1. Then $\mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\} = \mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{\theta}(k)\}$.

Proof: First, note that $\sigma(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$ is comprised of sets of the form $\{\omega : (\boldsymbol{x}(k)(\omega), \boldsymbol{\theta}(k)(\omega)) \in H\}$, $H \in \mathscr{B}(\mathbb{R}^n) \otimes \mathscr{B}(\mathbb{R}^m)$. Fix any $H \in \mathscr{B}(\mathbb{R}^n) \otimes \mathscr{B}(\mathbb{R}^m)$ and observe that

$$\begin{aligned}\{\omega :&(\boldsymbol{x}(k)(\omega), \boldsymbol{\theta}(k)(\omega)) \in H\} \\ &= \{\omega : \boldsymbol{x}(k)(\omega) \in H_1\} \cap \{\omega : \boldsymbol{\theta}(k)(\omega) \in H_2\}, \quad (7)\end{aligned}$$

for some $H_1 \in \mathscr{B}(\mathbb{R}^n)$ and $H_2 \in \mathscr{B}(\mathbb{R}^m)$. Next, note that (2) yields

$$\begin{aligned}\boldsymbol{x}(k) &= F_{k-1}(\boldsymbol{x}(k-1), \boldsymbol{\theta}(k-1)) \\ &= F_{k-1}(F_{k-2}(\boldsymbol{x}(k-2), \boldsymbol{\theta}(k-2)), \boldsymbol{\theta}(k-1)) \\ &\vdots \\ &= G(\boldsymbol{\theta}(k-1), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0)),\end{aligned}$$

where G is the indicated composition of the functions $F_0, \ldots, F_{k-1}$. Thus, (7) can be expressed as

$$\begin{aligned}&\{\omega : (\boldsymbol{x}(k)(\omega), \boldsymbol{\theta}(k)(\omega)) \in H\} \\ &= \{\omega : (\boldsymbol{\theta}(k)(\omega), \ldots, \boldsymbol{\theta}(0)(\omega), \boldsymbol{x}(0)(\omega)) \in H_2 \times G^{-1}(H_1)\}.\end{aligned}$$

Since G is measurable, it follows that

$$H_2 \times G^{-1}(H_1) \in \underbrace{\mathscr{B}(\mathbb{R}^m) \otimes \cdots \otimes \mathscr{B}(\mathbb{R}^m)}_{\text{k+1 copies}} \otimes \mathscr{B}(\mathbb{R}^n),$$

which in turn implies that $\{\omega : (\boldsymbol{x}(k)(\omega), \boldsymbol{\theta}(k)(\omega)) \in H\} \in \sigma(\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0))$. But since H is arbitrary, it follows that $\sigma(\boldsymbol{x}(k), \boldsymbol{\theta}(k)) \subseteq \sigma(\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0))$. This conclusion, Lemma 3.2, Corollary 3.1, and the Markov property of $\boldsymbol{\theta}(k)$ yield

$$\begin{aligned}\mathbf{E}\{&\boldsymbol{\theta}(k+1)|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\} \\ &= \mathbf{E}\{\mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0)\}|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\} \\ &= \mathbf{E}\{\mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0)\}|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\} \\ &= \mathbf{E}\{\mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{\theta}(k)\}|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}.\end{aligned}$$

Finally, note that $\mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{\theta}(k)\}$ is $\sigma(\boldsymbol{\theta}(k))$-measurable. Also note that $\sigma(\boldsymbol{\theta}(k)) \subseteq \sigma(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$, so $\mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{\theta}(k)\}$ is also $\sigma(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$-measurable. Thus, Lemma 3.2 yields $\mathbf{E}\{\mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{\theta}(k)\}|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\} = \mathbf{E}\{\boldsymbol{\theta}(k+1)|\boldsymbol{\theta}(k)\}$, and the result follows. ∎

C. Proof of Theorem 2.2

Proof: First, observe that $f(\boldsymbol{\theta}(k+1))$ is an integrable random variable, so Theorem 3.6 yields

$$\begin{aligned}\mathbf{E}\{&f(\boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} \\ &= \mathbf{E}\{f(\boldsymbol{\theta}(k+1))|\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}(0)\}. \quad (8)\end{aligned}$$

Next, note that $f(\boldsymbol{\theta}(k+1))$ is $\sigma(\boldsymbol{\theta}(k+1))$-measurable (Theorem 3.1), so $\sigma(f(\boldsymbol{\theta}(k+1)), \boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0)) \subseteq \sigma(\boldsymbol{\theta}(k+1), \ldots, \boldsymbol{\theta}(0))$. Furthermore, this and the hypothesis of Theorem 2.2 imply that $\boldsymbol{x}(0)$ is independent of $(f(\boldsymbol{\theta}(k+$

$1)), \boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0))$. Hence, (8), Theorem 3.5, and the Markov nature of $\boldsymbol{\theta}(k)$ yield

$$\mathbf{E}\{f(\boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\} = \mathbf{E}\{f(\boldsymbol{\theta}(k+1))|\boldsymbol{\theta}(k)\}. \quad (9)$$

From Theorem 3.7, $\sigma(\boldsymbol{x}(k), \boldsymbol{\theta}(k)) \subseteq \sigma(\boldsymbol{\theta}(k), \ldots, \boldsymbol{\theta}(0), \boldsymbol{x}_0)$. Thus, the argument in the proof of Theorem 3.7 yields $\mathbf{E}\{f(\boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\} = \mathbf{E}\{f(\boldsymbol{\theta}(k+1))|\boldsymbol{\theta}(k)\}$, which together with (9) confirm identity (3). ∎

IV. Proof of Theorem 2.4

Proof: Observe that $\mathcal{H}$ is a subset of the linear space of bounded, real-valued functions over Φ. Moreover, if $f_1, f_2 \in \mathcal{H}$ and $\alpha_1, \alpha_2 \in \mathbb{R}$, it follows that $\alpha_1 f_1 + \alpha_2 f_2$ satisfies (4). Thus, $\mathcal{H}$ is a subspace of a linear space, which in turn implies that it is a linear space.

Now, to show (i) in Theorem 2.3, suppose $f(\phi) = a$ for all $\phi \in \Phi$ and some $a \in \mathbb{R}$. Note that $f(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))$ is a constant random variable. Also, recall that for any σ-algebra $\mathscr{G}$, $\mathbf{E}\{f|\mathscr{G}\} = a$ (a.e.) [13, page 215], which in turn implies that f satisfies (4). Since this is true for any $a \in \mathbb{R}$, then $\mathcal{H}$ contains all the constant functions.

To show (ii), suppose that $\{f_n\} \in \mathcal{H}$ converges uniformly to f, and let L_n be finite constants such that $|f_n(\phi)| < L_n$ for all $\phi \in \Phi$. Next, fix $\epsilon > 0$ and observe from the hypothesis that there exists $N(\epsilon)$ such that $|f_n(\phi) - f(\phi)| < \epsilon$ for all $n \geq N(\epsilon)$ and for all $\phi \in \Phi$. In particular, $|f_{N(\epsilon)}(\phi) - f(\phi)| < \epsilon$ implies that $|f(\phi)| < L_{N(\epsilon)} + \epsilon$ for every $\phi \in \Phi$. Note that the uniform convergence of $\{f_n\}$ implies that it also converges pointwise almost everywhere to f. This in turn shows f is measurable [15, Corollary 2.2.4]. Hence f is a bounded Borel function.

Set $L = L_{N(\epsilon)} + \epsilon$ and observe that $|f_n(\phi) - f(\phi)| < \epsilon$, $n \geq N(\epsilon)$ implies that $|f_n(\phi)| < L + \epsilon$ for all $\phi \in \Phi$ and all $n \geq N(\epsilon)$. Next, define the function $g : \Phi \to \mathbb{R}$ as $g(\phi) = \max\{L_1, \ldots, L_{N(\epsilon)-1}, L + \epsilon\}$, for all $\phi \in \Phi$, and let $\boldsymbol{y}_n = f_n(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))$ and $\boldsymbol{y} = g(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))$. Observe that $\{\boldsymbol{y}_n\}$ converges pointwise to $f(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))$, that $|\boldsymbol{y}_n| < \boldsymbol{y}$, and that $\mathbf{E}\{\boldsymbol{y}\} = \max\{L_1, \ldots, L_{N(\epsilon)-1}, L + \epsilon\} < \infty$. Hence, it follows that for any σ-algebra $\mathscr{G}$, $\mathbf{E}\{\boldsymbol{y}_n|\mathscr{G}\}$ converges to $\mathbf{E}\{f(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))|\mathscr{G}\}$ [13, page 218]. Thus,

$$\lim_{n\to\infty} \mathbf{E}\{\boldsymbol{y}_n|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\}$$
$$= \mathbf{E}\{f(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\}.$$

But $\{f_n\} \in \mathcal{H}$. Thus, (4) implies that

$$\lim_{n\to\infty} \mathbf{E}\{\boldsymbol{y}_n|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\}$$
$$= \lim_{n\to\infty} \mathbf{E}\{\boldsymbol{y}_n|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}$$
$$= \mathbf{E}\{f(\boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}.$$

Hence, $f \in \mathcal{H}$, and (ii) follows.

To show (iii), let $\{f_n\}$ be a monotone sequence of functions in $\mathcal{H}$ such that $0 \leq f_n \leq M < \infty$, and note that $\{f_n\}$ converges pointwise to a function f, which is in turn a bounded Borel function. Thus, (iii) follows by using the same argument as in (ii).

To verify (iv), consider the set $\mathcal{C}$ composed of all the bounded separable functions $c : \Phi \to \mathbb{R}$ of the form $c(\phi) = \hat{c}_1(\gamma)\hat{c}_2(\lambda)$, $\phi = (\gamma, \lambda)$, where $\hat{c}_1 : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $\hat{c}_2 : \mathbb{R}^m \to \mathbb{R}$ are bounded Borel functions.

It is necessary to show that if $c(\phi), d(\phi) \in \mathcal{C}$ then $c(\phi)d(\phi) \in \mathcal{C}$, and that $\mathcal{C} \subseteq \mathcal{H}$. Note that $c(\phi)d(\phi) = (\hat{c}_1(\gamma)\hat{d}_1(\gamma))(\hat{c}_2(\lambda)\hat{d}_2(\lambda))$. But $\hat{c}_1(\gamma)\hat{d}_1(\gamma)$ and $\hat{c}_2(\lambda)\hat{d}_2(\lambda)$ are real, bounded Borel functions in, respectively, $\mathbb{R}^n \times \mathbb{R}^m$ and $\mathbb{R}^m$. Thus $c(\phi)d(\phi) \in \mathcal{C}$.

Finally, to see that $\mathcal{C} \in \mathcal{H}$, recall from [13, page 216] that if $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are random variables such that $\mathbf{E}\{|\boldsymbol{\xi}|\} < \infty$, $\mathbf{E}\{|\boldsymbol{\xi\eta}|\} < \infty$, and $\boldsymbol{\eta}$ is $\mathscr{G}$-measurable, then $\mathbf{E}\{\boldsymbol{\xi\eta}|\mathscr{G}\} = \boldsymbol{\eta}\mathbf{E}\{\boldsymbol{\xi}|\mathscr{G}\}$. Thus let $c(\phi) = \hat{c}_1(\gamma)\hat{c}_2(\lambda) \in \mathcal{C}$, $\phi = (\gamma, \lambda)$, and observe from Theorem 2.2 that

$$\mathbf{E}\{c((\boldsymbol{x}(k), \boldsymbol{\theta}(k)), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\}$$
$$= \hat{c}_1(\boldsymbol{x}(k), \boldsymbol{\theta}(k))\mathbf{E}\{\hat{c}_2(\boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k), \ldots, \boldsymbol{x}(0), \boldsymbol{\theta}(0)\}$$
$$= \hat{c}_1(\boldsymbol{x}(k), \boldsymbol{\theta}(k))\mathbf{E}\{\hat{c}_2(\boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}$$
$$= \mathbf{E}\{\hat{c}_1(\boldsymbol{x}(k), \boldsymbol{\theta}(k))\hat{c}_2(\boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}$$
$$= \mathbf{E}\{c((\boldsymbol{x}(k), \boldsymbol{\theta}(k)), \boldsymbol{\theta}(k+1))|\boldsymbol{x}(k), \boldsymbol{\theta}(k)\}.$$

Thus, c satisfies (4). Since this is true for every $c \in \mathcal{C}$, then $\mathcal{C} \in \mathcal{H}$ and (iv) follows. This completes the proof. ∎

V. Examples

This section applies Theorem 2.1 to three examples: A Markovian jump linear system, a jump linear system driven by a finite state machine with a Markovian input, and a hybrid jump linear system with a Markovian input.

A. Markov Jump Linear System

Consider the jump linear system

$$\boldsymbol{x}(k+1) = A_{\boldsymbol{\theta}(k)} x(k), \quad (10)$$

where $\boldsymbol{\theta}(k) \in \mathcal{I}_{l_\theta} \triangleq \{1, \ldots, l_\theta\}$ is a Markov chain [2], $A_i, i \in \mathcal{I}_N$ are $n \times n$ matrices, and $\boldsymbol{x}(0)$ is a second order random vector independent of $\boldsymbol{\theta}(k)$. As stated in Section I, it is well known that $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$ is a Markov chain. However, it is interesting to note how this example fits in our framework. Observe that (10) can be written as

$$\boldsymbol{x}(k+1) = \sum_{i=1}^{l_\theta} A_i \mathbf{1}_{\{\boldsymbol{\theta}(k)=i\}} x(k) \triangleq F(\boldsymbol{x}(k), \boldsymbol{\theta}(k)), \quad (11)$$

where $\mathbf{1}_{\{\cdot\}}$ is the indicator function. In (11), each summand is a product of two measurable functions, which shows that F is measurable. Since this satisfies the hypothesis of Theorem 2.1, then $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$ is a Markov chain.

B. Jump Linear System driven by a Finite State Machine

These class of systems has been used to model computer upsets in digital flight controllers [11]. In such models, the output of a FSM, which represents a computer algorithm, selects the operating mode of a linear control system. Thus, in (10), $\boldsymbol{\theta}(k)$ represents a FSM's output, which may not be Markovian in general. Thus, $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$ may not be a Markov chain as well. To produce a Markov sequence, one has to include information about the FSM's input.

[2] In the sequel, $\mathcal{I}_l \triangleq \{1, \ldots, l\}$ for any integer $l \geq 1$.

A FSM can be viewed as a 5-tuple $\mathcal{M} = (\mathcal{I}_N, \Sigma_S, \mathcal{I}_\theta, \delta, \varpi)$, where $\mathcal{I}_N$ is the input symbol set; $\Sigma_S \triangleq \{e_0, \ldots, e_s\}$, $e_j = [0, \ldots, 0, \underbrace{0, 1, 0}_{j\text{-th position}}, \ldots, 0]^T$, is the FSM's state set; $\mathcal{I}_\theta$ is the output symbol set; and $\delta : \mathcal{I}_N \times \Sigma_S \to \Sigma_S$ and $\varpi : \Sigma_S \to \mathcal{I}_O$ are, respectively, the state transition and the output maps [11]. Let the FSM's input sequence be given by $\boldsymbol{N}(k)$, a Markov chain in $\mathcal{I}_N$, and let $\boldsymbol{z}(k)$ represent the states of the FSM. Then, δ is given by

$$\boldsymbol{z}(k) = S_{\boldsymbol{N}(k-1)} \boldsymbol{z}(k-1),$$

where S_j, $j \in \mathcal{I}_N$, are deterministic transition matrices, i.e., matrices with a single 1 in each column and zeros in the other entries. The output sequence, $\boldsymbol{\theta}(k)$, is given by $\boldsymbol{\theta}(k) = \varpi(\boldsymbol{z}(k))$. For simplicity, it will be assumed that ϖ is an isomorphism such that $\varpi(e_i) = i$. With a slight abuse of notation, the dynamics of $\boldsymbol{\theta}(k)$ are derived as follows

$$\begin{aligned}\boldsymbol{\theta}(k) &= S_{\boldsymbol{N}(k-1)} \boldsymbol{\theta}(k-1) = \sum_{j=1}^{l_N} S_j \mathbf{1}_{\{\boldsymbol{N}(k-1)=j\}} \boldsymbol{\theta}(k-1) \\ &\triangleq G(\boldsymbol{\theta}(k-1), \boldsymbol{N}(k-1)). \qquad (12)\end{aligned}$$

Combining (11) and (12) yields

$$\begin{aligned}\begin{bmatrix} \boldsymbol{x}(k+1) \\ \boldsymbol{\theta}(k) \end{bmatrix} &= \begin{bmatrix} F(\boldsymbol{x}(k), G(\boldsymbol{\theta}(k-1), \boldsymbol{N}(k-1))) \\ G(\boldsymbol{\theta}(k-1), \boldsymbol{N}(k-1)) \end{bmatrix} \\ &\triangleq H(\boldsymbol{x}(k), \boldsymbol{\theta}(k-1), \boldsymbol{N}(k-1)).\end{aligned}$$

Clearly, H is a measurable function. Furthermore, by defining $\boldsymbol{y}(k) = [\boldsymbol{x}(k+1)^T, \boldsymbol{\theta}(k)]^T$, the equation above becomes

$$\boldsymbol{y}(k) = H(\boldsymbol{y}(k-1), \boldsymbol{N}(k-1)).$$

Thus, provided that $\boldsymbol{x}(1)$ and $\boldsymbol{\theta}(0)$ are independent from $\boldsymbol{N}(k)$, Theorem 2.1 shows that the process $(\boldsymbol{x}(k+1), \boldsymbol{\theta}(k), \boldsymbol{N}(k))$ is a Markov chain.

C. Hybrid Jump Linear System

The structure of a HJLS is similar to that of jump linear system driven by a finite state machine, except that the algorithm is allowed to make decisions based on state vector information [12]. This is achieved through a measurable quantization map $\psi : \mathbb{R}^n \to \mathcal{I}_{l_\psi}$, which divides $\mathbb{R}^n$ in l_ψ mutually exclusive subregions R_i, defined as

$$\psi(x) = i, \text{when } x \in R_i, i \in \mathcal{I}_{l_\psi}.$$

The algorithm gets information about the plant's state vector by making $\boldsymbol{\nu}(k) = \psi(\boldsymbol{x}(k))$ one of its inputs. Hence, the evolution of the FSM's output sequence is given by

$$\begin{aligned}\boldsymbol{\theta}(k) &= S_{\boldsymbol{N}(k-1)\boldsymbol{\nu}(k-1)} \boldsymbol{\theta}(k-1) \\ &= \sum_{i=1}^{l_N} \sum_{j=1}^{l_\psi} S_{ij} \mathbf{1}_{\{\boldsymbol{N}(k-1)=i, \psi(\boldsymbol{x}(k-1))=j\}} \boldsymbol{\theta}(k-1) \\ &\triangleq \tilde{G}(\boldsymbol{x}(k-1), \boldsymbol{\theta}(k-1), \boldsymbol{N}(k-1)).\end{aligned}$$

This expression and (11) yield, after introducing an auxiliary state vector,

$$\begin{aligned}\begin{bmatrix} \boldsymbol{x}(k+1) \\ \boldsymbol{x}_A(k+1) \\ \boldsymbol{\theta}(k) \end{bmatrix} &= \begin{bmatrix} F(\boldsymbol{x}(k), \tilde{G}(\boldsymbol{x}(k-1), \boldsymbol{\theta}(k-1), \boldsymbol{N}(k-1))) \\ \boldsymbol{x}(k) \\ \tilde{G}(\boldsymbol{x}(k-1), \boldsymbol{\theta}(k-1), \boldsymbol{N}(k-1)) \end{bmatrix} \\ &\triangleq \tilde{H}(\boldsymbol{x}(k), \boldsymbol{x}_A(k), \boldsymbol{\theta}(k-1), \boldsymbol{N}(k-1)).\end{aligned}$$

By defining $\tilde{\boldsymbol{y}}(k) = [\boldsymbol{x}(k+1)^T, \boldsymbol{x}_A(k+1)^T, \boldsymbol{\theta}(k)]^T$, the equation above can be expressed as

$$\tilde{\boldsymbol{y}}(k) = \tilde{H}(\tilde{\boldsymbol{y}}(k-1), \boldsymbol{N}(k-1)).$$

Thus, provided that $\boldsymbol{x}(1)$, $\boldsymbol{x}(0)$, and $\boldsymbol{\theta}(0)$ are independent from $\boldsymbol{N}(k)$, Theorem 2.1 shows that the process $(\boldsymbol{x}(k+1), \boldsymbol{x}(k), \boldsymbol{\theta}(k), \boldsymbol{N}(k))$ is a Markov chain.

Note that in the last two examples, the random vectors $\boldsymbol{y}(k)$ and $\tilde{\boldsymbol{y}}(k)$ represent the internal 'state' of the system. In this sense, $\boldsymbol{y}(k)$ and $\tilde{\boldsymbol{y}}(k)$ are equivalent to $\boldsymbol{x}(k)$ in the first example.

VI. CONCLUSIONS

A complete proof of the Markovian property of the state of general hybrid systems, $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$, has been presented. The proof is based on first principles and gives insight on the mechanism that produces the Markovianess of $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$. Further research is ongoing to determine the kernel functions of $(\boldsymbol{x}(k), \boldsymbol{\theta}(k))$ for particular HJLS's. These functions will enable the ergodic analysis of their associated HJLS, which is needed to establish their stability properties.

ACKNOWLEDGEMENTS

This research was supported by the NASA Langley Research Center under grants NNL04AA03A and NCC-1-03026.

REFERENCES

[1] W. S. Gray, O. R. González, and M. Dǒgan, "Stability analysis of digital linear flight controllers subject to electromagnetic disturbances," *IEEE Trans. Aerosp. Electron. Syst.*, no. 4, pp. 1204–1218, October 2000.

[2] H. Blom and J. Lygeros, "Hybridge final project report," 2005. [Online]. Available: http://hosted.nlr.nl/public/hosted-sites/hybridge/documents.html

[3] M. H. A. Davis, *Markov Models and Optimization*. New York: Chapman & Hall/CRC, 1993.

[4] Y. Fang and K. A. Loparo, "Stabilization of continuous-time jump linear systems," *IEEE Trans. Automat. Contr.*, vol. 47, no. 10, pp. 1590–1603, October 2002.

[5] M. Ehrhardt and Kleinman, "Controllability of linear stochastic systems," *Systems & Control Letters*, no. 3, pp. 145–153, 1982.

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

[7] T. Morozan, "On the stability of stochastic discrete systems," in *Control theory and Topics in Functional Analysis: Lectures Presented at an International Seminar Course at Trieste from 11 Sept. to 29 Nov. 1974*, International Atomic Energy Agency, Vienna, 1976, pp. 225–254.

[8] L. Arnold, *Stochastic Differential Equations: Theory and Applications*. New York: John Wiley & Sons, 1974.

[9] R. Srichander and K. Walker, "Stochastic stability analysis for continuous-time fault tolerant control systems," *Int. J. Control*, vol. 57, no. 2, pp. 433–452, 1993.

[10] L. Arnold and W. Kleinmann, "Qualitative theory of stochastic systems," in *Probabilistic Analysis and Related Topics, Vol 3*, A. T. Bharucha-Reid, Ed. New York: Academic Press, 1982, pp. 1–79.

[11] S. Patilkulkarni, H. Herencia-Zapana, W. S. Gray, and O. R. González, "On the stability of jump-linear systems driven by finite state machines with Markovian inputs," in *Proc. of the 2004 American Control Conference*, Boston, MA, 2004, pp. 2534–2539.

[12] A. Tejada, O. R. González, and W. S. Gray, "Asymptotic and mean square stability conditions for hybrid jump linear system with performance supervision," in *Proc. of the 2005 American Control Conference*, Portland, OR, 2005, pp. 569–574.

[13] A. N. Shiryaev, *Probability, Second Edition*. New York: Springer, 1995.

[14] P. Billingsley, *Probability and Measure, Third Edition*. New York: John Wiley & Sons, 1995.

[15] A. Friedman, *Foundations of Modern Analysis*. New York: Dover Publications Inc., 1982.