



ELSEVIER

Contents lists available at SciVerse ScienceDirect

## Journal of Computer and System Sciences

www.elsevier.com/locate/jcss

Vector space formulation of probabilistic finite state automata<sup>☆</sup>Yicheng Wen<sup>1</sup>, Asok Ray<sup>\*</sup>

College of Engineering, Pennsylvania State University, University Park, PA 16802, United States

## ARTICLE INFO

## Article history:

Received 16 January 2010

Received in revised form 21 December 2011

Accepted 7 February 2012

Available online 10 February 2012

## Keywords:

Probabilistic finite state automata

Normed vector space

Pattern recognition

## ABSTRACT

This paper develops a vector space model of a class of probabilistic finite state automata (PFSA) that are constructed from finite-length symbol sequences. The vector space is constructed over the real field, where the algebraic operations of vector addition and the associated scalar multiplication operations are defined on a probability measure space, and implications of these algebraic operations are interpreted. The zero element of this vector space is semantically equivalent to a PFSA, referred to as *symbolic white noise*. A norm is introduced on the vector space of PFSA, which provides a measure of the information content. An application example is presented in the framework of pattern recognition for identification of robot motion in a laboratory environment.

© 2012 Elsevier Inc. All rights reserved.

## 1. Introduction

Probabilistic finite state automata (PFSA) have emerged as a tool for modeling and analysis of uncertain dynamical systems in a variety of applications such as anomaly detection [1], pattern recognition [2], and decision & control [3]; in this respect, symbolic-model-based techniques have been developed for modeling, analysis, and control of dynamical systems by several researchers (e.g., [4]). The key feature of the work reported in this paper is formal language-theoretic in the sense that symbolic modeling is used instead of classical continuous-domain modeling. The proposed approach makes use of symbolic dynamic filtering (SDF) [1] that partitions the (possibly pre-processed) time series or image data observed from the underlying system to generate a symbol string. Then, semantic models are constructed in the symbolic domain.

Many finite state machine models have been reported in the literature [5], such as probabilistic finite state automata (PFSA), hidden Markov models (HMM) [6,7], stochastic regular grammars [8], Markov chains [9], just to name a few. The rationale of having the PFSA structure of a semantic model is that, in general, PFSA are easier to learn in practice, although PFSA may not always be as powerful as other models like HMM [5]. For example, experimental results [10] show that the usage of a PFSA structure could make learning of a pronunciation model for spoken words to be 10–100 times faster than a corresponding HMM, and yet the performance of PFSA is slightly better. This advantage leads to a very wide application of PFSA in many areas such as pattern classification [2,11]. Among several PFSA construction algorithms reported in the literature, causal-state splitting reconstruction (CSSR) [12] assumes no a priori structure and computes optimal representations ( $\epsilon$ -machine) in sense of mutual information. In contrast,  $D$ -Markov [1] construction starts with an a priori known structure that can be freely chosen.

<sup>☆</sup> This work has been supported in part by the US Office of Naval Research under Grant No. N00014-09-1-0688, and by the US Army Research Laboratory and the US Army Research Office under Grant No. W911NF-07-1-0376.

<sup>\*</sup> Corresponding author.

E-mail addresses: ywen2000@gmail.com (Y. Wen), axr2@psu.edu (A. Ray).

<sup>1</sup> Currently with General Electric, Louisville, KY, United States.

### Nomenclature

$\Sigma$	Alphabet of symbols with finite cardinality $ \Sigma $	$\mathcal{P}_\infty^+$	The set of strictly positive probability measures with finite norm
$\Sigma^*$	The set of all finite-length strings on $\Sigma$	$\mathcal{A}$	The set of all PFSA over alphabet $\Sigma$
$\Sigma^\omega$	The set of all strictly infinite-length strings on $\Sigma$	$\mathcal{A}^+$	The set of all PFSA with strictly positive probability morph function
$\epsilon$	Null string with $ \epsilon  = 0$	$\mathbb{H}$	Mapping from all PFSA to probability measures
$Q$	Set of state of PFSA	$\mathbb{H}_{-1}$	Mapping from probability measures to PFSA
$\delta$	State transition function of PFSA	$\mathbb{H}^+$	Restriction of the map $\mathbb{H}$ on $\mathcal{A}^+$
$\delta^*$	Extended state transition function of PFSA	$\mathbb{H}_{-1}^+$	Restriction of the map $\mathbb{H}_{-1}$ on $\mathcal{P}_f^+$
$\tilde{\pi}$	Probability morph function of PFSA	$\cong_{-1}$	Equivalence relation under $\mathbb{H}$
$\tilde{\Pi}$	Probability morph matrix of PFSA	$+$	Addition operation of PFSA
$\mathcal{B}_\Sigma$	The smallest $\sigma$ -algebra on the set $\{x \in \Sigma^\omega \text{ where } x \in \Sigma^*\}$	$\oplus$	Addition operation of two probability measures on $\mathcal{B}_\Sigma$
$\mathcal{N}_p$	Probabilistic Nerode equivalence on a probability space $(\Sigma^\omega, \mathcal{B}_\Sigma, p)$	$\odot$	Scalar multiplication operation of a probability measure on $\mathcal{B}_\Sigma$ over $\mathbb{R}$
$[x]_p$	Probabilistic Nerode equivalence class of a string $x$ on $\Sigma^*$	$\cdot$	Scalar multiplication operation $+$ of PFSA over $\mathbb{R}$
$\mathcal{P}$	The set of all probability measures on $\mathcal{B}_\Sigma$	$\otimes$	Synchronous composition of PFSA
$\mathcal{P}^+$	The set of all strictly positive probability measures on $\mathcal{B}_\Sigma$	$\underline{e}$	Symbolic white noise
$\mathcal{P}_f^+$	The set of strictly positive probability measures having finite probabilistic Nerode equivalence classes	$\ \cdot\ $	Norm on $\mathcal{P}_\infty^+$
		$\ \cdot\ _{\mathcal{A}^+}$	Norm on $\mathcal{A}^+$
		$h(\cdot)$	Entropy rate
		$KL(\cdot\ \cdot)$	K-L divergence

In general, it would be desirable to be able to treat PFSA models or stochastic languages as vectors in a normed vector space over the real field  $\mathbb{R}$  for signal processing, pattern recognition, and decision & control applications. For example, in pattern recognition, if PFSA are used as feature vectors (e.g., [11]), then the lack of a precise mathematical structure on the feature space (i.e., the space of PFSA) prevents the employment of classical machine learning algorithms [13,14]. Similarly, for information fusion [15], PFSA models can be used to compress the information derived from sensor time series, but the critical issue is how to fuse the information from heterogeneous sources, which requires mathematical operations on PFSA. 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 [16]. Barfoot and D'Eleuterio [17] 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. Apparently, no other prior work exists for defining a mathematical structure in the vector space of PFSA.

This vector space formalism enriches the current theory for PFSA by taking into account disparate automaton structures and probability measures as well as enhances the understanding of stochastic regular languages for solving the associated problems involving symbolic dynamics [18]. From these perspectives, the major contributions of the paper are delineated below.

1. The formulation of a normed vector space representation over the real field  $\mathbb{R}$  for a class of PFSA that are constructed from finite-length symbol sequences.
2. The experimental validation of the above normed vector space representation in the framework of pattern recognition for identification of mobile robots.

The paper is organized in seven sections. Section 2 presents the background information on the formal language theory and measure theory. The vector space on probability measures is constructed in Section 3. Section 4 establishes an isomorphism between the space of probability measures and the space of PFSA, defines new operations on PFSA, and formulates a method to compute those operations in terms of PFSA. Section 5 presents a physical interpretation of the algebraic operations in the vector space of PFSA. Section 6 experimentally validates some of the underlying concepts in a laboratory environment for identification of robot behavior patterns. Section 7 summarizes and concludes the paper with recommendations for future research.

## 2. Background information

This section summarizes the mathematical preliminaries in the formal language theory and measure theory, which are necessary to develop the theory of vector space representation of probabilistic finite state automata (PFSA).

### 2.1. Preliminaries on formal language theory

In the formal language theory [19], an alphabet  $\Sigma$  is a (non-empty finite) set of symbols, i.e., the alphabet's cardinality  $|\Sigma| \in \mathbb{N}$ . A string  $x$  over  $\Sigma$  is a finite-length sequence of symbols in  $\Sigma$ . The length of a string  $x$ , denoted by  $|x|$ , represents the number of symbols in  $x$ . The Kleene closure of  $\Sigma$ , denoted by  $\Sigma^*$ , is the set of all finite-length strings of events including the null string  $\epsilon$ ; cardinality of  $\Sigma^*$  is countable. The set of all strictly infinite-length strings is denoted as  $\Sigma^\omega$ ; cardinality of  $\Sigma^\omega$  is uncountable. The string  $xy$  is called concatenation of  $x$  and  $y$ . It is obvious that a null string  $\epsilon$  is an identity for concatenation.

**Definition 2.1 (PFSA).** A probabilistic finite state automaton (PFSA) is a tuple  $G = (Q, \Sigma, \delta, q_0, \tilde{\pi})$ , where the underlying finite state automaton (FSA) is assumed to be deterministic and complete.

- $Q$  is a (non-empty) finite set, called set of states.
- $\Sigma$  is a (non-empty) finite set, called input alphabet.
- $\delta : Q \times \Sigma \rightarrow Q$  is the state transition function. The transition map  $\delta$  naturally induces an extended transition function  $\delta^* : Q \times \Sigma^* \rightarrow Q$  such that  $\delta^*(q, \epsilon) = q$  and  $\delta^*(q, x\tau) = \delta(\delta^*(q, x), \tau)$  for  $q \in Q, x \in \Sigma^*$  and  $\tau \in \Sigma$ .
- $q_0 \in Q$  is the start state. All states are assumed to be reachable from the start state; otherwise, the non-reachable states could be removed from  $Q$ .
- $\tilde{\pi} : Q \times \Sigma \rightarrow [0, 1]$  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 probability morph function  $\tilde{\pi}$  is represented in a matrix form as  $\tilde{\Pi}$  with the element  $\tilde{\Pi}_{ij} \triangleq \tilde{\pi}(q_i, \sigma_j)$ , where  $q_i \in Q$  and  $\sigma_j \in \Sigma$ ,

where it is mandated that  $\tilde{\pi}$  and  $\delta$  are compatible with each other, i.e., if  $\delta(q, \sigma)$  is defined, then  $\tilde{\pi}(q, \sigma) > 0$  and vice versa.

In this paper, we consider a specific class of PFSA, namely the complete PFSA, whose state transition function  $\delta$  is a total function. This is equivalent to the probability morph matrix  $\tilde{\Pi}$  in Definition 2.1 being elementwise strictly positive. As explained below, the rationale for imposing this restriction is availability of only finite-length experimental (e.g., time series or image) data for generation of symbol strings.

For a finite-length symbol sequence  $\mathbb{S}$  over an alphabet  $\Sigma$ , there exist several PFSA construction algorithms (e.g., [12,1]) to discover the underlying PFSA model  $G$ . These algorithms start with identifying the structure of  $G$ , i.e.,  $(Q, \Sigma, \delta, q_0)$ . Then, a  $|Q| \times |\Sigma|$  count matrix  $C$  is initialized with each of its elements to 1. Let  $N_{ij}$  denote the number of times a symbol  $\sigma_j$  is emanated from the state  $q_i$  by observation of the symbol sequence  $\mathbb{S}$ . In this way, the estimated morph matrix for the PFSA  $G$  is computed as

$$\tilde{\Pi}_{ij} \triangleq \frac{C_{ij}}{\sum_k C_{ik}} = \frac{1 + N_{ik}}{|\Sigma| + N_{ik}}. \tag{1}$$

The rationale for initializing all elements of  $C$  to 1 is that if a state  $q_i$  is never encountered by observing the finitely many symbols in  $\mathbb{S}$ , then there should be no preference to any specific symbols emanating from  $q_i$ . Therefore, it is logical to initialize  $\tilde{\Pi}_{ij} = \frac{1}{|\Sigma|}$ , i.e., the uniform distribution for the  $i$ -th row of the morph matrix  $\tilde{\Pi}$ . (It is shown later in the paper that a morph matrix with all elements equal to  $\frac{1}{|\Sigma|}$  serves as the zero element in the vector space and is referred to as the *symbolic white noise*.) The count matrix  $C$  is updated as more symbols are observed from the symbol sequence  $\mathbb{S}$ . This procedure guarantees that each element of the morph matrix  $\tilde{\Pi}$  is strictly positive for any finite-length symbol sequence  $\mathbb{S}$ .

### 2.2. Preliminaries on measure theory

Relevant definitions in the measure theory are recalled in this subsection.

**Definition 2.2 ( $\sigma$ -Algebra).** (See [20].) A collection  $\mathfrak{M}$  of subsets of a non-empty set  $X$  is said to be a  $\sigma$ -algebra in  $X$  if  $\mathfrak{M}$  has the following properties:

1.  $X \in \mathfrak{M}$ .
2. If  $A \in \mathfrak{M}$ , then  $A^c \in \mathfrak{M}$  where  $A^c$  is the complement of  $A$  relative to  $X$ , i.e.,  $A^c = X \setminus A$ .
3. If  $A_n \in \mathfrak{M}$  for  $n \in \mathbb{N}$ , then  $\bigcup_{n=1}^\infty A_n \in \mathfrak{M}$ .

**Definition 2.3** (Measure). (See [20].) A finite (non-negative) measure is a countably additive function  $\mu$ , defined on a  $\sigma$ -algebra  $\mathfrak{M}$ , whose range is  $[0, K]$  for some  $K \in \mathbb{R}$ . Countable additivity means that if  $\{A_i\}$  is a pairwise disjoint countable collection of members of  $\mathfrak{M}$ , then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \quad (2)$$

**Definition 2.4** (Probability measure space). A probability measure on a non-empty set in a specified  $\sigma$ -algebra  $\mathfrak{M}$  is a finite non-negative measure on  $\mathfrak{M}$ . Although not required by the theory, a probability measure is defined on the unit interval  $[0, 1]$  as its range. A probability measure space is a triple  $(X, \mathfrak{M}, p)$  where  $X$  is a non-empty set,  $\mathfrak{M}$  is a  $\sigma$ -algebra in  $X$  and  $p$  is a finite non-negative measure on  $\mathfrak{M}$ .

**Definition 2.5** (Probability measure space for PFSA). Given an alphabet  $\Sigma$ , the set  $\mathcal{B}_\Sigma \triangleq 2^{\Sigma^*} \Sigma^\omega$  is defined to be the  $\sigma$ -algebra generated by the set  $\{L: L = x\Sigma^\omega \text{ where } x \in \Sigma^*\}$ , i.e., the smallest  $\sigma$ -algebra on the set  $\Sigma^\omega$ , which contains the set  $\{L: L = x\Sigma^\omega \text{ where } x \in \Sigma^*\}$ .

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

**Definition 2.6** (Probabilistic Nerode relation). Given an alphabet  $\Sigma$ , any two strings  $x, y \in \Sigma^*$  are said to satisfy the probabilistic Nerode relation  $\mathcal{N}_p$  on a probability space  $(\Sigma^\omega, \mathcal{B}_\Sigma, p)$ , denoted by  $x\mathcal{N}_p y$ , if

$$p(u|x) = p(u|y), \quad \forall u \in \Sigma^*, \quad (3)$$

where the conditional probability  $p(u|x)$  is defined as  $p(u|x) \triangleq \frac{p(xu)}{p(x)}$ .

It has been proven in [21] that the probabilistic Nerode relation defined above is a right-invariant equivalence relation. In the sequel, this is referred to as probabilistic Nerode equivalence and is denoted as the probabilistic Nerode equivalence class of a string  $x$  on  $\Sigma^*$  by  $[x]_p$ , i.e.,  $[x]_p = \{z \in \Sigma^*: x\mathcal{N}_p z\}$ . That is, two probabilistic Nerode equivalent strings  $x$  and  $y$  have the same probability distribution in the future.

**Remark 2.1.** The probabilistic Nerode equivalence of a measure  $p$  induces a partition of strings on  $\Sigma^*$ . In general, such a partition of  $\Sigma^*$  could consist of infinitely many equivalence classes but there must be finitely many equivalence classes for a probability measure that is encoded by a PFSA as described later in this paper.

### 3. Vector space of probability measures

Given the probability measure space  $(\Sigma^\omega, \mathcal{B}_\Sigma, p)$ , let  $\mathcal{P}$  denote the space of all probability measures on  $\mathcal{B}_\Sigma$ . Let  $\mathcal{P}^+ \triangleq \{p \in \mathcal{P}: p(x) > 0, \forall x \in \Sigma^*\}$ , which is a proper subset of  $\mathcal{P}$ . Each element of  $\mathcal{P}^+$  is a strictly positive probability measure that assigns a non-zero probability to any string on  $\mathcal{B}_\Sigma$ .  $|\Sigma|$  is the cardinality of the alphabet  $\Sigma$ , i.e. the number of symbols in  $\Sigma$ .

**Definition 3.1** (Vector addition). The addition operation  $\oplus: \mathcal{P}^+ \times \mathcal{P}^+ \rightarrow \mathcal{P}^+$  is defined by  $p_3 \triangleq p_1 \oplus p_2$ ,  $\forall p_1, p_2 \in \mathcal{P}^+$  such that

$$\begin{aligned} 1. & p_3(\epsilon) = 1; \\ 2. & \forall x \in \Sigma^* \text{ and } \tau \in \Sigma, p_3(\tau|x) = \frac{p_1(\tau|x)p_2(\tau|x)}{\sum_{\alpha \in \Sigma} p_1(\alpha|x)p_2(\alpha|x)}. \end{aligned}$$

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

**Proposition 3.1.**  $(\mathcal{P}^+, \oplus)$  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:**

To show  $(p_1 \oplus p_2) \oplus p_3 = p_1 \oplus (p_2 \oplus p_3)$ ,  $\forall x \in \Sigma^*$  and  $\tau \in \Sigma$ , we proceed as:

$$\begin{aligned} ((p_1 \oplus p_2) \oplus p_3)(\tau|x) &= \frac{(p_1 \oplus p_2)(\tau|x)p_3(\tau|x)}{\sum_{\beta \in \Sigma} (p_1 \oplus p_2)(\beta|x)p_3(\beta|x)} \\ &= \frac{\sum_{\alpha \in \Sigma} \frac{p_1(\tau|x)p_2(\tau|x)}{p_1(\alpha|x)p_2(\alpha|x)} p_3(\tau|x)}{\sum_{\beta \in \Sigma} \frac{p_1(\beta|x)p_2(\beta|x)}{\sum_{\alpha \in \Sigma} p_1(\alpha|x)p_2(\alpha|x)} p_3(\beta|x)} = \frac{p_1(\tau|x)p_2(\tau|x)p_3(\tau|x)}{\sum_{\beta \in \Sigma} p_1(\beta|x)p_2(\beta|x)p_3(\beta|x)} \\ &= \frac{p_1(\tau|x) \sum_{\alpha \in \Sigma} \frac{p_2(\tau|x)p_3(\tau|x)}{p_2(\alpha|x)p_3(\alpha|x)}}{\sum_{\beta \in \Sigma} p_1(\beta|x) \frac{p_2(\beta|x)p_3(\beta|x)}{\sum_{\alpha \in \Sigma} p_2(\alpha|x)p_3(\alpha|x)}} = \frac{p_1(\tau|x)(p_2 \oplus p_3)(\tau|x)}{\sum_{\beta \in \Sigma} p_1(\beta|x)(p_2 \oplus p_3)(\beta|x)} \\ &= (p_1 \oplus (p_2 \oplus p_3))(\tau|x). \end{aligned}$$

• **Existence of identity:**

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

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

The above relations imply that  $p \oplus \underline{e} = \underline{e} \oplus p = p$  by Definition 3.1 and by commutativity. Therefore,  $\underline{e}$  is the identity of the monoid  $(\mathcal{P}^+, \oplus)$ .

• **Existence of inverse:**

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

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

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

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

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

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

**Definition 3.2 (Scalar multiplication).** The scalar multiplication operation  $\odot : \mathbb{R} \times \mathcal{P}^+ \rightarrow \mathcal{P}^+$  is defined as follows:

1.  $(k \odot p)(\epsilon) = 1$ ;
2.  $(k \odot p)(\tau|x) = \frac{p^k(\tau|x)}{\sum_{\alpha \in \Sigma} p^k(\alpha|x)}$

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

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

**Remark 3.1.** By convention, 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 3.1.**  $(\mathcal{P}^+, \oplus, \odot)$  defines a vector space over the real field  $\mathbb{R}$ .

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

- To show that  $k \odot p_1 \oplus k \odot p_2 = k \odot (p_1 \oplus p_2)$ , we proceed as:

$$\begin{aligned}
 (k \odot p_1 \oplus k \odot p_2)(\tau|x) &= \frac{(k \odot p_1)(\tau|x) \cdot (k \odot p_2)(\tau|x)}{\sum_{\alpha \in \Sigma} [(k \odot p_1)(\alpha|x) \cdot (k \odot p_2)(\alpha|x)]} \\
 &= \frac{\frac{p_1^k(\tau|x)}{\sum_{\alpha \in \Sigma} p_1^k(\alpha|x)} \frac{p_2^k(\tau|x)}{\sum_{\alpha \in \Sigma} p_2^k(\alpha|x)}}{\sum_{\alpha \in \Sigma} \frac{p_1^k(\alpha|x)}{\sum_{\beta \in \Sigma} p_1^k(\beta|x)} \frac{p_2^k(\alpha|x)}{\sum_{\beta \in \Sigma} p_2^k(\beta|x)}} = \frac{p_1^k(\tau|x) p_2^k(\tau|x)}{\sum_{\alpha \in \Sigma} p_1^k(\alpha|x) p_2^k(\alpha|x)} \\
 &= \frac{(p_1 \oplus p_2)^k(\tau|x)}{\sum_{\alpha \in \Sigma} (p_1 \oplus p_2)^k(\alpha|x)} = (k \odot (p_1 \oplus p_2))(\tau|x).
 \end{aligned}$$

- To show that  $(k_1 + k_2) \odot p = k_1 \odot p \oplus k_2 \odot p$ , we proceed as:

$$\begin{aligned}
 ((k_1 + k_2) \odot p)(\tau|x) &= \frac{p^{k_1+k_2}(\tau|x)}{\sum_{\alpha \in \Sigma} p^{k_1+k_2}(\alpha|x)} = \frac{\frac{p^{k_1}(\tau|x)}{\sum_{\gamma \in \Sigma} p^{k_1}(\gamma|x)} \frac{p^{k_2}(\tau|x)}{\sum_{\gamma \in \Sigma} p^{k_2}(\gamma|x)}}{\sum_{\alpha \in \Sigma} \frac{p^{k_1}(\alpha|x)}{\sum_{\gamma \in \Sigma} p^{k_1}(\gamma|x)} \frac{p^{k_2}(\alpha|x)}{\sum_{\gamma \in \Sigma} p^{k_2}(\gamma|x)}} \\
 &= \frac{(k_1 \odot p)(\tau|x) \cdot (k_2 \odot p)(\tau|x)}{\sum_{\alpha \in \Sigma} [(k_1 \odot p)(\alpha|x) \cdot (k_2 \odot p)(\alpha|x)]} = (k_1 \odot p \oplus k_2 \odot p)(\tau|x).
 \end{aligned}$$

- To show that  $k_1 \odot (k_2 \odot p) = (k_1 k_2) \odot p$ , we proceed as:

$$\begin{aligned}
 (k_1 \odot (k_2 \odot p))(\tau|x) &= \frac{(k_2 \odot p)^{k_1}(\tau|x)}{\sum_{\alpha \in \Sigma} (k_2 \odot p)^{k_1}(\alpha|x)} = \frac{\left(\frac{p^{k_2}(\tau|x)}{\sum_{\beta \in \Sigma} p^{k_2}(\beta|x)}\right)^{k_1}}{\sum_{\alpha \in \Sigma} \left(\frac{p^{k_2}(\alpha|x)}{\sum_{\beta \in \Sigma} p^{k_2}(\beta|x)}\right)^{k_1}} \\
 &= \frac{p^{k_1 k_2}(\tau|x)}{\sum_{\beta \in \Sigma} p^{k_1 k_2}(\beta|x)} = ((k_1 k_2) \odot p)(\tau|x).
 \end{aligned}$$

- The equality  $1 \odot p = p$  follows directly from Definition 3.2.  $\square$

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

**Definition 3.3** (Subspace  $\mathcal{P}_\infty^+$ ). The subspace  $\mathcal{P}_\infty^+$  of the vector space  $\mathcal{P}^+$  is defined as:

$$\mathcal{P}_\infty^+ = \left\{ p \in \mathcal{P}^+ : \sup_{x \in \Sigma^*, \tau \in \Sigma} p(\tau|x) < 1 \right\}. \tag{4}$$

**Theorem 3.2.** The function  $\|\cdot\| : \mathcal{P}_\infty^+ \rightarrow [0, \infty)$  defined as

$$\|p\| \triangleq \sup_{x \in \Sigma^*} \log \left( \frac{p(\tau_{\max}|x)}{p(\tau_{\min}|x)} \right) \tag{5}$$

is a norm on the vector space  $\mathcal{P}_\infty^+$ , where

$$p(\tau_{\max}|x) \triangleq \max_{\tau \in \Sigma} p(\tau|x) \quad \text{and} \quad p(\tau_{\min}|x) \triangleq \min_{\tau \in \Sigma} p(\tau|x). \tag{6}$$

**Proof.** Let  $p \in \mathcal{P}_\infty^+$ . First, it is easy to see that Eq. (4) guarantees the function is well defined on  $\mathcal{P}_\infty^+$ . Then the following properties are established:

- **Strict positivity:**  
 Since  $\frac{p(\tau_{\max}|x)}{p(\tau_{\min}|x)} \geq 1$ , it follows that  $\|p\| \geq 0$ . Clearly for the zero element  $\mathfrak{e}$ ,  $\frac{\mathfrak{e}(\tau_{\max}|x)}{\mathfrak{e}(\tau_{\min}|x)} = 1$  for all  $x \in \Sigma^*$  and thus  $\|\mathfrak{e}\| = 0$ .  
 Conversely, if  $\|p\| = 0$  then it forces that  $\frac{p(\tau_{\max}|x)}{p(\tau_{\min}|x)} = 1$  for all  $x \in \Sigma^*$ .  
 It follows that  $p(\tau|x) = \frac{1}{|\Sigma|}$  for all  $x \in \Sigma^*$  and  $\tau \in \Sigma$ . Indeed,  $p = \mathfrak{e}$ .
- **Homogeneity:**  
 A non-negative  $k$  preserves the order of  $p(x\tau)$  for any fixed  $x$  and a negative  $k$  reverses the order. Therefore, for  $k \geq 0$ ,

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

and for  $k < 0$ ,

$$\begin{aligned} \|k \odot p\| &= \sup_{x \in \Sigma^*} \log \left( \frac{(k \odot p)(\tau_{\max}|x)}{(k \odot p)(\tau_{\min}|x)} \right) = \sup_{x \in \Sigma^*} \log \left( \frac{p(\tau_{\min}|x)}{p(\tau_{\max}|x)} \right)^k \\ &= \sup_{x \in \Sigma^*} \log \left( \frac{p(\tau_{\max}|x)}{p(\tau_{\min}|x)} \right)^{-k} = |k| \cdot \|p\|. \end{aligned}$$

• Triangular inequality:

$$\begin{aligned} \|p_1 \oplus p_2\| &= \sup_{x \in \Sigma^*} \log \left( \frac{(p_1 \oplus p_2)(\tau_{\max}|x)}{(p_1 \oplus p_2)(\tau_{\min}|x)} \right) \leq \sup_{x, y \in \Sigma^*} \log \left( \frac{p_1(\tau_{\max}|x)p_2(\tau_{\max}|y)}{p_1(\tau_{\min}|x)p_2(\tau_{\min}|y)} \right) \\ &= \|p_1\| + \|p_2\|. \end{aligned}$$

The proof is now complete. □

Note that in Eq. (5), there may exist more than one  $\tau$  that achieves  $p(\tau_{\max}|x)$  for each  $x$ .

**Remark 3.2.** Convergence of sequences in a normed space directly follows from the fact that the norm  $\|\cdot\|$  serves as a metric. For example, as  $n \rightarrow \infty$ , the sequence  $\{\frac{1}{n} \odot p\}$  converges to  $\underline{e}$ , the symbolic white noise, under this metrizable topology, because

$$\left\| \frac{1}{n} \odot p - \underline{e} \right\| = \frac{1}{n} \|p\| \rightarrow 0 \quad \text{as } n \rightarrow \infty. \tag{7}$$

On the other hand, the sequence  $\{n \odot p\}$  does not converge since  $\|n \odot p\| \rightarrow \infty$ . To see what the final distribution  $p^*$  looks like, we compute  $\forall x$  and  $\forall \tau \in \Sigma$ ,

$$\frac{(n \odot p)(\tau|x)}{(n \odot p)(\tau_{\max}|x)} = \left( \frac{p(\tau|x)}{p(\tau_{\max}|x)} \right)^n \rightarrow \begin{cases} 0 & \text{if } p(\tau|x) < p(\tau_{\max}|x), \\ 1 & \text{if } p(\tau|x) = p(\tau_{\max}|x). \end{cases} \tag{8}$$

In other words,

$$p^*(\tau|x) = \begin{cases} 1/N(x) & \text{if } \tau \text{ achieves } p(\tau_{\max}|x), \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

where  $N(x)$  is the number of  $\tau$ 's that achieve  $p(\tau_{\max}|x)$ . However, the final distribution  $p^*$  is not an element of  $\mathcal{P}_\infty^+$ ; it is not even an element of  $\mathcal{P}^+$ .

#### 4. Relationship between PFSA and probability measures

There is a close relationship between probabilistic finite state automata (PFSA) and probability measures on  $\mathcal{B}_\Sigma$ . A concept of equivalence is introduced in this section to establish an isomorphism between these two spaces. Let the space of PFSA be denoted by  $\mathcal{A}$ .

**Definition 4.1** (*Mapping of PFSA*). (See [21].) Let  $G = (Q, \Sigma, \delta, q_0, \tilde{\pi}) \in \mathcal{A}$  and  $p \in \mathcal{P}$ . Let a map  $\mathbb{H} : \mathcal{A} \rightarrow \mathcal{P}$  be defined as  $\mathbb{H}(G) = p$  such that

$$p(x) = \tilde{\pi}(q_0, \sigma_1) \prod_{k=1}^{r-1} \tilde{\pi}(\delta^*(q_0, \sigma_1 \cdots \sigma_k), \sigma_{k+1}) \tag{10}$$

where the event string  $x = \sigma_1 \cdots \sigma_r \in \Sigma^*$ , where  $r \in \mathbb{N}$ , the set of positive integers, and probability morph function  $\tilde{\pi} : Q \times \Sigma \rightarrow [0, 1]$ . Then, the PFSA  $G$  is said to be a perfect encoding of the measure space  $(\Sigma^\omega, \mathcal{B}_\Sigma, p)$ .

Conversely, however, there are many PFSA realizations (up to an isomorphism of state labeling) that encode the same probability measure on  $\mathcal{B}_\Sigma$  due to non-minimal realization. For example, in Fig. 1 where the alphabet  $\Sigma = \{a, b\}$ , the two PFSA essentially encode the same measure on  $\mathcal{B}_\Sigma$  but clearly have different representations. The right one is a non-minimal realization of the left top one since the states  $q_1$  and  $q_2$  are the same. They can be combined to obtain the left one.

**Definition 4.2** (*PFSA equivalence*). Two PFSA  $G_1$  and  $G_2$  are said to be equivalent if  $\mathbb{H}(G_1) = \mathbb{H}(G_2)$ . This equivalence relation is denoted as:  $G_1 \cong G_2$ .

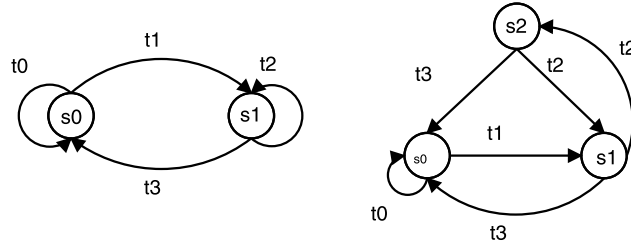


Fig. 1. Different PFSA realizations of the same probability measure.

**Remark 4.1.** Definition 4.2 is interpreted as follows. There is no distinction among the PFSA that encode the same probability measure on  $\mathcal{B}_\Sigma$  because it is not important how the measure is encoded but what the measure itself is. In the sequel, the notation  $[G]$  is used to represent the equivalence class of  $G$ ; more explicitly,  $[G] \triangleq \{P \in \mathcal{A} : P \cong G\}$ . In the sequel,  $[G]$  is denoted simply as  $G$  for brevity, similar to what is done to (equality almost everywhere) equivalence classes of functions in  $L_p$ -spaces [20].

**Definition 4.3** (Map  $\mathbb{H}_{-1}$ ). Given a probability measure  $(\Sigma^\omega, \mathcal{B}_\Sigma, p)$  with a finite number of Nerode equivalence classes (see Definition 2.6), a PFSA can be represented as:

- Let  $Q = \{q_j : j \in \mathbb{N}\}$  be the set of Nerode equivalence classes;
- The initial state  $q_0 = [\varepsilon]_p \in Q$ ;
- The state transition  $\delta$  is  $\delta(q_i, \sigma) = [x\sigma]_p, \forall \sigma \in \Sigma$  where  $x \in q_i$ ;
- The morph  $\tilde{\pi}(q_i, \sigma) = p(\sigma|x)$  with any  $x \in q_i$ .

This procedure is denoted as  $\mathbb{H}_{-1} : \mathcal{P} \rightarrow \mathcal{A}$ .

With a restriction to minimal PFSA,  $\mathbb{H}_{-1}$  becomes the inverse of the map  $\mathbb{H}$ . It follows from Definitions 3.3 and 4.3 that any  $p \in \mathcal{P}_\infty^+$  can be encoded into a PFSA if and only if the probabilistic Nerode equivalence  $\mathcal{N}_p$  is of finite index. Let  $\mathcal{P}_f^+$  denote the set of probability measures, which has only finitely many Nerode equivalence classes, and  $\mathcal{P}_f^+ \subset \mathcal{P}^+$ .

Next we present Proposition 4.1 and Corollary 4.1 to show that  $\mathcal{P}_f^+$  is a subspace of the normed space  $(\mathcal{P}_\infty^+, \|\cdot\|)$ .

**Proposition 4.1.** Let  $p_1, p_2 \in \mathcal{P}^+$  and  $x, y \in \Sigma^*$ .

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

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

$$p_3(u_1|z_1) = \frac{p_1(u_1|z_1)p_2(u_1|z_1)}{\sum_{\alpha \in \Sigma} p_1(\alpha|z_1)p_2(\alpha|z_1)} = \frac{p_1(u_1|z_2)p_2(u_1|z_2)}{\sum_{\alpha \in \Sigma} p_1(\alpha|z_2)p_2(\alpha|z_2)} = p_3(u_1|z_2).$$

Now for the inductive step,

$$p_3(u_{n+1}|z_1) = p_3(u_n|z_1)p_3(\tau_{n+1}|z_1 u_n) = p_3(u_n|z_2)p_3(\tau_{n+1}|z_2 u_n) = p_3(u_{n+1}|z_2).$$

The second identity can be derived exactly the same way.  $\square$

**Corollary 4.1.**  $\mathcal{P}_f^+$  is a subspace of the normed space  $(\mathcal{P}_\infty^+, \|\cdot\|)$  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  $p_1, p_2 \in \mathcal{P}_f^+$  induce  $M_1$  and  $M_2$  equivalence classes, respectively, where  $M_1, M_2 \in \mathbb{N}$  (see Remark 2.1). From Proposition 4.1, 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  $p \in \mathcal{P}_f^+$  has only finitely many equivalence classes, the supremum in Eq. (5) is replaced by its maximum:

$$\|p\| = \max_{x \in \Sigma^*} \log \left( \frac{p(\tau_{\max}|x)}{p(\tau_{\min}|x)} \right) \tag{11}$$

as there are only finitely many values  $\frac{p(\tau_{\max}|x)}{p(\tau_{\min}|x)}$  can take. It is obvious that  $\|p\| < \infty, \forall p \in \mathcal{P}_f^+$ . Now, by virtue of Theorem 3.2 with the maximum instead of the supremum,  $\|\cdot\|$  is indeed a valid norm on  $\mathcal{P}_f^+$ .  $\square$

Since  $\mathcal{A}^+ \triangleq \{G = (Q, \Sigma, \delta, q_0, \tilde{\pi}) : \tilde{\pi}(q, \sigma) > 0 \text{ for all } q \in Q \text{ and all } \sigma \in \Sigma\}$ , is a proper subset of  $\mathcal{A}$ , it follows that the transition map of any PFSA in the subset  $\mathcal{A}^+$  is a total function. Let the map  $\mathbb{H} : \mathcal{A} \rightarrow \mathcal{P}$  be restricted on a smaller domain  $\mathcal{A}^+$ , that is,  $\mathbb{H}^+ : \mathcal{A}^+ \rightarrow \mathcal{P}_f^+$  with  $\mathbb{H}^+ = \mathbb{H}|_{\mathcal{A}^+}$ . Similarly,  $\mathbb{H}_{-1}$  is restricted on  $\mathcal{P}_f^+$ , i.e.  $\mathbb{H}_{-1}^+ = \mathbb{H}_{-1}|_{\mathcal{P}_f^+}$ . Since the map  $\mathbb{H}^+$  is bijective, we can define it to be an isometric isomorphism between the two spaces  $\mathcal{A}^+$  and  $\mathcal{P}_f^+$  so that a normed vector space structure can be established on  $\mathcal{A}^+$ .

**Definition 4.4** (Isometric isomorphism). Let  $G_1, G_2 \in \mathcal{P}_f^+$  and  $k \in \mathbb{R}$ . Then:

- The addition operation  $+: \mathcal{A}^+ \times \mathcal{A}^+ \rightarrow \mathcal{A}^+$  is defined as

$$G_1 + G_2 = \mathbb{H}_{-1}^+(\mathbb{H}^+(G_1) \oplus \mathbb{H}^+(G_2)).$$

- The scalar multiplication operation  $\cdot: \mathbb{R} \times \mathcal{A}^+ \rightarrow \mathcal{A}^+$  is defined as

$$k \cdot G_1 = \mathbb{H}_{-1}^+(k \odot (\mathbb{H}^+(G_1))).$$

- The norm  $\|\cdot\|_{\mathcal{A}^+}$  on the vector space  $\mathcal{A}^+$  is defined as

$$\|G\|_{\mathcal{A}^+} = \|\mathbb{H}^+(G)\|, \quad \forall G \in \mathcal{A}^+.$$

**Remark 4.2.** By adopting the same rule as in the space  $(\mathcal{P}^+, \oplus, \odot)$  (see Remark 3.1), the multiplication operation  $\cdot$  takes precedence over the addition operation  $+$  in the space  $(\mathcal{A}^+, +, \cdot)$ . For example,  $k \cdot G_1 + G_2$  means  $(k \cdot G_1) + G_2$  rather than  $k \cdot (G_1 + G_2)$ . Furthermore,  $(-1) \cdot G$  is denoted by  $-G$ . For brevity, the scalar multiplication  $\cdot$  is omitted in the sequel.

Definition 4.4 does not give any efficient ways of computing the operations. An alternative means is presented below to express those operations in terms of PFSA only.

**Definition 4.5** (Similar PFSA structure). Two PFSA  $G_i = (Q_i, \Sigma, \delta_i, q_0^i, \tilde{\pi}_i) \in \mathcal{A}^+, i = 1, 2$ , are defined 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 4.2.** If two PFSA  $G_1, G_2 \in \mathcal{A}^+$  are of the same structure, i.e.  $G_i = (Q, \Sigma, \delta, q_0, \tilde{\pi}_i), i = \{1, 2\}$ , then it follows that  $G_1 + G_2 \cong (Q, \Sigma, \delta, q_0, \tilde{\pi})$  where

$$\tilde{\pi}(q, \sigma) = \frac{\tilde{\pi}_1(q, \sigma)\tilde{\pi}_2(q, \sigma)}{\sum_{\alpha \in \Sigma} \tilde{\pi}_1(q, \alpha)\tilde{\pi}_2(q, \alpha)}. \tag{12}$$

**Proof.** Denoting by  $p_i = \mathbb{H}^+(G_i), i = \{1, 2\}$ , since  $G_1$  and  $G_2$  have the same structure from Eq. (10), it follows that

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

for all string  $x$  in state  $q \in Q$  and all  $\sigma \in \Sigma$ .

Now, by Definitions 3.1 and 4.1,

$$\tilde{\pi}(q, \sigma) = (p_1 \oplus p_2)(\sigma|x) = \frac{p_1(\sigma|x)p_2(\sigma|x)}{\sum_{\alpha \in \Sigma} p_1(\alpha|x)p_2(\alpha|x)} = \frac{\tilde{\pi}_1(q, \sigma)\tilde{\pi}_2(q, \sigma)}{\sum_{\alpha \in \Sigma} \tilde{\pi}_1(q, \alpha)\tilde{\pi}_2(q, \alpha)}.$$

The proof is now complete.  $\square$

**Definition 4.6** (Synchronous composition). The binary operation of synchronous composition of two PFSA  $G_i = (Q_i, \Sigma, \delta, q_0^{(i)}, \tilde{\pi}_i) \in \mathcal{A}^+$  where  $i = 1, 2$ , denoted by  $\otimes: \mathcal{A}^+ \times \mathcal{A}^+ \rightarrow \mathcal{A}^+$ , is defined as:

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

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

$$\delta'((q_i, q_j), \sigma) = (\delta_1(q_i, \sigma), \delta_2(q_j, \sigma)) \quad \text{and} \quad \tilde{\pi}'((q_i, q_j), \sigma) = \tilde{\pi}_1(q_i, \sigma).$$

**Proposition 4.3.** *If  $G_1, G_2 \in \mathcal{A}^+$ , then  $\mathbb{H}^+(G_1) = \mathbb{H}^+(G_1 \otimes G_2)$  and therefore  $G_1 \cong G_1 \otimes G_2$ . Similarly,  $G_2 \cong G_2 \otimes G_1$ .*

**Proof.** See Theorem 4.5 in [21].  $\square$

**Remark 4.3.** It is clear that  $G_1 \otimes G_2$  and  $G_2 \otimes G_1$  have the same structure up to the state relabeling.

**Theorem 4.1.** *Given two PFSA  $G = (Q, \Sigma, \delta, q_0, \tilde{\pi}) \in \mathcal{A}^+$ ,  $R \in \mathcal{A}^+$  and  $k \in \mathbb{R}$ . Then*

1.  $G + R$  can be computed via Proposition 4.2 and Definition 4.6 as follows

$$G + R \cong G \otimes R + R \otimes G. \quad (13)$$

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

$$\tilde{\pi}'(q, \sigma) = \frac{(\tilde{\pi}(q, \sigma))^k}{\sum_{\alpha \in \Sigma} (\tilde{\pi}(q, \alpha))^k} \quad (14)$$

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

3. The norm of  $G$  is

$$\|G\|_{\mathcal{A}^+} = \max_{q \in Q} \log \left( \frac{\max_{\sigma_1 \in \Sigma} \tilde{\pi}(q, \sigma_1)}{\min_{\sigma_2 \in \Sigma} \tilde{\pi}(q, \sigma_2)} \right). \quad (15)$$

**Proof.**

1. It follows from Propositions 4.3 and Remark 4.3 that

$$\begin{aligned} \mathbb{H}^+(G + R) &= \mathbb{H}^+(G) \oplus \mathbb{H}^+(R) \\ &= \mathbb{H}^+(G \otimes R) \oplus \mathbb{H}^+(R \otimes G) = \mathbb{H}^+(G \otimes R + R \otimes G). \end{aligned}$$

2. By Proposition 4.1, scalar multiplication by  $k$  does not change the structure of  $G$  and therefore the transition function  $\delta$  and the start state also remain unchanged. Denoting  $p = \mathbb{H}^+(G)$ , it follows from Eq. (10) that

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

for all string  $x$  in state  $q \in Q$  and all  $\sigma \in \Sigma$ . By Definition 3.2, it follows that

$$\begin{aligned} \tilde{\pi}'(q, \sigma) &= (k \odot p)(\sigma|x) = \frac{p^k(\sigma|x)}{\sum_{\alpha \in \Sigma} p^k(\alpha|x)} = \frac{p^k(\sigma|x)}{\sum_{\alpha \in \Sigma} p^k(\alpha|x)} \\ &= \frac{(\tilde{\pi}(q, \sigma))^k}{\sum_{\alpha \in \Sigma} (\tilde{\pi}(q, \alpha))^k}. \end{aligned}$$

3. It directly follows from Definitions 3.3 and 4.1.

The proof is now complete.  $\square$

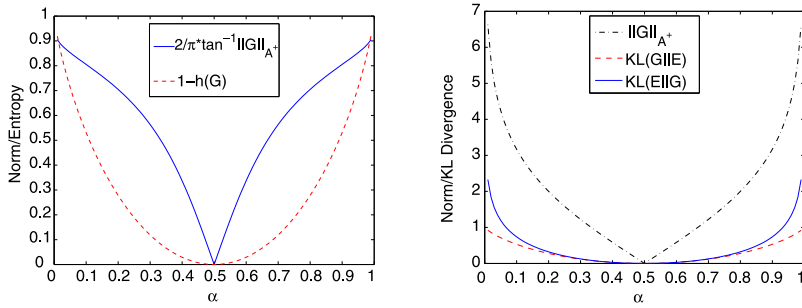
## 5. Interpretation of algebraic operations

The probabilistic finite state automata, constructed from a given alphabet  $\Sigma$ , are regarded as semantic models of the underlying physical process. The zero element  $\underline{e}$ , called *symbolic white noise*, in the vector space  $\mathcal{P}^+$  corresponds to the uniform distribution on  $\mathcal{B}_\Sigma$  and is perfectly encoded by the PFSA  $E \in \mathcal{A}^+$ , expressed as:

$$E = \mathbb{H}_{-1}^+(\underline{e}) = \{ \{q\}, \Sigma, \delta, \{q\}, \tilde{\pi} \}$$

where  $\delta(q, q) = q$  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 are independent 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



(a) Comparison of entropy and norm of an i.i.d process  $G$  (b) Comparison of  $\|G\|_{\mathcal{A}^+}$ ,  $KL(G\|E)$ , and  $KL(E\|G)$  versus the parameter  $\alpha$  when  $E$  is the *Symbolic White Noise*

**Fig. 2.** Comparison of norm and metric with quantities in Information Theory.

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  $p \in \mathcal{P}_f^+$  on  $\mathcal{B}_\Sigma$ . Multiplication by  $k \in \mathbb{R}$  alters the probability assigned to strings in the sense that original higher probable strings may now have lower probabilities and vice versa. As  $k \rightarrow +\infty$ , the distribution  $p$  approaches the delta distribution; similarly, as  $k \rightarrow 0$ , the distribution  $p$  approaches  $\underline{e}$  that is the uniform distribution. In Eq. (5), 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.

The vector sum  $p_1 \oplus p_2$  of two probability measures  $p_1$  and  $p_2$  increases the probability of the strings that are most likely to occur in both  $p_1$  and  $p_2$ . If  $p_1$  and  $p_2$  are distributions of positively (negatively) correlated processes, then the distribution  $p_1 \oplus p_2$  approaches a delta (uniform) distribution. In the extreme case, when  $p_1$  and  $p_2$  are negatively correlated, i.e.  $p_1 = -p_2$ , 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 norm of PFSA, defined in Eq. (15), is now interpreted by analogy with the entropy rate of the PFSA  $G$  in information theory [22] as:

$$h(G) = - \sum_{q \in Q} \wp(q) \left[ \sum_{\sigma \in \Sigma} \tilde{\pi}(q, \sigma) \log \tilde{\pi}(q, \sigma) \right]. \tag{16}$$

While the norm of PFSA in Eq. (15) is the maximum log-ratio over the states, the entropy rate in Eq. (16) is the expectation over the entropy of the states. Considering an independent and identically distributed (i.i.d.) process, namely, a single-state PFSA  $G$  over the binary alphabet  $\Sigma = \{a, b\}$ , let the probabilities of generating the symbols  $a$  and  $b$  be  $\alpha$  and  $(1 - \alpha)$ , respectively, with  $\alpha \in (0, 1)$ .

Fig. 2(a) compares  $(1 - h(G))$  (dashed line) and  $\frac{2}{\pi} \tan^{-1}(\|G\|_{\mathcal{A}^+})$  (solid line), where the range of entropy rate  $h(G)$  is  $[0, 1]$  and the range of the norm  $\|G\|_{\mathcal{A}^+}$  is converted from  $[0, \infty)$  to  $[0, 1]$  by homeomorphism via the arc tangent function. It is observed that the profiles of  $(1 - h(G))$  and  $\frac{2}{\pi} \tan^{-1}(\|G\|_{\mathcal{A}^+})$  are qualitatively similar. Hence, by drawing an analogy, it is possible to interpret the norm in Eq. (15) as a measure of certainty or information contained in the PFSA  $G$ .

Let two processes be represented by PFSA  $\tilde{G}$  and  $G$ , whose probability mass functions are  $\tilde{P}$  and  $P$ , respectively. Then, a diversity between  $\tilde{G}$  and  $G$  is defined as the K-L diversity [22] of  $\tilde{P}$  and  $P$ .

$$KL(\tilde{G}\|G) \triangleq \sum_i \tilde{P}(i) \log \frac{\tilde{P}(i)}{P(i)}. \tag{17}$$

Setting the PFSA  $\tilde{G}$  to be the *symbolic white noise* PFSA  $E$ , Fig. 2(b) compares  $\|G\|_{\mathcal{A}^+}$  (dash-dot line), the Kullback-Leibler (K-L) divergence  $KL(E\|G)$  (solid line), and the K-L divergence  $KL(G\|E)$  (dashed line) as function of the probability distribution parameter  $\alpha$ . It is seen that these three curves are qualitatively similar as all of them approach infinity when  $\alpha$  approaches 0 or 1 and achieve the minimum at 0 if  $\alpha = 0.5$ . An advantage of the proposed PFSA norm  $\|\bullet\|_{\mathcal{A}^+}$  is that it induces a true metric, whereas K-L divergence does not.

Following Remark 3.2, the limiting behavior of PFSA  $G$  can be studied. For example,  $\frac{1}{n} \cdot G$  converges to the zero-PFSA  $E$  as  $n \rightarrow \infty$ . Also  $n \cdot G$  does not form a Cauchy sequence in this topology and the following corollary is immediate.

**Corollary 5.1.** Given a PFSA  $G = (Q, \Sigma, \delta, q_0, \tilde{\pi}) \in \mathcal{A}^+$ , if each row of the morph matrix  $\tilde{\Pi}$  (recall Definition 2.1) has a unique maximum element, then  $n \cdot G$  becomes a deterministic process as  $n \rightarrow \infty$ .

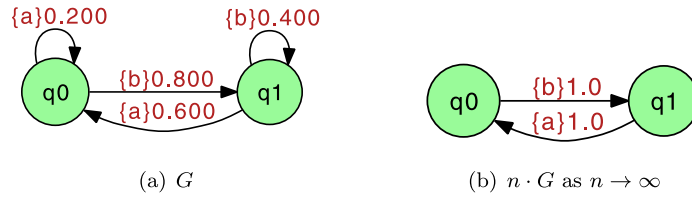


Fig. 3. An example of  $n \cdot G$  as  $n \rightarrow \infty$ .

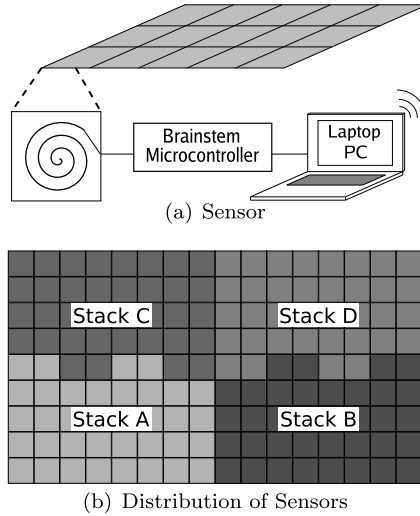


Fig. 4. Sensor layout in the laboratory environment.

Fig. 3 depicts an example of  $n \cdot G$  as  $n \rightarrow \infty$  over the alphabet  $\{a, b\}$ , which describes a deterministic process  $bababa \dots$  instead of a stochastic process.

### 6. Pattern recognition for identification of the robot type

This section presents an application of the theory on vector space representation of PFSA to pattern recognition for identification of the robot type [11]. The underlying concept is experimentally validated in a laboratory environment; the experimental apparatus is built upon a wireless network of distributed sensors, which interconnects mobile robots and robot simulators. The main objective of this experimentation is to identify the type of robot based on the statistical patterns of robot behavior that are affected by both parametric and nonparametric uncertainties such as:

1. Small variations in the robot mass that includes unloaded base weights of the platform itself and its payload.
2. Uncertainties in friction coefficients for robot traction.
3. Fluctuations in the robot motion due to small random delays in commands, that is largely a combination of communication delays and computational processing delays especially if the processor is heavily loaded.
4. Sensor uncertainties due to random noise in the A/D converter channels of the microprocessor.

#### 6.1. The distributed sensor network

The sensor network is imbedded underneath a pressure-sensitive floor that consists of piezoelectric wires serving as arrays of distributed pressure sensors. A coil of piezoelectric wire is placed under each of the  $0.65m \times 0.65m$  square floor tiles as shown in Fig. 4(a) such that the sensor generates an analog voltage due to pressure applied on it. This voltage is sensed by a Brainstem microcontroller using one of its 10-bit A/D channels, thereby yielding sensor readings in the range of 0 to 1023. A total of 144 sensors are placed in a  $9 \times 16$  grid of floor tiles to cover the entire laboratory environment as shown in Fig. 4(b). The sensors are grouped into four quadrants, each being connected to a stack consisting of 8 networked Brainstem microcontrollers for data acquisition. The microcontrollers are, in turn, connected to two laptop computers running Player servers [23] that collect the raw sensor data and distribute them to any client over the wireless network for further processing.



Fig. 5. Robot hardware: Pioneer 2AT (on left) and Segway RMP (on right).

### 6.2. Robot platforms

The robot hardware consists of two types of mobile robots, namely, Pioneer 2AT robots and Segway RMP. Fig. 5 shows a pair of Pioneer robots and a Segway RMP that have the following features:

- Pioneer 2AT is a four-wheeled robot that is equipped with a differential drive train system and has an approximate weight of  $\sim 35$  kilograms.
- Segway RMP is a two-wheeled robot (with inverted pendulum dynamics) that has a zero turn radius and has an approximate weight of  $\sim 70$  kilograms.

Both Pioneer 2AT robots and the Segway RMP robot are commanded to execute three different motion trajectories, namely, square motion, circular motion, and random motion. Since the spatial distribution of the pressure-sensitive coils underneath the floor (see Fig. 4) is statistically homogeneous, the decisions on detection of robot behavior patterns are statistically independent of the robot location (e.g., orientation of the square, center of the circle, and mean of the distribution for random motion). As the Pioneer robots are lighter than Segway and (four-wheeled) Pioneer's load on the floor is more evenly distributed than (two-wheeled) Segway's, the statistical behavior of these two types of robots are dissimilar. Furthermore, since their kinematics and dynamics are different, the texture of the respective pressure sensor signals are also different. From these perspectives, the following parameters are selected to specify different motions identically for both types of robots:

1. Edge length of the square motion: 3 meters.
2. Radius of the circular motion: 2 meters.
3. Distribution of the random motion: Uniform in the range of 1 to 4 meters in the  $x$ -direction and 1 to 4 meters in the  $y$ -direction.

### 6.3. Problem statement and experimental results

In the presence of the afore-mentioned uncertainties, a complete solution of the pattern identification problem may not be possible in a deterministic setting, because the pattern measure would not be identical even for similar robots behaving similarly. Therefore, the problem is posed in the statistical setting, where a family of pattern measures is generated from multiple experiments conducted under identical operating conditions. The requirement is to generate a family of patterns for each class of robot behavior that needs to be recognized. Each member of a family represents the pattern measure of a single experiment of one robot executing a particular motion profile.

By representing the movements of a robot with an alphabet of three symbols (i.e.,  $|\Sigma| = 3$ ) corresponding to three different types of motion, Fig. 6 depicts the following two PFSA:

$G_1$  for Pioneer 2AT and  $G_2$  for Segway RMP,

each of which is constructed as a  $D$ -Markov machine with the depth  $D = 1$  [1]. Given a PFSA  $P$  over the same symbol alphabet  $\Sigma$  and the same number of states, which is constructed from time series of robot movements by a robot of unknown type, the problem is to classify the unknown pattern  $P$  into exactly one of the two known patterns:  $G_1$  and  $G_2$  in terms of the norm in Eq. (11).

Since all three PFSA have the same structure, it suffices to consider their morph matrices according to Proposition 4.2.

$$\begin{aligned} \tilde{\Pi}^P &= \begin{pmatrix} 0.68 & 0.14 & 0.18 \\ 0.61 & 0.14 & 0.25 \\ 0.68 & 0.14 & 0.18 \end{pmatrix} \xrightarrow[\text{termwise}]{\text{Take inverse}} \begin{pmatrix} 1/0.68 & 1/0.14 & 1/0.18 \\ 1/0.61 & 1/0.14 & 1/0.25 \\ 1/0.68 & 1/0.14 & 1/0.18 \end{pmatrix} \\ &\xrightarrow[\text{Rows}]{\text{Normalize}} \begin{pmatrix} 0.10 & 0.51 & 0.39 \\ 0.13 & 0.55 & 0.32 \\ 0.10 & 0.51 & 0.39 \end{pmatrix} = \tilde{\Pi}^{-P}. \end{aligned}$$

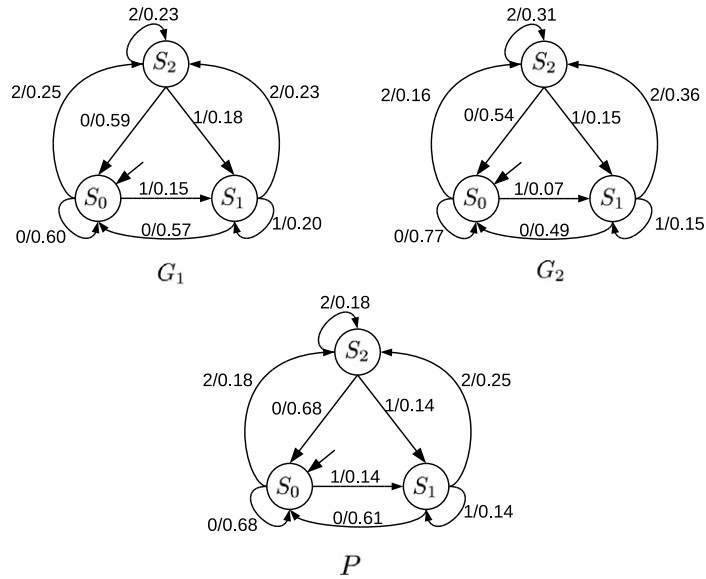


Fig. 6. PFSA constructed from different types of robot motion (all probabilities are rounded).

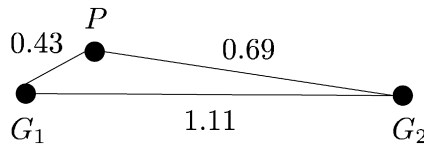


Fig. 7. Embedding of three PFSA in the 2-dimensional Euclidean space.

To verify that the morph matrix of the PFSA  $G_1 - P = G_1 + (-P)$ , elementwise multiplication (denoted by  $\cdot$ )  $\tilde{\Pi}^{G_1}$  is used with  $\tilde{\Pi}^{-P}$  and then the rows are normalized.

$$\tilde{\Pi}^{G_1} \cdot \tilde{\Pi}^{-P} = \begin{pmatrix} 0.60 \times 0.10 & 0.15 \times 0.51 & 0.25 \times 0.39 \\ 0.57 \times 0.13 & 0.20 \times 0.55 & 0.23 \times 0.32 \\ 0.59 \times 0.10 & 0.18 \times 0.51 & 0.23 \times 0.39 \end{pmatrix} \xrightarrow[\text{Rows}]{\text{Normalize}} \begin{pmatrix} 0.27 & 0.32 & 0.41 \\ 0.29 & 0.43 & 0.28 \\ 0.25 & 0.38 & 0.37 \end{pmatrix} = \tilde{\Pi}^{G_1 - P}.$$

By virtue of part 3 in Theorem 4.1, it follows that the norm  $\|G_1 - P\| = \log(\frac{0.43}{0.28}) = 0.43$ , where natural logarithm is used. Similarly, the other two pairs of distances are evaluated as:  $\|G_2 - P\| = 0.69$  and  $\|G_1 - G_2\| = 1.11$ ; it is noticed that the triangular inequality holds. Since  $\|G_1 - P\| < \|G_2 - P\|$ , it is concluded that  $P$  is closer to  $G_1$  and hence the robot motion is classified to be  $G_1$ . Fig. 7 shows the embedding of the three PFSA geometrically in a plane.

One can also find out the middle point  $\tilde{G}$  of  $G_1$  and  $G_2$  by computing  $\tilde{G} = \frac{G_1 + G_2}{2}$  by the following procedure.

$$\tilde{\Pi}^{G_1 + G_2} = \begin{pmatrix} 0.90 & 0.02 & 0.08 \\ 0.78 & 0.05 & 0.17 \\ 0.81 & 0.05 & 0.14 \end{pmatrix} \xrightarrow[\text{termwise}]{\text{Square root}} \begin{pmatrix} \sqrt{0.90} & \sqrt{0.02} & \sqrt{0.08} \\ \sqrt{0.78} & \sqrt{0.05} & \sqrt{0.17} \\ \sqrt{0.81} & \sqrt{0.05} & \sqrt{0.14} \end{pmatrix} \xrightarrow[\text{Rows}]{\text{Normalize}} \begin{pmatrix} 0.68 & 0.12 & 0.20 \\ 0.59 & 0.14 & 0.27 \\ 0.60 & 0.15 & 0.25 \end{pmatrix} = \tilde{\Pi}^{\tilde{G}}.$$

### 7. Conclusions and future work

This paper presents the construction of a normed vector space over the real field  $\mathbb{R}$  for a class of probabilistic finite state automata (PFSA) in the measure-theoretic setting. The operations of vector addition and scalar multiplication operations and the norm on the space of PFSA are defined by establishing an isometric isomorphism between the spaces of probability measures and PFSA, and the physical significance of these operations is interpreted. An application of this mathematical framework is pattern recognition for identification of the robot motion in a laboratory environment. While there are many areas of future work for both theoretical and application-oriented research, a few examples of potential research topics are presented below.

- *Extension of the vector space model to the full range  $\mathcal{A}$  of PFSA:* This will require removal of restriction of the morph matrix  $\tilde{T}$  being strictly positive.
- *Development of methods for construction of norms (and inner products) on the vector space of PFSA:* Different metrics could be introduced on the space  $\mathcal{A}^+$ , including the norm defined in this paper and metrics introduced in earlier publications [21,5]. The research issue is to develop an appropriate norm for the problem at hand.
- *Formulation of one or more methods of decision and control synthesis based to the vector space model of PFSA:* This work could be to some extent analogous to what has been done for classical design in finite-dimensional vector spaces.

## Acknowledgments

The authors acknowledge benefits of discussion with their colleagues Dr. Ishanu Chattopadhyay and Dr. Shashi Phoha.

## References

- [1] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, *Signal Process.* 84 (7) (2004) 1115–1130.
- [2] X. Jin, S. Gupta, K. Mukherjee, A. Ray, Wavelet-based feature extraction using probabilistic finite state automata for pattern classification, *Pattern Recognit.* 44 (7) (2011) 1343–1356.
- [3] M. Egerstedt, E. Frazzoli, G. Pappas, Special section on symbolic methods for complex control systems, *IEEE Trans. Automat. Control* 51 (6) (2006) 921–923.
- [4] G. Pola, P. Tabuada, Symbolic models for nonlinear control systems: Alternating approximate bisimulations, *SIAM J. Control Optim.* 48 (2) (2009) 719–733.
- [5] E. Vidal, F. Thollard, C. de la Higuera, F. Casacuberta, R. Carrasco, Probabilistic finite-state machines – part I, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (7) (2005) 1013–1025.
- [6] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech processing, *Proc. IEEE* 77 (2) (1989) 257–286.
- [7] M. Vidyasagar, The complete realization problem for hidden Markov models: a survey and some new results, *Math. Control Signals Systems* 23 (2011) 1–65, doi:10.1007/s00498-011-0066-7.
- [8] R. Carrasco, J. Oncina, Learning stochastic regular grammars by means of a state merging method, in: *Proceedings of Second International Colloquium on Grammatical Inference and Applications*, Alicante, Spain, 1994, pp. 139–152.
- [9] L. Saul, F. Pereira, Aggregate and mixed-order Markov models for statistical language processing, in: *Empirical Methods in Natural Language Processing*, 1997, pp. 81–89.
- [10] D. Ron, Y. Singer, N. Tishby, On the learnability and usage of acyclic probabilistic finite automata, *J. Comput. System Sci.* 56 (2) (1998) 133–152.
- [11] G. Mallapragada, A. Ray, X. Jin, Symbolic dynamic filtering and language measure for behavior identification of mobile robots, *IEEE Trans. Syst. Man Cybernet. Part B: Cybernet.*, doi:10.1109/TSMCB.2011.2172419, in press.
- [12] C. Shalizi, K. Shalizi, Blind construction of optimal nonlinear recursive predictors for discrete sequences, in: *AUAI'04: Proceedings of the 20th Conference on Uncertainty in Artificial Intelligence*, AUAI Press, Arlington, VA, USA, 2004, pp. 504–511.
- [13] R. Duda, P. Hart, D. Stork, *Pattern Classification*, 2nd edition, John Wiley & Sons, Inc., New York, NY, USA, 2001.
- [14] C.M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, NY, USA, 2006.
- [15] S. Iyengar, P. Varshney, T. Damarla, A parametric copula-based framework for hypothesis testing using heterogeneous data, *IEEE Trans. Signal Process.* 59 (5) (2011) 2308–2319.
- [16] A. Ray, Signed real measure of regular languages for discrete event supervisory control, *Internat. J. Control* 78 (12) (2005) 949–967.
- [17] T. Barfoot, G. D'Eleuterio, An algebra for the control of stochastic systems: Exercises in linear algebra, in: *Proceedings of the 5th International Conference on Dynamics and Control of Structures in Space (DCSS)*, Cambridge, England, 2002 pp. 14–18.
- [18] D. Lind, B. Marcus, *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, Cambridge, UK, 1995.
- [19] J.E. Hopcroft, R. Motwani, J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, 2nd edition, Addison–Wesley, Boston, MA, USA, 2001.
- [20] W. Rudin, *Real and Complex Analysis*, 3rd edition, McGraw–Hill, New York, NY, USA, 1987.
- [21] I. Chattopadhyay, A. Ray, Structural transformations of probabilistic finite state machines, *Internat. J. Control* 81 (5) (2008) 820–835.
- [22] T. Cover, J. Thomas, *Elements of Information Theory*, 2nd edition, Wiley–Interscience, New York, NY, USA, 2006.
- [23] B. Gerkey, R. Vaughan, A. Howard, The player/stage project: Tools for multi-robot and distributed sensor systems, in: *Proceedings of the International Conference on Advanced Robotics (ICAR 2003)*, Coimbra, Portugal, June 30–July 3, 2003, pp. 317–323.