2011 American Control Conference
on O'Farrell Street, San Francisco, CA, USA
June 29 - July 01, 2011

Modeling of Symbolic Systems: Part I - Vector Space Representation of Probabilistic Finite State Automata★

Yicheng Wen yxw167@psu.edu
Asok Ray axr2@psu.edu
Ishanu Chattopadhyay ixc128@psu.edu
Shashi Phoha sxp26@psu.edu

The Pennsylvania State University
University Park, PA 16802, USA

Abstract— **This paper, which is the first of two parts, brings in the notions of vector addition and the associated scalar multiplication operations on probabilistic finite state automata (PFSA). A class of PFSA is shown to constitute a vector space over the real field $\mathbb{R}$, where the zero element is semantically equivalent to a subclass of PFSA, referred to as *symbolic white noise*. A norm is introduced on the vector space of PFSA and it quantifies the non-probabilistic behavior of a PFSA. The second part constructs a family of inner products on this vector space and presents numerical examples and applications.**

I. Introduction

Probabilistic finite state automata (PFSA) have emerged as a tool for modeling uncertain dynamical systems [1][2]. In this respect, symbolization-based techniques have been developed for probabilistic representation of dynamical systems to compensate for certain inadequacies of classical time-domain and frequency-domain system identification. The key feature of the work reported in this paper, which is the first of two parts, is formal language-theoretic and symbolic modeling instead of classical continuous-domain modeling. The basic approach is symbolic dynamic filtering (SDF) [3] that partitions the (possibly pre-processed) time series or image data observed from the underlying system to generate a string of symbols. Then, semantic models are constructed in the symbolic domain.

Although many algorithms have been proposed for constructing PFSA models from time series data, the theory of how to algebraically manipulate two PFSA has not been explored except for a few cases. The notion of vector space construction for finite state automata over the finite field $GF(2)$ was reported by Ray [4]. Barfoot and D'leuterio [5] proposed an algebraic construction for control of stochastic systems, where the algebra is defined for $m \times n$ stochastic matrices, which is only directly applicable to PFSA of the same structure (see Definition IV.6). A structural manipulation of PFSA models of dynamical systems has been addressed by Chattopadhyay and Ray [6], where the ability to project a PFSA model to an arbitrary structure is critical for synthesis of supervisory control algorithms for symbolic models.

★ This work has been supported in part by the U.S. Office of Naval Research under Grant No. N00014-09-1-0688, and the U.S. Army Research Laboratory and the U.S. Army Research Office under Grant No. W911NF-07-1-0376.

The major contribution of this paper is formulation of a normed vector space of a certain class of PFSA over the real field $\mathbb{R}$. The vector space formalism enriches the current theory for PFSA by taking into account disparate automaton structures and probability measures, as needed for many engineering applications (e.g., information fusion in sensor networks).

II. Preliminaries

In the formal language theory, an alphabet Σ is a (non-empty finite) set of symbols whose cardinality is denoted as $|\Sigma|$. A string x over Σ is a finite-length sequence of symbols in Σ, and its length, denoted by $|x|$, represents the number of symbols in x. The Kleene closure of Σ, denoted by $\Sigma^\star$, is the set of all finite-length strings of events including the null string ϵ; cardinality of $\Sigma^\star$ is $\aleph_0$. The set of all strictly infinite-length strings is denoted as Σ^ω; cardinality of Σ^ω is $\aleph_1$. The string xy is called concatenation of x and y. It is obvious that the null string ϵ is the identity for the concatenation operation.

Definition II.1 (PFSA) *A probabilistic finite state automaton (PFSA) is a tuple $G = (Q, \Sigma, \delta, q_0, \tilde{\pi})$, where*

- *Q is a (nonempty) finite set, called set of states;*
- *Σ is a (nonempty) finite set, called input alphabet;*
- *$\delta : Q \times \Sigma \to Q$ is the state transition function;*
- *$q_0 \in Q$ is the start state;*
- *$\tilde{\pi} : Q \times \Sigma \to [0, 1]$ is an output mapping which is known as a probability morph function and satisfies the condition $\sum_{\tau \in \Sigma} \tilde{\pi}(q_j, \tau) = 1$ for all $q_j \in Q$.*

The transition map δ naturally induces an extended transition function $\delta^\star : Q \times \Sigma^\star \to Q$ such that $\delta^\star(q, \epsilon) = q$ and $\delta^\star(q, x\tau) = \delta(\delta^\star(q, x), \tau)$ for $q \in Q$, $x \in \Sigma^\star$ and $\tau \in \Sigma$.

The probability morph function $\tilde{\pi}$ is represented in a matrix form as $\widetilde{\Pi}$ with the element $\widetilde{\Pi}_{ij} \triangleq \tilde{\pi}(q_i, \sigma_j)$, where $q_i \in Q$ and $\sigma_j \in \Sigma$. This paper assumes that all states in a PFSA are reachable from the start state. Otherwise, the non-reachable states should be removed from Q.

Definition II.2 (Probability Measure Space) *Given an alphabet Σ, the set $\mathscr{B}_\Sigma \triangleq 2^{\Sigma^\star}\Sigma^\omega$ is defined to be the σ-algebra [7] generated by the set $\left\{x\Sigma^\omega : x \in \Sigma^\star\right\}$.*

For brevity, the probability $p(x\Sigma^\omega)$ is denoted as $p(x)$, $\forall x \in \Sigma^\star$, in the sequel. That is, $p(x)$ is the probability of occurrence of all (infinitely long) strings with x as the prefix.

Definition II.3 (Probabilistic Nerode Relation [6]) *Given an alphabet Σ, any two strings $x, y \in \Sigma^\star$ are said to satisfy the probabilistic Nerode relation $\mathcal{N}_p$ on a probability space $(\Sigma^\omega, \mathscr{B}_\Sigma, p)$, and are denoted by $x\mathcal{N}_p y$, if either of the following conditions is true:*

1) $p(x) = p(y) = 0$;
2) $\forall \sigma \in \Sigma^\star$, $\frac{p(x\sigma)}{p(x)} = \frac{p(y\sigma)}{p(y)}$ *if* $p(x) \neq 0$ *and* $p(y) \neq 0$.

It is proven in [6] that the probabilistic Nerode relation $\mathcal{N}_p$ forms a right-invariant equivalence class. In the sequel, the probabilistic Nerode equivalence class of a string x on $\Sigma^\star$ is denoted by $[x]_p$, i.e., $[x]_p \triangleq \{z \in \Sigma^\star : x\mathcal{N}_p z\}$.

Remark II.1 *The probabilistic Nerode equivalence of a measure p induces a partition of strings on $\Sigma^\star$. In general, such a partition of $\Sigma^\star$ could have infinite index, i.e., consist of infinitely many equivalence classes. However, there must be finitely many equivalence classes for a probability measure that is encoded by a PFSA [6]. In other words, each PFSA must have a finite index as discussed later in this paper at the end of Section III.*

III. Vector Space of Probability Measures on $\mathscr{B}_\Sigma$

Given the probability measure space $(\Sigma^\omega, \mathscr{B}_\Sigma, p)$, let $\mathscr{P}$ denote the space of all probability measures on $\mathscr{B}_\Sigma$. Let $\mathscr{P}^+ \triangleq \{p \in \mathscr{P} : p(x) > 0, \forall x \in \Sigma^\star\}$, which is a proper subset of $\mathscr{P}$. Thus, each element of $\mathscr{P}^+$ is a strictly positive probability measure that assigns a non-zero probability to any string on $\mathscr{B}_\Sigma$.

Definition III.1 (Vector Addition) *The operation of vector addition $\oplus : \mathscr{P}^+ \times \mathscr{P}^+ \to \mathscr{P}^+$ is defined as $p_3 \triangleq p_1 \oplus p_2, \forall p_1, p_2 \in \mathscr{P}^+$ such that*

1) $p_3(\epsilon) = 1$;
2) $\forall x \in \Sigma^\star$ *and* $\tau \in \Sigma$, $\frac{p_3(x\tau)}{p_3(x)} = \frac{p_1(x\tau)p_2(x\tau)}{\sum_{\alpha\in\Sigma} p_1(x\alpha)p_2(x\alpha)}$.
3) *For all countable pairwise disjoint sets* $\{x_i\Sigma^\omega\}$, $p_3(\bigcup_i \{x\Sigma^\omega\}) = \sum_i p_3(x_i)$

In the above equation, p_3 is a probability measure on $\mathscr{P}^+$ because $\Sigma_{\tau\in\Sigma} p_3(x\tau) = \Sigma_{\tau\in\Sigma} \frac{p_1(x\tau)p_2(x\tau)}{\sum_{\alpha\in\Sigma} p_1(x\alpha)p_2(x\alpha)} p_3(x) = p_3(x)\ \forall x \in \Sigma^\star$.

Proposition III.1 $(\mathscr{P}^+, \oplus)$ *forms an Abelian group.*

Proof: The closure and commutativity properties are obvious. Associativity, existence of identity, and existence of the inverse element are established below.

- Associativity:

Following Condition (2) in Definition III.1, it suffices to show $(p_1 \oplus p_2) \oplus p_3 = p_1 \oplus (p_2 \oplus p_3)\ \forall x \in \Sigma^\star$ and $\tau \in \Sigma$.

$$\frac{((p_1 \oplus p_2) \oplus p_3)(x\tau)}{((p_1 \oplus p_2) \oplus p_3)(x)} = \frac{(p_1 \oplus p_2)(x\tau)p_3(x\tau)}{\sum_{\beta\in\Sigma}(p_1 \oplus p_2)(x\beta)p_3(x\beta)}$$
$$= \frac{p_1(x\tau)p_2(x\tau)p_3(x\tau)}{\sum_{\beta\in\Sigma} p_1(x\beta)p_2(x\beta)p_3(x\beta)}$$
$$= \frac{p_1(x\tau)(p_2 \oplus p_3)(x\tau)}{\sum_{\beta\in\Sigma} p_1(x\beta)(p_2 \oplus p_3)(x\beta)}$$
$$= \frac{(p_1 \oplus (p_2 \oplus p_3))(x\tau)}{(p_1 \oplus (p_2 \oplus p_3))(x)}$$

- Existence of identity:

Let a probability measure $\underline{e}$ of symbol strings be defined such that $\underline{e}(x) \triangleq \left(\frac{1}{|\Sigma|}\right)^{|x|}\ \forall x$, where $|x|$ denotes the length of a string $x \in \Sigma^\star$. It follows that $\forall \tau \in \Sigma$, $\frac{\underline{e}(x\tau)}{\underline{e}(x)} = \frac{1}{|\Sigma|}$. Then, for a probability measure $p \in \mathscr{P}^+$ and $\forall \tau \in \Sigma$,

$$\frac{(p \oplus \underline{e})(x\tau)}{(p \oplus \underline{e})(x)} = \frac{p(x\tau)\underline{e}(x\tau)}{\sum_{\alpha\in\Sigma} p(x\alpha)\underline{e}(x\alpha)} = \frac{p(x\tau)\frac{1}{|\Sigma|}}{\frac{1}{|\Sigma|}\sum_{\alpha\in\Sigma} p(x\alpha)}$$
$$= \frac{p(x\tau)}{p(x)}$$

The above relations imply that $p \oplus \underline{e} = \underline{e} \oplus p = p$ by Definition III.1 and by commutativity. Therefore, $\underline{e}$ is the identity of the monoid $(\mathscr{P}^+, \oplus)$.

- Existence of inverse:

$\forall p \in \mathscr{P}^+$, $\forall x \in \Sigma^\star$ and $\forall \tau \in \Sigma$, let a probability measure $-p$ be defined as:

$$(-p)(\epsilon) \triangleq 1 \quad \text{and} \quad \frac{(-p)(x\tau)}{(-p)(x)} \triangleq \frac{p^{-1}(x\tau)}{\sum_{\alpha\in\Sigma} p^{-1}(x\alpha)}$$

where $p^{-1}(x\tau) = \frac{1}{p(x\tau)}$. Then, it follows that

$$\frac{(p \oplus (-p))(x\tau)}{(p \oplus (-p))(x)} = \frac{p(x\tau)(-p)(x\tau)}{\sum_{\alpha\in\Sigma} p(x\alpha)(-p)(x\alpha)}$$
$$= \frac{\frac{p(x\tau)p^{-1}(x\tau)}{\sum_{\beta\in\Sigma} p^{-1}(x\beta)}}{\sum_{\alpha\in\Sigma} p(x\alpha)\frac{p^{-1}(x\alpha)}{\sum_{\beta\in\Sigma} p^{-1}(x\beta)}} = \frac{1}{|\Sigma|}$$

The above expression yields $p \oplus (-p) = \underline{e}$ and hence $(\mathscr{P}^+, \oplus)$ is an Abelian group. ■

In the sequel, the zero-element $\underline{e}$ of the Abelian group $(\mathscr{P}^+, \oplus)$ is denoted as *symbolic white noise*. Next, the scalar multiplication operation is defined over the real field $\mathbb{R}$.

Definition III.2 (Scalar Multiplication) *The scalar multiplication operation $\odot : \mathbb{R} \times \mathscr{P}^+ \to \mathscr{P}^+$ is defined as follows:*

1) $(k \odot p)(\epsilon) = 1$;
2) $\frac{(k\odot p)(x\tau)}{(k\odot p)(x)} = \frac{p^k(x\tau)}{\sum_{\alpha\in\Sigma} p^k(x\alpha)}$
3) *For all countable pairwise disjoint sets* $\{x_i\Sigma^\omega\}$, $(k \odot p)(\bigcup_i \{x\Sigma^\omega\}) = \sum_i (k \odot p)(x_i)$

where $p^k(x\tau) = [p(x\tau)]^k, k \in \mathbb{R}$, $p \in \mathscr{P}^+$, $x \in \Sigma^\star$, *and* $\tau \in \Sigma$.

It follows that $k \odot p$ is a valid probability measure on $\mathscr{P}^+$.

Remark III.1 *By convention, it is assumed that the scalar multiplication operation has a higher precedence than the addition operation. For example,* $k \odot p_1 \oplus p_2$ *implies* $(k \odot p_1) \oplus p_2$.

Theorem III.1 (Main Result) $(\mathscr{P}^+, \oplus, \odot)$ *defines a vector space over the real field* $\mathbb{R}$.

Proof: Let $p, p_1, p_2 \in \mathscr{P}^+$, $k, k_1, k_2 \in \mathbb{R}$, $x \in \Sigma^\star$, and $\tau \in \Sigma$. We check the following equalities:

- It suffices to show that $k \odot p_1 \oplus k \odot p_2 = k \odot (p1 \oplus p_2)$.

$$\frac{(k \odot p_1 \oplus k \odot p_2)(x\tau)}{(k \odot p_1 \oplus k \odot p_2)(x)} = \frac{(k \odot p_1)(x\tau) \cdot (k \odot p_2)(x\tau)}{\Sigma_{\alpha\in\Sigma}\left[(k \odot p_1)(x\alpha) \cdot (k \odot p_2)(x\alpha)\right]}$$
$$= \frac{\frac{p_1^k(x\tau)}{\sum_{\alpha\in\Sigma} p_1^k(x\alpha)} \frac{p_2^k(x\tau)}{\sum_{\alpha\in\Sigma} p_2^k(x\alpha)}}{\sum_{\alpha\in\Sigma} \frac{p_1^k(x\alpha)}{\sum_{\beta\in\Sigma} p_1^k(x\beta)} \frac{p_2^k(x\alpha)}{\sum_{\beta\in\Sigma} p_2^k(x\beta)}}$$
$$= \frac{p_1^k(x\tau) p_2^k(x\tau)}{\sum_{\alpha\in\Sigma} p_1^k(x\alpha) p_2^k(x\alpha)} = \frac{(p_1 \oplus p_2)^k(x\tau)}{\sum_{\alpha\in\Sigma}(p_1 \oplus p_2)^k(x\alpha)}$$
$$= \frac{(k \odot (p_1 \oplus p_2))(x\tau)}{(k \odot (p_1 \oplus p_2))(x)}$$

- It suffices to show that $(k_1 + k_2) \odot p = k_1 \odot p \oplus k_2 \odot p$.

$$\frac{((k_1 + k_2) \odot p)(x\tau)}{((k_1 + k_2) \odot p)(x)} = \frac{p^{k_1+k_2}(x\tau)}{\sum_{\alpha\in\Sigma} p^{k_1+k_2}(x\alpha)}$$
$$= \frac{\frac{p^{k_1}(x\tau)}{\sum_{\gamma\in\Sigma} p^{k_1}(x\gamma)} \frac{p^{k_2}(x\tau)}{\sum_{\gamma\in\Sigma} p^{k_2}(x\gamma)}}{\sum_{\alpha\in\Sigma} \frac{p^{k_1}(x\alpha)}{\sum_{\gamma\in\Sigma} p^{k_1}(x\gamma)} \frac{p^{k_2}(x\alpha)}{\sum_{\gamma\in\Sigma} p^{k_2}(x\gamma)}}$$
$$= \frac{(k_1 \odot p)(x\tau) \cdot (k_2 \odot p)(x\tau)}{\sum_{\alpha\in\Sigma}\left[(k_1 \odot p)(x\alpha) \cdot (k_2 \odot p)(x\alpha)\right]}$$
$$= \frac{(k_1 \odot p \oplus k_2 \odot p)(x\tau)}{(k_1 \odot p \oplus k_2 \odot p)(x)}$$

- It suffices to show that $k_1 \odot (k_2 \odot p) = (k_1 k_2) \odot p$.

$$\frac{(k_1 \odot (k_2 \odot p))(x\tau)}{(k_1 \odot (k_2 \odot p))(x)} = \frac{(k_2 \odot p)^{k_1}(x\tau)}{\sum_{\alpha\in\Sigma}(k_2 \odot p)^{k_1}(x\alpha)}$$
$$= \frac{\left(\frac{p^{k_2}(x\tau)}{\sum_{\beta\in\Sigma} p^{k_2}(x\beta)}\right)^{k_1}}{\sum_{\alpha\in\Sigma}\left(\frac{p^{k_2}(x\alpha)}{\sum_{\beta\in\Sigma} p^{k_2}(x\beta)}\right)^{k_1}}$$
$$= \frac{p^{k_1 k_2}(x\tau)}{\sum_{\beta\in\Sigma} p^{k_1 k_2}(x\beta)} = \frac{((k_1 k_2) \odot p)(x\tau)}{((k_1 k_2) \odot p)(x)}$$

- The equality $1 \odot p = p$ follows from Definition III.2. ■

So far an algebraic structure on $\mathscr{P}^+$ has been established. Now, a topological structure is introduced on a subspace of the vector space $\mathscr{P}^+$ with an appropriate norm.

Definition III.3 (Subspace $\mathscr{P}^+_\infty$) *The subspace* $\mathscr{P}^+_\infty$ *of the vector space* $\mathscr{P}^+$ *is defined as:*

$$\mathscr{P}^+_\infty = \left\{ p \in \mathscr{P}^+ : \sup_{x\in\Sigma^\star} \log\left(\frac{p(x\tau_{max})}{p(x\tau_{min})}\right) < \infty \right\} \quad (1)$$

where $p(x\tau_{max}) \triangleq \max_{\tau\in\Sigma}\{p(x\tau)\}$ *and* $p(x\tau_{min}) \triangleq \min_{\tau\in\Sigma}\{p(x\tau)\}$,

Theorem III.2 *A function* $\|\cdot\| : \mathscr{P}^+_\infty \to [0, \infty)$ *defined as:*

$$\|p\| = \sup_{x\in\Sigma^\star} \log\left(\frac{p(x\tau_{max})}{p(x\tau_{min})}\right) \quad (2)$$

is a norm on the vector space $\mathscr{P}^+_\infty$.

Proof: Let $p \in \mathscr{P}^+_\infty$. The following properties are established:

- Strict positivity: Since $\frac{p(x\tau_{max})}{p(x\tau_{min})} \geq 1$, it follows that $\|p\| \geq 0$. Clearly for the zero element $\underline{e}$, $\frac{\underline{e}(x\tau_{max})}{\underline{e}(x\tau_{min})} = 1$ for all $x \in \Sigma^\star$ and thus $\|\underline{e}\| = 0$.
Conversely, if $\|p\| = 0$ then it forces that $\frac{p(x\tau_{max})}{p(x\tau_{min})} = 1$ for all $x \in \Sigma^\star$.
It follows that $\frac{p(x\tau)}{p(x)} = \frac{1}{|\Sigma|}$ for all $x \in \Sigma^\star$ and $\tau \in \Sigma$. Indeed, $p = \underline{e}$.

- Homogeneity: A non-negative real k preserves the order of $p(x\tau)$ for any fixed x and a negative real k reverses the order. Therefore, for $k \geq 0$,

$$\|k \odot p\| = \sup_{x\in\Sigma^\star} \log\left(\frac{(k \odot p)(x\tau_{max})}{(k \odot p)(x\tau_{min})}\right) = \sup_{x\in\Sigma^\star} \log\left(\frac{p(x\tau_{max})}{p(x\tau_{min})}\right)^k = |k| \cdot \|p\|$$

and for $k < 0$,

$$\|k \odot p\| = \sup_{x\in\Sigma^\star} \log\left(\frac{(k \odot p)(x\tau_{max})}{(k \odot p)(x\tau_{min})}\right) = \sup_{x\in\Sigma^\star} \log\left(\frac{p(x\tau_{min})}{p(x\tau_{max})}\right)^k = \sup_{x\in\Sigma^\star} \log\left(\frac{p(x\tau_{max})}{p(x\tau_{min})}\right)^{-k} = |k| \cdot \|p\|$$

- Triangular inequality:

$$\|p_1 \oplus p_2\| = \sup_{x\in\Sigma^\star} \log\left(\frac{(p_1 \oplus p_2)(x\tau_{max})}{(p_1 \oplus p_2)(x\tau_{min})}\right) \leq \sup_{x,y\in\Sigma^\star} \log\left(\frac{p_1(x\tau_{max}) p_2(y\tau_{max})}{p_1(x\tau_{min}) p_2(y\tau_{min})}\right) = \|p_1\| + \|p_2\|$$

The proof is now complete. ■

It follows from Theorem 4 in [6] that a probability measure $p \in \mathscr{P}^+_\infty$ can be encoded into a PFSA if and only if the probabilistic Nerode equivalence $\mathcal{N}_p$ is of finite index. Let $\mathscr{P}^+_f$ denote the set of positive probability measures, which has only finitely many Nerode equivalence classes.

IV. PFSA Space and Probability Measure Space

There is a close relationship between the space of probabilistic finite state automata (PFSA) and the space of probability measures on $\mathscr{B}_\Sigma$. A concept of equivalence is introduced in this section to establish a bijection between these two spaces to admit similar vector space structures.

Following Definition II.1, let $\widetilde{\mathscr{A}} \triangleq \{G = (Q, \Sigma, \delta, q_0, \tilde{\pi}) : \tilde{\pi}(q, \sigma) > 0$ for all $q \in Q$ and all $\sigma \in \Sigma\}$.

Definition IV.1 (Mapping of PFSA [6]) *Let a PFSA $G = (Q, \Sigma, \delta, q_0, \tilde{\pi}) \in \widetilde{\mathscr{A}}$ and the associated probability measure $p \in \mathscr{P}_f^+$. Then, a map $\widetilde{\mathbb{H}} : \widetilde{\mathscr{A}} \to \mathscr{P}_f^+$ is defined as $\widetilde{\mathbb{H}}(G) = p$ such that*

$$p(x) = \tilde{\pi}(q_0, \sigma_1) \prod_{k=0}^{r-1} \tilde{\pi}(\delta^\star(q_0, \sigma_0\sigma_1 \cdots \sigma_k), \sigma_{k+1}) \tag{3}$$

where $x \triangleq \sigma_1 \cdots \sigma_r \in \Sigma^\star$ is a symbol string of length $r \in \mathbb{N}$ and the set $\{\sigma_0\}$ signifies the empty string ϵ.

Algorithm 1 Construction of PFSA G from the probability measure p associated with the measurable space $(\Sigma^\omega, \mathscr{B}_\Sigma)$

Input: $(\Sigma^\omega, \mathscr{B}_\Sigma, p)$ such that $\mathcal{N}_p$ is of finite index $n \in \mathbb{N}$;
Output: G;
Let $Q = \{q_j : j \in \{1, \cdots, n\}\}$ be the set of equivalence classes of the relation $\mathcal{N}_p$;
Set the initial state of G as $q_o \in Q$ such that the null string ϵ belongs to the equivalence class q_o;
for each $q_j \in Q$ **do**
  Pick an arbitrary string $x \in q_j$;
  for each $\sigma \in \Sigma$ **do**
    if $x\sigma \in q_k$ **then**
      Set $\delta(q_j, \sigma) = q_k$;
      Set $\tilde{\pi}(q_j, \sigma) = \frac{p(x\sigma)}{p(x)}$;
    end if
  end for
end for

Algorithm 1 is constructed in the context of the probabilistic Nerode equivalence $\mathcal{N}_p$ (see Definition II.3) such that the map $\widetilde{\mathbb{H}} : \widetilde{\mathscr{A}} \to \mathscr{P}_f^+$ in Definition IV.1 is surjective. However, $\widetilde{\mathbb{H}}$ may not be injective as explained below.

There may exist different PFSA realizations that encode the same probability measure on $\mathscr{B}_\Sigma$ due to two reasons: (i) non-minimal realization and (ii) state relabeling. In this respect, the equivalence of two PFSA is addressed as follows.

Definition IV.2 (PFSA equivalence) *Two PFSA $\widetilde{G}$ and G are said to be equivalent if the associated probabilities are equal, i.e., if $\widetilde{\mathbb{H}}(\widetilde{G}) = \widetilde{\mathbb{H}}(G)$. The equivalence class of G is denoted as $\Xi(G) \triangleq \{\widetilde{G} \in \mathscr{A} : \widetilde{\mathbb{H}}(\widetilde{G}) = \widetilde{\mathbb{H}}(G)\}$.*

Definition IV.2 is interpreted as follows. There is no distinction among the PFSA that encode the same probability measure on $\mathscr{B}_\Sigma$, because it is not important how the measure is encoded but what the measure itself is. This formulation of equivalence classes bears a direct analogy with that of the L_p space, where for each $f \in L_p$, the vector f means the equivalence class of f defined as $\{g \in L_p : f = g$ a.e.$\}$.

Next we define the quotient space $\mathscr{A} \triangleq \widetilde{\mathscr{A}}/\Xi$ and the associated quotient map

$$\mathbb{H} : \mathscr{A} \longrightarrow \mathscr{P}_f^+ \tag{4}$$

is well defined because any member in the equivalence class $\Xi(G)$ yields the same function value. The equivalence class $\Xi(G)$ is simply denoted as G for brevity in the sequel.

The quotient map $\mathbb{H} : \mathscr{A} \longrightarrow \mathscr{P}_f^+$ in Eq. (4) is a bijection, because the original map $\widetilde{\mathbb{H}} : \widetilde{\mathscr{A}} \to \mathscr{P}_f^+$ is constructed to be surjective (see Algorithm 1) and the quotient map is naturally injective. The bijection $\mathbb{H}$ permits a similar definition of a normed vector space on $\mathscr{A}$. In this setting, the inverse of the map $\mathbb{H}$ is denoted as:

$$\mathbb{F} : \mathscr{P}_f^+ \to \mathscr{A} \quad \text{such that } \mathbb{H}\mathbb{F} = \mathcal{I}d : \mathscr{P}_f^+ \to \mathscr{P}_f^+ \tag{5}$$

where $\mathcal{I}d$ is the identity map. Note that the inverse map $\mathbb{F}$ is generated by Algorithm 1.

Definition IV.3 (Perfect Encoding) *Given an alphabet Σ, a PFSA $G = (Q, \Sigma, \delta, q_0, \tilde{\pi})$ is said to be a perfect encoding of the measure space $(\Sigma^\omega, \mathscr{B}_\Sigma, p)$ if $p = \mathbb{H}(G)$.*

Next we present Proposition IV.1 and Corollary IV.1, which together show that $\mathscr{P}_f^+$ is a subspace of the normed space $(\mathscr{P}_\infty^+, \|\cdot\|)$ (See Theorem III.2).

Proposition IV.1 *Let $p_1, p_2 \in \mathscr{P}^+$ and $x, y \in \Sigma^\star$.*

1) *If $z_1, z_2 \in [x]_{p_1} \cap [y]_{p_2}$, then $z_1 \mathcal{N}_{p_1 \oplus p_2} z_2$;*
2) *If $z_1, z_2 \in [x]_{p_1}$ and $k \in \mathbb{R}$, then $z_1 \mathcal{N}_{k \odot p_1} z_2$*

where $[x]_p \triangleq \{z \in \Sigma^\star : x\mathcal{N}_p z\}$.

Proof: Let $u_n = \tau_1\tau_2 \ldots \tau_n \in \Sigma^\star$ and $p_3 = p_1 \oplus p_2$ where $\tau_i \in \Sigma$. For Eq. (1), it will be proven that $\frac{p_3(z_1 u_n)}{p_3(z_1)} = \frac{p_3(z_2 u_n)}{p_3(z_2)}$ for any $u_n \in \Sigma^\star$. This can be achieved by induction.

$$\begin{aligned}
\frac{p_3(z_1 u_1)}{p_3(z_1)} &= \frac{p_1(z_1 u_1) p_2(z_1 u_1)}{\sum_{\alpha \in \Sigma} p_1(z_1 \alpha) p_2(z_1 \alpha)} \\
&= \frac{\frac{p_1(z_1 u_1)}{p_1(z_1)} \frac{p_2(z_1 u_1)}{p_2(z_1)}}{\sum_{\alpha \in \Sigma} \frac{p_1(z_1 \alpha)}{p_1(z_1)} \frac{p_2(z_1 \alpha)}{p_2(z_1)}} \\
&= \frac{\frac{p_1(z_2 u_1)}{p_1(z_2)} \frac{p_2(z_2 u_1)}{p_2(z_2)}}{\sum_{\alpha \in \Sigma} \frac{p_1(z_2 \alpha)}{p_1(z_2)} \frac{p_2(z_2 \alpha)}{p_2(z_2)}} = \frac{p_3(z_2 u_1)}{p_3(z_2)}
\end{aligned}$$

Now for the inductive step,

$$\begin{aligned}
\frac{p_3(z_1 u_{n+1})}{p_3(z_1)} &= \frac{p_3(z_1 u_n)}{p_3(z_1)} \frac{p_3(z_1 u_{n+1})}{p_3(z_1 u_n)} \\
&= \frac{p_3(z_2 u_n)}{p_3(z_2)} \frac{p_3(z_2 u_{n+1})}{p_3(z_2 u_n)} = \frac{p_3(z_2 u_{n+1})}{p_3(z_2)}
\end{aligned}$$

The second identity is derived in the same way. ∎

Corollary IV.1 *$\mathscr{P}_f^+$ is a subspace of the normed space $(\mathscr{P}_\infty^+, \|\cdot\|)$ over $\mathbb{R}$ with $\oplus$ and $\odot$ being the vector addition and scalar multiplication operations, respectively.*

Proof: Let $k \in \mathbb{R}$ and let the probability measures $p_1, p_2 \in \mathscr{P}_f^+$ induce M_1 and M_2 equivalence classes, respectively, where $M_1, M_2 \in \mathbb{N}$ (see Remark II.1). From Proposition IV.1, it follows that $p_1 \oplus p_2$ has at most $M_1 \cdot M_2$ equivalence classes and $k \odot p_1$ has at most M_1 equivalence classes. Hence the closure property is satisfied.

Next it is shown that $\|\cdot\|$ is a valid norm. Since every $p \in \mathscr{P}_f^+$ has only finitely many equivalence classes, the supremum in Eq. (III.3) is replaced by its maximum:

$$\|p\| = \max_{x\in\Sigma^\star} \log\left(\frac{p(x\tau_{max})}{p(x\tau_{min})}\right) \tag{6}$$

as there are only finitely many values $\frac{p(x\tau_{max})}{p(x\tau_{min})}$ can take. It is obvious that $\|p\| < \infty, \forall p \in \mathscr{P}_f^+$. Now, by virtue of Theorem III.2 with the maximum instead of the supremum, $\|\cdot\|$ is indeed a valid norm on $\mathscr{P}_f^+$. ■

So far we have established the vector space $(\mathscr{P}_f^+, \oplus, \odot)$. By use of the bijection $\mathbb{H}$ and its inverse $\mathbb{F}$, we introduce new vector addition and scalar multiplication operations on $\mathscr{A}$.

Definition IV.4 (Vector space $\mathscr{A}$) *Let $G_1, G_2 \in \mathscr{A}$ and $k \in \mathbb{R}$. Then,*

- *The addition operation $+ : \mathscr{A} \times \mathscr{A} \to \mathscr{A}$ is defined as*

$$G_1 + G_2 = \mathbb{F}(\mathbb{H}(G_1) \oplus \mathbb{H}(G_2))$$

- *The scalar multiplication operation $\cdot : \mathbb{R} \times \mathscr{A} \to \mathscr{A}$ is defined as*

$$k \cdot G_1 = \mathbb{F}(k \odot (\mathbb{H}(G_1))$$

Remark IV.1 *It follows from the Definition IV.4 that $\mathbb{H}(G_1+G_2) = \mathbb{H}(G_1)\oplus\mathbb{H}(G_2)$ and $\mathbb{H}(k\cdot G_1) = k\odot\mathbb{H}(G_1)$. This construction implies that the bijection $\mathbb{H}$ is linear and hence the bijection $\mathbb{H}$ becomes an isomorphism between the vector spaces $(\mathscr{P}_f^+, \oplus, \odot)$ and $(\mathscr{A}, +, \cdot)$.*

A norm on the vector space $\mathscr{A}$ is defined in the same way.

Definition IV.5 (Normed Space $\mathscr{A}$) *The norm $\|\cdot\|_A$ on the vector space $\mathscr{A}$ is defined as*

$$\|G\|_A = \|\mathbb{H}(G)\| \qquad \forall G \in \mathscr{A}$$

This makes the quotient map $\mathbb{H} : \mathscr{A} \longrightarrow \mathscr{P}_f^+$ in Eq. (4) an isometric isomorphism between the two normed spaces.

For brevity, the scalar multiplication $\cdot$ is omitted in the expression. Again, by using the same precedence rule as before, multiplication takes precedence over addition. For example, $kG_1 + G_2$ implies $(k \cdot G_1) + G_2$ rather than $k \cdot (G_1 + G_2)$. Furthermore, $(-1) \cdot G$ is denoted by $-G$.

Definition IV.4 does not provide an efficient way of computing the algebraic operations. An alternative means is explored to express these operations in terms of PFSA only.

Definition IV.6 (Structural Similarity) *Two PFSA $G_i = (Q_i, \Sigma, \delta_i, q_0^i, \tilde{\pi}_i) \in \mathscr{A}$, $i = \{1, 2\}$, are said to have the same structure if $Q_1 = Q_2$, $q_0^1 = q_0^2$ and $\delta_1(q,\sigma) = \delta_2(q,\sigma)$ $\forall q \in Q_1$ and $\forall \sigma \in \Sigma$.*

Proposition IV.2 *If two PFSA $G_1, G_2 \in \mathscr{A}$ are of the same structure, i.e., $G_i = (Q, \Sigma, \delta, q_0, \tilde{\pi}_i), i = \{1, 2\}$, then $G_1 + G_2 = (Q, \Sigma, \delta, q_0, \tilde{\pi})$ where*

$$\tilde{\pi}(q,\sigma) = \frac{\tilde{\pi}_1(q,\sigma)\tilde{\pi}_2(q,\sigma)}{\sum_{\alpha\in\Sigma}\tilde{\pi}_1(q,\alpha)\tilde{\pi}_2(q,\alpha)} \tag{7}$$

Proof: Denoting $p_i = \mathbb{H}(G_i)$, $i = \{1, 2\}$, since G_1 and G_2 have the same structure, it follows from Eq. (3) that

$$\frac{p_i(x\sigma)}{p_i(x)} = \tilde{\pi}_i(\delta^\star(q_0, x), \sigma) = \tilde{\pi}_i(q, \sigma)$$

for all string x in state $q \in Q$ and all $\sigma \in \Sigma$.
Now, by Definition III.1 and Definition IV.1,

$$\begin{aligned}\tilde{\pi}(q,\sigma) &= \frac{(p_1 \oplus p_2)(x\sigma)}{(p_1 \oplus p_2)(x)} = \frac{p_1(x\sigma)p_2(x\sigma)}{\sum_{\alpha\in\Sigma} p_1(x\alpha)p_2(x\alpha)} \\ &= \frac{\frac{p_1(x\sigma)p_2(x\sigma)}{p_1(x)p_2(x)}}{\sum_{\alpha\in\Sigma}\frac{p_1(x\alpha)p_2(x\alpha)}{p_1(x)p_2(x)}} = \frac{\tilde{\pi}_1(q,\sigma)\tilde{\pi}_2(q,\sigma)}{\sum_{\alpha\in\Sigma}\tilde{\pi}_1(q,\alpha)\tilde{\pi}_2(q,\alpha)}\end{aligned}$$

The proof is now complete. ■

Definition IV.7 (Synchronous Composition) *The binary operation of synchronous composition of two PFSA $G_i = (Q_i, \Sigma, \delta, q_0^{(i)}, \tilde{\pi}_i) \in \mathscr{A}$ where $i = 1, 2$, denoted by $\otimes : \mathscr{A} \times \mathscr{A} \to \mathscr{A}$ is defined as:*

$$G_1 \otimes G_2 = (Q_1 \times Q_2, \Sigma, \delta', (q_0^{(1)}, q_0^{(2)}), \tilde{\pi}')$$

where $\forall q_i \in Q_1$ $\forall q_j \in Q_2$ $\forall \sigma \in \Sigma$,

$$\delta'((q_i, q_j), \sigma) = (\delta_1(q_i,\sigma), \delta_2(q_j,\sigma)) \quad \text{and}$$

$$\tilde{\pi}'((q_i, q_j), \sigma) = \tilde{\pi}_1(q_i, \sigma)$$

Proposition IV.3 *If $G_1, G_2 \in \mathscr{A}$, then $\mathbb{H}(G_1) = \mathbb{H}(G_1 \otimes G_2)$ and therefore $G_1 = G_1 \otimes G_2$.*

Proof: See Theorem 4.5 in [6]. ■

Proposition IV.4 *$G_1 \otimes G_2$ and $G_2 \otimes G_1$ have the same structure up to state relabeling.*

Proof: Let us define a function $T : Q \times Q \to Q \times Q$ such that $T(p, q) = (q, p)$. It is noted that T is a bijection. ■

Theorem IV.1 *Given two PFSA $G = (Q, \Sigma, \delta, q_0, \tilde{\pi}) \in \mathscr{A}$, $\widetilde{G} \in \mathscr{A}$ and $k \in \mathbb{R}$. Then*

1) *$G + \widetilde{G}$ can be computed via Proposition IV.2 and Definition IV.7 as follows*

$$G + \widetilde{G} = G \otimes \widetilde{G} + \widetilde{G} \otimes G \tag{8}$$

2) *$kG = (Q, \Sigma, \delta, q_0, \tilde{\pi}')$ where*

$$\tilde{\pi}'(q,\sigma) = \frac{(\tilde{\pi}(q,\sigma))^k}{\sum_{\alpha\in\Sigma}(\tilde{\pi}(q,\alpha))^k} \tag{9}$$

for all $q \in Q$ and $\sigma \in \Sigma$.

3) *The norm of E is*

$$\|G\|_A = \max_{q \in Q} \log \left(\frac{\max_{\sigma_1 \in \Sigma} \tilde{\pi}(q, \sigma_1)}{\min_{\sigma_2 \in \Sigma} \tilde{\pi}(q, \sigma_2)} \right) \quad (10)$$

Proof:

1) It follows from Proposition IV.4 and Proposition IV.3 that

$$\mathbb{H}(G + \widetilde{G}) = \mathbb{H}(G) \oplus \mathbb{H}(\widetilde{G})$$
$$= \mathbb{H}(G \otimes \widetilde{G}) \oplus \mathbb{H}(\widetilde{G} \otimes G) = \mathbb{H}(G \otimes \widetilde{G} + \widetilde{G} \otimes G)$$

2) By Proposition IV.1, scalar multiplication by k does not change the structure of G and therefore the transition function δ and the start state also remain unchanged. Denoting $p = \mathbb{H}(G)$, it follows from Eq. (3) that

$$\frac{p(x\sigma)}{p(x)} = \tilde{\pi}(\delta^\star(q_0, x), \sigma) = \tilde{\pi}(q, \sigma)$$

for all string x in state $q \in Q$ and all $\sigma \in \Sigma$. By Definition III.2, it follows that

$$\tilde{\pi}'(q, \sigma) = \frac{k \odot p(x\sigma)}{k \odot p(x)} = \frac{p^k(x\sigma)}{\sum_{\alpha \in \Sigma} p^k(x\alpha)}$$
$$= \frac{\frac{p^k(x\sigma)}{p^k(x)}}{\sum_{\alpha \in \Sigma} \frac{p^k(x\alpha)}{p^k(x)}} = \frac{(\tilde{\pi}(q, \sigma))^k}{\sum_{\alpha \in \Sigma} (\tilde{\pi}(q, \alpha))^k}$$

3) It directly follows from Definitions III.3 and IV.1.

The proof is now complete. ∎

V. Interpretation of Algebraic Operations

The probabilistic finite state automata (PFSA), constructed from an alphabet Σ, are regarded as semantic models of the underlying physical process. Interpretations of the algebraic operations in the vector space of PFSA are presented below.

The vector sum $p_1 \oplus p_2$ of two probability measures p_1 and p_2 increases the probability of the symbol strings that are most likely to occur in both p_1 and p_2. If p_1 and p_2 are probability distributions of positively (negatively) correlated processes, then the distribution $p_1 \oplus p_2$ approaches a delta (uniform) distribution. In the extreme case, if $p_1 = -p_2$ (i.e., perfect negative correlation), then their vector addition exactly yields $\underline{e}$ that represents the uniform distribution. Hence, the norm of the sum of two measures p_1 and p_2 reflects the correlation between these two measures. The zero element $\underline{e}$, called *symbolic white noise*, in the vector space $\mathscr{P}_f^+$ corresponds to the uniform distribution on $\mathscr{B}_\Sigma$ and is perfectly encoded by the PFSA $E \in \mathscr{A}$, expressed as:

$$E = \mathbb{F}(\underline{e}) = \{\{q\}, \Sigma, \delta, \{q\}, \tilde{\pi}\}$$

where $\delta(q, \sigma) = q$ and $\tilde{\pi}(q, \sigma) = \frac{1}{|\Sigma|}, \forall \sigma \in \Sigma$.

Every string of the same length has equal probability of occurrence in the PFSA E that has only one state, where the symbols occur independently of each other and have equal probability of occurrence. The knowledge of the history does not provide any information for predicting the future of any symbol sequence generated by E. Thus, E is viewed as a semantic model for *symbolic white noise* in a dynamical system, because no additional information is provided through vector addition of E to any PFSA.

Scalar multiplication relates to reshaping the probability distribution $p \in \mathscr{P}_f^+$ on $\mathscr{B}_\Sigma$. For example, multiplication by $k > 0$ alters the probability assigned to strings in the sense that a string with a higher probability will now have a higher probability and vice versa for $k < 0$. As $k \to +\infty$ or $k \to -\infty$, the distribution p approaches the delta distribution; similarly, as $k \to 0$, the distribution p approaches $\underline{e}$ that is the uniform distribution. In Definition III.3, the uniform distribution yields a zero norm, while PFSA whose distributions are close to the delta distribution would have their respective norms close to infinity. With the increase of k, $\|k \odot p\|$ is a non-decreasing function of k. The norm provides a uniform bound on the deviation from the uniform distribution to quantify the non-probabilistic behavior of a PFSA.

VI. Summary and Conclusions

This paper, which is the first of two parts, constructs a vector space for a class of probabilistic finite state automata (PFSA) in the measure-theoretic setting. The operations of vector addition and scalar multiplication are introduced by establishing an isomorphism between the space of probability measures and the quotient space of PFSA relative to a specified equivalence relation. This isomorphism is made isometric by constructing appropriate norms on the respective vector spaces. This mathematical framework is motivated by various applications in dynamical systems modeling, analysis, and control in the stochastic setting. Significance of the algebraic operations is interpreted in terms of the semantic models of the underlying process as the vector space of PFSA. The second part [8] of this two-part paper constructs a family of inner products for model identification and order reduction.

References

[1] R. Carrasco and J. Oncina, "Learning stochastic regular grammars by means of a state merging method," *Proceedings of Second International Colloquium on Grammatical Inference and Applications, Alicante, Spain*, pp. 139–152, September 21-23, 1994.

[2] E. Vidal, F. Thollard, C. de la Higuera, F. Casacuberta, and R. Carrasco, "Probabilistic finite-state machines – Parts I and II," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 7, pp. 1013–1039, 2005.

[3] A. Ray, "Symbolic dynamic analysis of complex systems for anomaly detection," *Signal Processing*, vol. 84, no. 7, pp. 1115–1130, 2004.

[4] A. Ray, "Signed real measure of regular languages for discrete event supervisory control," *International Journal of Control*, vol. 78, no. 12, pp. 949–967, 2005.

[5] T. Barfoot and G. D'Eleuterio, "An algebra for the control of stochastic systems: Exercises in linear algebra," *Proceedings of the 5th International Conference on Dynamics and Control of Structures in Space (DCSS), Cambridge, England*, pp. 14–18, 2002.

[6] I. Chattopadhyay and A. Ray, "Stuctural transformations of probabilistic finite state machines," *International Journal of Control*, vol. 81, no. 5, pp. 820–835, 2008.

[7] W. Rudin, *Real and Complex Analysis, 3rd edition*. McGraw-Hill, New York, NY, USA, 1987.

[8] Y. Wen, A. Ray, I. Chattopadhyay, and S. Phoha, "Modeling of symbolic systems: Part II - Hilbert space construction for model identification and order reduction," in *2011 American Control Conference, San Francisco, CA, USA*, June-July.