Signal Processing 93 (2013) 2594–2611

Contents lists available at SciVerse ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Hilbert space formulation of symbolic systems for signal representation and analysis ☆

Yicheng Wen, Asok Ray*, Shashi Phoha

The Pennsylvania State University, University Park, PA 16802, USA

ARTICLE INFO

Article history:
Received 26 June 2012
Accepted 2 February 2013
Available online 18 February 2013

Keywords:
Probabilistic finite state automata
Inner product space
Signal representation
Model order reduction

ABSTRACT

Probabilistic finite state automata (PFSA) have been widely used as an analysis tool for signal representation and modeling of physical systems. This paper presents a new method to address these issues by bringing in the notion of vector-space formulation of symbolic systems in the setting of PFSA. In this context, a link is established between the formal language theory and functional analysis by defining an inner product space over a class of stochastic regular languages, represented by PFSA models that are constructed from finite-length symbol sequences. The norm induced by the inner product is interpreted as a measure of the information contained in the respective PFSA. Numerical examples are presented to illustrate the computational steps in the proposed method and to demonstrate model order reduction via orthogonal projection from a general Hilbert space of PFSA onto a (closed) Markov subspace that belongs to a class of shifts of finite type. These concepts are validated by analyzing time series of ultrasonic signals, collected from an experimental apparatus, for fatigue damage detection in polycrystalline alloys.

1. Introduction

Symbolization-based techniques have been developed for probabilistic analysis of physical signals in terms of stochastic regular languages as a convenient framework to achieve a common treatment of heterogeneous models of dynamical systems [1]. Examples are signal representation and modeling in dynamical systems [2] and pattern recognition [3,4]. The key idea here is that, by partitioning the (possibly pre-processed) time series or image data observed from the underlying physical system, a string of symbols are generated to construct a finite-state stochastic language model.

In the context of symbolic systems, construction of several probabilistic finite state models has been reported in the literature [5,6]; examples are probabilistic finite state automata (PFSA), hidden Markov models (HMM) [7–9], stochastic regular grammars [10], and Markov chains [11]. In this paper, PFSA models have been used to serve as recognizers of stochastic regular languages. A major advantage of using a PFSA model is that, in general, it is easier to learn from a dynamical system [12–14] although PFSA may not be as powerful as other models like HMM [5]. The definition of PFSA (see Definition 2.1), adopted in this paper, is slightly different from that used in [5]. In the sequel, the notion of PFSA is as stated in Definition 2.1.

Symbolic models are abstract descriptions of continuously varying systems in which symbols represent aggregates of continuous states. During the last one and a half

☆ This work has been supported in part by the U.S. Army Research Laboratory and the U.S. Army Research Office under Grant no. W911NF-07-1-0376 and by the U.S. Office of Naval Research under Grant no. N00014-09-1-0688. Any opinions, findings and conclusions or recommendations expressed in this publication are those of the authors and do not necessarily reflect the views of the sponsoring agencies.

* Corresponding author.
E-mail addresses: ywen2000@gmail.com (Y. Wen), axr2@psu.edu (A. Ray), sxp26@psu.edu (S. Phoha).

0165-1684/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.sigpro.2013.02.002

Nomenclature

$+$	addition operation of PFSA [Definition 4.3]
$\oplus$	addition of two probability measures on $\mathcal{B}_\Sigma$ [Definition 3.1]
$\cdot$	scalar multiplication operation of PFSA over $\mathbb{R}$ [Definition 4.3]
$\odot$	scalar multiplication of a probability measure on $\mathcal{B}_\Sigma$ over $\mathbb{R}$ [Definition 3.2]
$\otimes$	synchronous composition of PFSA [Definition 2.4]
$\langle \cdot,\cdot \rangle$	inner product of two probability measures [Definition 3.4]
$\langle \cdot,\cdot \rangle_A$	inner product of two PFSA [Definition 4.4]
$\sim$	equivalence relation in the space of probability measures [Definition 3.3]
δ	state transition function of PFSA [Definition 2.1]
$\delta^\star$	extended state transition function of PFSA [Remark 2.2]
Ξ	equivalence relation in the space of PFSA [Definition 4.2]
μ	normalized finite measure defined on the σ-algebra $2^{\Sigma^\star}$ [Section 3.2]
Π	state transition matrix of PFSA [Definition 2.2]
$\tilde{\Pi}$	probability morph matrix of PFSA [Definition 2.1]
$\tilde{\pi}$	probability morph function of PFSA [Definition 2.1]
$\|\Sigma\|$	cardinality of the alphabet Σ [Section 2]
Σ^ω	the set of all strictly infinite-length strings on Σ [Section 2]
$\Sigma^\star$	the set of all finite-length strings on Σ [Section 2]
$\mathcal{B}_\Sigma$	smallest σ-algebra on the set $\{x\Sigma^\omega$ where $x \in \Sigma^\star\}$ [Definition 2.5]
$\mathcal{D}_d$	space of D-Markov machines with depth d [Example 7.2]
$\mathbb{F}$	inverse of the map $\mathbb{H}$ [Definition 4.3]
$\mathcal{G}$	quotient space of PFSA over the alphabet Σ [Equation (18)]
$\tilde{\mathcal{G}}$	the set of all strictly positive PFSA [Definition 2.3]
$\mathbb{H}$	isomorphism between the vector spaces $(\mathcal{G},+,\cdot)$ and $(\mathcal{Q}_f,\oplus,\odot)$ [Eq. (18)]
$\tilde{\mathbb{H}}$	mapping from the space $\tilde{\mathcal{G}}$ to the space $\mathcal{Q}_f$ [Definition 2.8]
m_n	function mapping from $2^{\Sigma^\star}$ to $[0,1]$ [Definition 6.1]
$\mathcal{N}_p$	probabilistic Nerode equivalence on $(\Sigma^\omega,\mathcal{B}_\Sigma,p)$ [Definition 2.7]
$\mathbf{N}_p$	the set of all Nerode equivalence classes on $\Sigma^\star$ induced by $\mathfrak{N}_p$ [Remark 2.5]
$\mathbb{P}_{\mathcal{G}_2}$	orthogonal projection from $\mathcal{G}_1 \subseteq \mathcal{G}$ onto $\mathcal{G}_2 \subseteq \mathcal{G}$ [Example 7.2]
$\mathcal{P}$	the set of all strictly positive probability measures on $\mathcal{B}_\Sigma$ [Definition 2.6]
$\mathcal{Q}$	quotient space $\mathcal{P}/_\sim$ [Definition 3.3]
$\mathcal{Q}_f$	subspace of $\mathcal{Q}$ with a finite number of Nerode equivalence classes [Definition 4.1]
$\mathbb{U}(G)$	uniformizer of PFSA G [Definition 6.2]
$[x]_p$	the set of all equivalent strings of x for a measure p [Proposition 4.1]

decade, there has been a growing interest in the use of symbolic models as an analytical tool for signal representation, modeling, pattern recognition, and decision and control of interacting dynamical systems (e.g., planning and navigation of autonomous robots in uncertain environments, fault detection in aerospace systems, and military intelligence, surveillance and reconnaissance). In these systems, the notion of classic objectives (e.g., robust stability and performance) is augmented with system-level issues and their hardware and software implementation on remotely located computational platforms that are interconnected over a communication network. Along this line, several researchers (e.g., [2]) have reported theoretical work that includes stability analysis and synthesis of both linear and nonlinear time-delayed control systems that may be subjected to disturbances.

In general, it would be desirable to be able to treat PFSA models or stochastic languages as vectors in a Hilbert space for applications in pattern recognition and information fusion. For example, in pattern recognition, if PFSA are used as feature vectors (e.g., [3,15]), then the lack of a precise mathematical structure on the feature space (i.e., the space of PFSA in this case) may not allow direct usage of classical signal processing and machine learning tools [16]. Similarly, for enhancement of information fusion [17] and information source localization [18] tools that are often computation-intensive, PFSA models are capable of efficiently compressing the information derived from sensor time series [19]; but the problem is how to fuse these heterogeneous sources of compressed information. An example is to construct a linear combination of PFSA models with larger weights assigned to more reliable ones to increase the signal-to-noise ratio if the sensors are of the same type. For heterogeneous sources, one may project the PFSA models onto a subspace defined over a common alphabet for correlation analysis. It is also useful to perform model order reduction on PFSA models for higher level fusion, such as situation assessment. Nevertheless mathematical operations on PFSA are required for such feature level fusion.

The theory of how to algebraically manipulate two PFSA has not been explored except for a few cases. Ray [20] introduced the notion of vector space construction for finite state automata over the finite field $GF(2)$. Barfoot and D'Leuterio [21] 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 4.5). Recently, Wen and Ray [22] introduced the concept of a vector space of PFSA over the real field $\mathbb{R}$, and Adenis et al. [23] used an algebraic approach to construct an inner-product space structure over $\mathbb{R}$ for a class of PFSA. Apparently, no other prior work exists for rigorously defining a mathematical structure in the vector space of PFSA.

The major contribution of this paper is the construction of a mathematical structure in the space of PFSA using a measure-theoretic approach as a generalization to the algebraic approach proposed by Adenis et al. [23]. Along this line, a vector space is constructed over $\mathbb{R}$ for a class of stochastic languages, which can be realized as a set of dynamic models represented by PFSA that are generated from finite-length symbol sequences derived from time series of physical signals. A family of inner products is introduced on this vector space, which is configurable through a user-selectable measure. A vector subspace is also constructed by a quotient map on PFSA, where the algebraic and topological operations in the vector space are physically interpreted; especially, an analogy is drawn between the norm induced from the inner product and the entropy rate in an information-theoretic setting [24]. Potential applications of this formulation are discussed through examples of model order reduction. The proposed analytical approach has the following potential benefits.

1. Development of a mathematical structure for solving problems of signal representation, modeling and analysis (e.g., model identification, model order reduction, and system performance analysis) in the setting of symbolic dynamics [25].
2. Establishment of a link between the theories of formal languages [26] and functional analysis [27] to enhance the tools of solving signal analysis problems in physical processes.

The paper is organized in eight sections. Section 2 presents the preliminary concepts and notations in the formal language theory and related previous work. Section 3 uses a measure-theoretic approach to construct the vector space by defining algebraic operations (e.g., vector addition and scalar multiplication) and introduces a family of inner products to provide a topological structure in a general setting of stochastic languages. Section 4 focuses on stochastic regular languages, where an isometric isomorphism is defined between the vector space of stochastic regular languages and the vector space of PFSA. Section 5 presents physical interpretations of the algebraic operations and their applications to system modeling via stochastic regular languages. Section 6 addresses the choice and computation of the measure μ for constructing the inner product. Section 7 presents two numerical examples to illustrate the underlying concepts and to demonstrate model order reduction based on experimental data. The paper is concluded with recommendations for future work in Section 8.

2. Preliminaries

In the formal language theory [26], an alphabet Σ is a (non-empty finite) set of symbols, i.e., the alphabet's cardinality $|\Sigma| \in \mathbb{N}$, the set of positive integers. A string x over Σ is a finite-length sequence of symbols in Σ. The length of a string x, 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 symbols including the null string ϵ that has zero length, i.e., $|\epsilon| = 0$; cardinality of $\Sigma^\star$ is $\aleph_0$ (countably infinite). The notation Σ^+ means the subset of $\Sigma^\star$ without ϵ, i.e., $\Sigma^+ = \Sigma^\star \backslash \{\epsilon\}$; and Σ^ω represents the set of all strictly infinite-length strings over Σ, where the cardinality of Σ^ω is uncountable.

2.1. PFSA model for symbolic systems

This subsection introduces basic notations and definitions related to PFSA models that are used in the sequel.

Definition 2.1 (*PFSA*). A probabilistic finite state automaton (PFSA) is a tuple $G = (Q, \Sigma, \delta, q_0, \tilde{\pi})$, where

- Q is a (non-empty) finite set, called the set of states;
- Σ is a (non-empty) finite set, called the input alphabet;
- $\delta : Q \times \Sigma \rightarrow Q$ is the state transition function;
- $q_0 \in Q$ is the start state;
- $\tilde{\pi} : Q \times \Sigma \rightarrow [0,1]$ is the probability morph function and satisfies the condition $\sum_{\sigma \in \Sigma} \tilde{\pi}(q,\sigma) = 1$ for all $q \in Q$. The probability morph function $\tilde{\pi}$ is represented in a matrix form as $\tilde{\Pi}$ with the element $\tilde{\Pi}_{ij} \triangleq \tilde{\pi}(q_i, \sigma_j)$.

Note: All states in a PFSA are reachable from the start state. Otherwise, any non-reachable states are removed from Q.

Remark 2.1. The PFSA model defined in this paper is slightly different from that in [5], as mentioned earlier in Section 1. One of the major differences is that, in this paper, the PFSA models do not have final (or terminating) state probabilities. The rationale is that the statistical behavior of dynamical systems under consideration is quasi-static in nature.

Remark 2.2. The transition map δ naturally induces an extended transition function $\delta^\star : Q \times \Sigma^\star \rightarrow Q$ such that $\delta^\star(q,\epsilon) = q$ and $\delta^\star(q, x\sigma) = \delta(\delta^\star(q,x),\sigma)$ $\forall q \in Q$, $\forall x \in \Sigma^\star$, and $\forall \sigma \in \Sigma$.

Remark 2.3. The symbol sequence $\{\sigma_k\}_{k=1}^{\infty}$ of a PFSA can be realized from the respective rows of the morph matrix as follows:

1. generate a symbol $\sigma \in \Sigma$ according to the probability mass function $\tilde{\pi}(q_{k-1}, \cdot)$, where $k > 0$;
2. cause the kth state transition to generate $q_k = \delta(q_{k-1}, \sigma_k)$;
3. increment k and go to Step 1.

Note that, in general, the realizations of individual PFSA are different, but all such symbol sequences share the same statistics specified by the morph function $\tilde{\pi}$.

Remark 2.4. In PFSA, a state transition is modeled via occurrence of symbols; in contrast, in an uncontrolled Markov chain, a state transition is not specified. Therefore, it is usually difficult to make sense of comparing or combining two Markov chains unless the meaning of the states are related. However, as long as two PFSA are constructed over the same alphabet, their underlying semantics can be compared and combined.

Definition 2.2 (*State transition probability matrix*). For every PFSA $G=(Q,\Sigma,\delta,q_0,\tilde{\pi})$, there is an associated stochastic matrix $\Pi\in\mathbb{R}^{|Q|\times|Q|}$, called the state transition (probability) matrix, which is defined as follows:

$$\Pi_{jk}=\sum_{\sigma:\delta(q_j,\sigma)=q_k}\tilde{\pi}(q_j,\sigma) \tag{1}$$

Every PFSA G induces a Markov chain $\{X_n,X_n\in Q\}$ with the state transition probability matrix Π.

Given a finite-length symbol sequence $\mathbb{S}$ over an alphabet Σ, there exist several PFSA construction algorithms (e.g., [13,14]) to discover an underlying PFSA model G. These algorithms start with identifying the structure of G, i.e., (Q,Σ,δ,q_0). Then, a $|Q|\times|\Sigma|$ count matrix C is introduced with each of its elements initialized to 1. Let N_{ij} denote the number of times a symbol σ_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

$$\tilde{\Pi}_{ij}\triangleq\frac{C_{ij}}{\sum_{k=1}^{|\Sigma|}C_{ik}}=\frac{1+N_{ij}}{|\Sigma|+\sum_{k=1}^{|\Sigma|}N_{ik}} \tag{2}$$

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 the sequence $\mathbb{S}$, then there should be no preference to any specific symbols emanating from q_i. Therefore, it is logical to initialize $\tilde{\Pi}_{ij}=1/|\Sigma|$, i.e., the uniform distribution for the ith row of the morph matrix $\tilde{\Pi}$. (It is shown later in the paper that a morph matrix with all elements equal to $1/|\Sigma|$ 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 $\mathbb{S}$. In this setting, the ith row $\tilde{\Pi}_i$ of the morph matrix is a random vector that follows Dirichlet distribution [28] as described below:

$$f_{\tilde{\Pi}_i}(\boldsymbol{\theta})=\frac{1}{B(C_i)}\prod_{j=1}^{|\Sigma|}(\theta_{ij})^{C_{ij}-1} \tag{3}$$

where $\boldsymbol{\theta}_i\triangleq[\theta_{i1}\ \cdots\theta_{i|\Sigma|}]$ is a realization of $\tilde{\Pi}_i$ and the normalization constant is

$$B(C_i)\triangleq\frac{\prod_{j=1}^{|\Sigma|}\Gamma(C_{ij})}{\Gamma(\sum_{j=1}^{|\Sigma|}C_{ij})} \tag{4}$$

where $\Gamma(\bullet)$ is the standard gamma function. It is noted that the Dirichlet distribution approaches the δ-distribution as more and more symbols are observed [29]. This procedure guarantees that each element of the morph matrix $\tilde{\Pi}$ is strictly positive for any finite-length symbol sequence $\mathbb{S}$.

Definition 2.3 (*Set of strictly positive PFSA*). The set of strictly positive PFSA is defined as

$$\tilde{\mathcal{G}}\triangleq\{(Q,\Sigma,\delta,q_0,\tilde{\pi}):\tilde{\pi}(q,\sigma)>0\ \forall q\in Q\ \text{and}\ \forall\sigma\in\Sigma\}$$

Definition 2.4 (*Synchronous composition*). The synchronous composition of two PFSA $G_i\triangleq(Q_i,\Sigma,\delta,q_0^{(i)},\tilde{\pi}_i)\in\mathcal{G}$, $i=1,2$, denoted by $\otimes:\tilde{\mathcal{G}}\times\tilde{\mathcal{G}}\to\tilde{\mathcal{G}}$, is defined as

$$G_1\otimes G_2=(Q_1\times Q_2,\Sigma,\delta',(q_0^{(1)},q_0^{(2)}),\tilde{\pi}')\quad \forall q_i\in Q_1,\quad \forall q_j\in Q_2,\quad \forall\sigma\in\Sigma$$

where

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

The synchronous composition is not commutative since only the morph matrix of the first PFSA is used in the composition. Essentially, synchronous composition breaks down each state in the first PFSA into a set of “sub-states“ according to the structure of the second PFSA but still keeps the output probability distribution the same inside each set of “sub-states”.

2.2. Probability measures

This subsection develops the notion of probability measures by introducing the following definitions.

Definition 2.5 (*Probability measure space*). [30] :Given an alphabet Σ, the set $\mathcal{B}_\Sigma\triangleq 2^{\Sigma^\star}\Sigma^\omega$ is the σ-algebra generated by the set $\{L:L=x\Sigma^\omega$ *and* $x\in\Sigma^\star\}$. 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 a prefix. In particular, $p(\epsilon)$ is the probability of the entire set Σ^ω and thus it follows that $p(\epsilon)=1$ by the probability axioms.

Definition 2.6 (*Space $\mathcal{P}$ of probability measures*). Given the probability measure space $(\Sigma^\omega,\mathcal{B}_\Sigma,p)$, let $\mathcal{P}$ denote the space of strictly positive probability measures on $\mathcal{B}_\Sigma$, namely, $\mathcal{P}\triangleq\{p:\mathcal{B}_\Sigma\to[0,1]$ and such that $p(x)>0,\forall x\in\Sigma^\star\}$. Thus, each element of $\mathcal{P}$ assigns a non-zero probability to any finite string.

Definition 2.7 (*Probabilistic Nerode equivalence [30]*). 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 $(\Sigma^\omega,\mathcal{B}_\Sigma,p)$, denoted by $x\mathcal{N}_p y$, if exactly one of the following two conditions is true:

1. $p(x)=0$ and $p(y)=0$;
2. $\forall s\in\Sigma^\star$, $p(xs)/p(x)=p(ys)/p(y)$ for $p(x)\neq 0$ and $p(y)\neq 0$.

The probabilistic Nerode relation is a right-invariant equivalence relation [30] that is referred to as probabilistic Nerode equivalence in the sequel. The probabilistic Nerode equivalence $\mathcal{N}_p$ of a measure induces a partition $\mathbf{N}_p$ of the set $\Sigma^\star$.

2.3. Relationship between PFSA and probability measures

This subsection presents the relationship between PFSA and probability measures.

Definition 2.8 (*Mapping of PFSA [30]*). Following Definition 2.3, let $G \triangleq (Q,\Sigma,\delta,q_0,\tilde{\pi}) \in \tilde{\mathcal{G}}$ be a strictly positive PFSA and the associated probability measure be $p \in \mathcal{P}$. Then, a map $\tilde{\mathbb{H}} : \tilde{\mathcal{G}} \to \mathcal{P}$ is defined as $\tilde{\mathbb{H}}(G) = p$ such that

$$p(x) = \tilde{\pi}(q_0,\sigma_1) \prod_{k=1}^{r-1} \tilde{\pi}(\delta^\star(q_0,\sigma_1 \cdots \sigma_k),\sigma_{k+1}) \tag{5}$$

where the symbol string $x = \sigma_1 \cdots \sigma_r \in \Sigma^\star$ and $r = |x| \in \mathbb{N}$. Then, the measure p is said to be encoded by the PFSA G, or the PFSA G encodes the probability measure p.

Fig. 1 shows an example of a PFSA G over $\Sigma = \{a,b\}$ and its encoded probability measure $p = \tilde{\mathbb{H}}(G)$.

Proposition 2.1 (*Synchronous composition [30]*). *Following Definition2.4, if $G_1,G_2 \in \tilde{\mathcal{G}}$, then*

$$\mathbb{H}(G_1 \otimes G_2) = \mathbb{H}(G_1) \quad \text{and} \quad \mathbb{H}(G_2 \otimes G_1) = \mathbb{H}(G_2)$$

Proposition 2.1, whose proof is given in [30], implies that any two PFSA over the same alphabet can be transformed into a common structure without affecting their encoding measures by applying synchronous composition.

Remark 2.5. The map $\tilde{\mathbb{H}} : \tilde{\mathcal{G}} \to \mathcal{P}$ in Definition 2.8 may not be injective. In other words, there may exist different PFSA realizations that encode the same probability measure on $\mathcal{B}_\Sigma$ due to two reasons: (i) non-minimal realization, and (ii) state relabeling. For example, in Fig. 2, all three PFSA essentially encode the same measure on $\mathcal{B}_\Sigma$ in spite of their different representations. The PFSA on the right 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 top PFSA in Fig. 2. The left two PFSA are exactly the same with the exception that the states are relabeled; and this relabeling does not affect the underlying encoded measure.

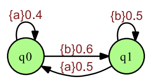
The probabilistic Nerode equivalence of a measure p induces a partition $\mathbf{N}_p$ of the set $\Sigma^\star$ (see Definition 2.7). Thus, the measure p can be encoded by a PFSA if and only if the partition $\mathbf{N}_p$ is finite [30]. However, all measures in $\mathcal{P}$ cannot be encoded by PFSA, because a PFSA only encodes measures that belong to finite probabilistic Nerode equivalence classes. Therefore, the map $\tilde{\mathbb{H}} : \tilde{\mathcal{G}} \to \mathcal{P}$ in Definition 2.8 may not be surjective.

Since the range of the map $\tilde{\mathbb{H}}$ is the space of measures with finite Nerode equivalence classes [30], restricting $\mathcal{P}$ to the range of $\tilde{\mathbb{H}}$ yields the right inverse of $\tilde{\mathbb{H}}$, denoted by $\tilde{\mathbb{F}}$, i.e., $\tilde{\mathbb{H}} \circ \tilde{\mathbb{F}} = I$. Then, it follows that if a measure $p(x)$ exists for each $x \in \Sigma^\star$, then the minimal representation of a PFSA G (e.g., see Fig. 1) can be generated by Algorithm 1.

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

Input: $(\Sigma^\omega,\mathcal{B}_\Sigma,p)$ such that $\mathfrak{N}_p$ is of finite index $n \in \mathbb{N}$;
Output: G;
Let $Q = \{q_j : j \in \{1,\ldots,n\}\}$ be the set of equivalence classes of the relation $\mathfrak{N}_p$;
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

a



b

x	a	b	aa	ab	ba	bb	…
p(x)	0.4	0.6	0.16	0.24	0.3	0.3	…

Fig. 1. PFSA G (with q_0 as the initial state) and its encoded measure p. (a) G, (b) $p = \tilde{\mathbb{H}}(G)$.

3. Inner product space of probability measures

This section first presents the algebraic construction of a vector space over the real field $\mathbb{R}$. Then, an inner product

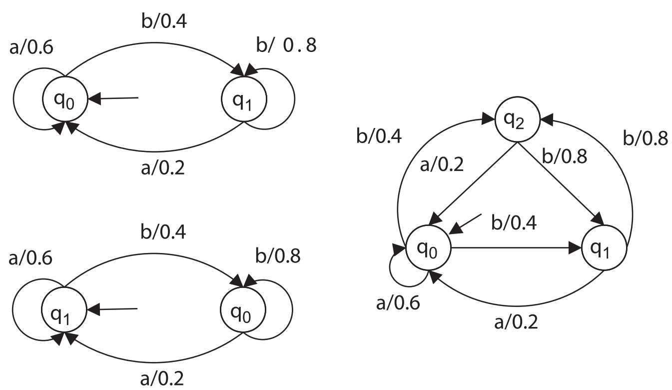


Fig. 2. Different PFSA realizations of the same probability measure.

is defined on this vector space to build a topological structure. Such a structure is useful for applications like model identification and model order reduction [31].

3.1. Construction of the vector space

This subsection constructs a vector space of PFSA over the real field $\mathbb{R}$. To this end, the notions of algebraic operations of vector addition and scalar multiplication are introduced below.

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

(1) $p_3(\epsilon) = 1$.
(2) $\forall x \in \Sigma^\star$ and $\sigma \in \Sigma$,

$$\frac{p_3(x\sigma)}{p_3(x)} = \frac{p_1(x\sigma)p_2(x\sigma)}{\sum_{\alpha\in\Sigma} p_1(x\alpha)p_2(x\alpha)} \tag{6}$$

(3) For all countable pairwise disjoint sets $\{x_i\Sigma^\omega\}$, $p_3(\bigcup_i\{x_i\Sigma^\omega\}) = \sum_i p_3(x_i)$.

In the above definition, it follows that p_3 is a probability measure on $\mathcal{P}$, because $\forall x \in \Sigma^\star$,

$$\sum_{\sigma\in\Sigma} p_3(x\sigma) = \left(\sum_{\sigma\in\Sigma}\frac{p_1(x\sigma)p_2(x\sigma)}{\sum_{\alpha\in\Sigma}p_1(x\alpha)p_2(x\alpha)}\right)p_3(x) = p_3(x)$$

Proposition 3.1. *The algebraic system* $(\mathcal{P},\oplus)$ *forms an Abelian group.*

Proof. The closure and commutativity properties follow directly from Definition 3.1. Associativity, existence of identity, and existence of the inverse element are established below.

- Associativity:

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 $\forall \sigma \in \Sigma$.

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

- Existence and uniqueness of the identity:

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

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

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 and uniqueness of an inverse:

$\forall p \in \mathcal{P}$, $\forall x \in \Sigma^\star$ and $\forall \sigma \in \Sigma$, let a probability measure $-p$ be defined as

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

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

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

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

Remark 3.1. In the sequel, the zero-element $\underline{e}$ of the Abelian group $(\mathcal{P},\oplus)$ is denoted as the *symbolic white noise*. For the symbolic white noise, every string of the same length has equal probability of occurrence and the knowledge of the history does not provide any information for predicting the future.

Next the scalar multiplication operation is defined over the real field $\mathbb{R}$.

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

(1) $(k \odot p)(\epsilon) = 1$;
(2) $\forall x \in \Sigma^\star$ and $\sigma \in \Sigma$

$$\frac{(k\odot p)(x\sigma)}{(k\odot p)(x)} = \frac{p^k(x\sigma)}{\sum_{\alpha\in\Sigma}p^k(x\alpha)} \tag{7}$$

(3) for all countable pairwise disjoint sets $\{x_i\Sigma^\omega\}$, $(k\odot p)(\bigcup_i\{x_i\Sigma^\omega\}) = \sum_i (k\odot p)(x_i)$

where $p^k(x\sigma) = [p(x\sigma)]^k$, $k \in \mathbb{R}$, $p \in \mathcal{P}$, $x \in \Sigma^\star$, and $\sigma \in \Sigma$.

Remark 3.2. It follows from Definitions 3.1 and 3.2 that $k \odot p$ is a valid probability measure on $\mathcal{P}$. By convention, it is asserted 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 3.1 (*Vector space construction*). $(\mathcal{P}, \oplus, \odot)$ *defines a vector space over the real field* $\mathbb{R}$.

Proof. Let $p,p_1,p_2 \in \mathcal{P}$; $k,k_1,k_2 \in \mathbb{R}$; $x \in \Sigma^\star$; and $\sigma \in \Sigma$. The following equalities are checked:

- To show that $k \odot p_1 \oplus k \odot p_2 = k \odot (p1 \oplus p_2)$.

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

- To show that $(k_1+k_2) \odot p = k_1 \odot p \oplus k_2 \odot p$.

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

- To show that $k_1 \odot (k_2 \odot p) = (k_1k_2) \odot p$.

$$\frac{(k_1 \odot (k_2 \odot p))(x\sigma)}{(k_1 \odot (k_2 \odot p))(x)} = \frac{(k_2 \odot p)^{k_1}(x\sigma)}{\sum_{\alpha\in\Sigma}(k_2 \odot p)^{k_1}(x\alpha)} = \frac{\left(\frac{p^{k_2}(x\sigma)}{\sum_{\beta\in\Sigma}p^{k_2}(x\beta)}\right)^{k_1}}{\sum_{\alpha\in\Sigma}\left(\frac{p^{k_2}(x\alpha)}{\sum_{\beta\in\Sigma}p^{k_2}(x\beta)}\right)^{k_1}} = \frac{p^{k_1k_2}(x\sigma)}{\sum_{\beta\in\Sigma}p^{k_1k_2}(x\beta)} = \frac{((k_1k_2) \odot p)(x\sigma)}{((k_1k_2) \odot p)(x)}$$

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

3.2. Construction of a family of inner products

In order to build a framework for generating a family of inner products, a measure space $(\Sigma^\star, 2^{\Sigma^\star}, \mu)$ is constructed, where the finite measure $\mu : 2^{\Sigma^\star} \to [0,1]$ has the following properties.

- $\mu(\Sigma^\star) = 1$;
- $\mu(\bigcup_{k=1}^{\infty}\{x_k\}) = \sum_{k=1}^{\infty}\mu(\{x_k\})$ where $x_k \in \Sigma^\star$.

The second condition in the above statement implies that a non-negative measure is assigned to each singleton string set and the null string ϵ, which are considered to be mutually disjoint measurable sets. Thus, for any collection of strings, $L \in 2^{\Sigma^\star}$ and $\mu(L) = \sum_{x\in L}\mu(\{x\})$.

Given a probability measure p, its conditional version is expressed as $p(\sigma|x) \triangleq p(x\sigma)/p(x)$ for all $\sigma \in \Sigma$ and $x \in \Sigma^\star$, which is another representation of the measure p. This is so because one representation can be recovered from the other by the chain rule of conditional probability.

The conditional version $p(\cdot|\cdot) : \Sigma \times \Sigma^\star \to [0,1]$ is treated as a ($|\Sigma|$-dimensional) vector-valued function for any given string $x \in \Sigma^\star$, which is denoted as $p(\cdot|x) : \Sigma^\star \to [0,1]^{|\Sigma|}$ such that, for every $x \in \Sigma^\star$,

$$p(\cdot|x) = [p(\sigma_1|x), p(\sigma_2|x), \ldots, p(\sigma_{|\Sigma|}|x)] \quad (8)$$

with the constraint $\sum_{j=1}^{|\Sigma|} p(\sigma_j|x) = 1$.

Definition 3.3 (*Probability equivalence*). Given $p_1,p_2 \in \mathcal{P}$, an equivalence relation $\sim$ is defined as: $p_1 \sim p_2$ if $p_1(\cdot|x) = p_2(\cdot|x)$, μ-almost everywhere (a.e.), i.e., if $\mu(\{x \in \Sigma^\star : p_1(\cdot|x) \neq p_2(\cdot|x)\}) = 0$. In this context, a quotient space is defined as $\mathcal{Q} = \mathcal{P}/_\sim$ based on the equivalence relation $\sim$.

Remark 3.3. If the following condition is imposed on the measure μ:

$$\mu(\{x\}) > 0 \quad \forall x \in \Sigma^\star \quad (9)$$

then the condition μ-a.e. in Definition 3.3 becomes μ-everywhere because, in this case, $\mu(\{x\}) \neq \mu(\{y\})\ \forall x \neq y$. That is, the relation $p_1 \sim p_2$ becomes equivalent to $p_1 = p_2$. In other words, $\mathcal{Q} = \mathcal{P}$ provided that Eq. (9) holds.

Proposition 3.2. *$(\mathcal{Q}, \oplus, \odot)$ is a well-defined vector subspace of $\mathcal{P}$.*

Proof. It follows from Definitions 3.1 and 3.2 that, for all $\sigma \in \Sigma$, $x \in \Sigma^\star$ and $k \in \mathbb{R}$,

$$(p_1 \oplus p_2)(\sigma|x) = \frac{p_1(\sigma|x)p_2(\sigma|x)}{\sum_{\alpha\in\Sigma}p_1(\alpha|x)p_2(\alpha|x)} \quad (10)$$

$$(k \odot p_1)(\sigma|x) = \frac{p_1^k(\sigma|x)}{\sum_{\alpha\in\Sigma}p_1^k(\alpha|x)} \quad (11)$$

Both the above equations are consistent under the equivalence relation $\sim$, i.e., $p_1 \sim p_1'$ and $p_2 \sim p_2'$ implies $(p_1 \oplus p_2) \sim (p_1' \oplus p_2')$ and $(k \odot p_1) \sim (k \odot p_1')$. □

Definition 3.4 (*Inner product*). On the vector space $\mathcal{Q}$ over the real field $\mathbb{R}$, a function $\langle \cdot, \cdot \rangle : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ is defined as

$$\langle p_1, p_2 \rangle \triangleq \frac{1}{2}\sum_{\sigma_i,\sigma_j\in\Sigma}\sum_{x\in\Sigma^\star}\log\frac{p_1(x\sigma_i)}{p_1(x\sigma_j)}\log\frac{p_2(x\sigma_i)}{p_2(x\sigma_j)}\mu(\{x\}) \quad (12)$$

The rationale for using the format $\log p_1(x\sigma_i)/p_1(x\sigma_j)$, instead of $\log p_1(x\sigma_i)$, in the definition of inner product in Eq. (12) is that the normalization constants (i.e., the denominators in Eqs. (6) and (7)) may not be the same for all strings $x \in \Sigma^\star$. That is why a division is required to eliminate the effect of this normalization constant.

Remark 3.4. In Eq. (12), the inner summation over $x \in \Sigma^\star$ could be recognized as an integration over $\Sigma^\star$ with the measure μ. In this case, the integration degenerates to a summation since $\Sigma^\star$ is countable.

Theorem 3.2 (*Pre-Hilbert space*). *In Definition3.4, the function $\langle \cdot, \cdot \rangle : \mathcal{Q} \times \mathcal{Q} \to \mathbb{R}$ is an inner product. That is, $(\mathcal{Q}, \oplus, \odot, \langle \cdot, \cdot \rangle)$ forms a pre-Hilbert space over the real field $\mathbb{R}$.*

Proof. While the symmetry property, i.e., $\langle p_1,p_2\rangle = \langle p_2,p_1\rangle$, follows directly from Definition 3.4, positive-definiteness is established as follows:

$$\langle p,p\rangle = \frac{1}{2}\sum_{\sigma_i,\sigma_j\in\Sigma}\sum_{x\in\Sigma^\star}\left(\log\frac{p(x\sigma_i)}{p(x\sigma_j)}\right)^2\mu(\{x\})\geq 0 \tag{13}$$

If $\langle p,p\rangle = 0$, non-negativity of each term in the summation mandates that, for μ-almost every $x\in\Sigma^\star$, $\log p(x\sigma_i)/p(x\sigma_j)=0$. Therefore, $p(\sigma_i|x)=1/|\Sigma|\ \forall\sigma_i\in\Sigma$, and it follows that $p\sim\underline{e}$.

The linearity property is established as follows:

$$\begin{aligned}\langle a\odot p_1,p_2\rangle &= \frac{1}{2}\sum_{\sigma_i,\sigma_j\in\Sigma}\sum_{x\in\Sigma^\star}\log\frac{(a\odot p_1)(x\sigma_i)}{(a\odot p_1)(x\sigma_j)}\log\frac{p_2(x\sigma_i)}{p_2(x\sigma_j)}\mu(\{x\})\\ &= \frac{1}{2}\sum_{\sigma_i,\sigma_j\in\Sigma}\sum_{x\in\Sigma^\star}\log\frac{p_1^a(x\sigma_i)}{p_1^a(x\sigma_j)}\log\frac{p_2(x\sigma_i)}{p_2(x\sigma_j)}\mu(\{x\})\\ &= \frac{a}{2}\sum_{\sigma_i,\sigma_j\in\Sigma}\sum_{x\in\Sigma^\star}\log\frac{p_1(x\sigma_i)}{p_1(x\sigma_j)}\log\frac{p_2(x\sigma_i)}{p_2(x\sigma_j)}\mu(\{x\})\\ &= a\langle p_1,p_2\rangle\end{aligned} \tag{14}$$

and

$$\begin{aligned}\langle p_1\oplus p_2,p_3\rangle &= \frac{1}{2}\sum_{\sigma_i,\sigma_j\in\Sigma}\sum_{x\in\Sigma^\star}\log\frac{(p_1\oplus p_2)(x\sigma_i)}{(p_1\oplus p_2)(x\sigma_j)}\log\frac{p_3(x\sigma_i)}{p_3(x\sigma_j)}\mu(\{x\})\\ &= \frac{1}{2}\sum_{\sigma_i,\sigma_j\in\Sigma}\sum_{x\in\Sigma^\star}\log\frac{p_1(x\sigma_i)p_2(x\sigma_i)}{p_1(x\sigma_j)p_2(x\sigma_j)}\log\frac{p_3(x\sigma_i)}{p_3(x\sigma_j)}\mu(\{x\})\\ &= \frac{1}{2}\sum_{\sigma_i,\sigma_j\in\Sigma}\sum_{x\in\Sigma^\star}\left(\log\frac{p_1(x\sigma_i)}{p_1(x\sigma_j)}+\log\frac{p_2(x\sigma_i)}{p_2(x\sigma_j)}\right)\\ &\quad\cdot\log\frac{p_3(x\sigma_i)}{p_3(x\sigma_j)}\mu(\{x\})\\ &= \langle p_1,p_3\rangle+\langle p_2,p_3\rangle\end{aligned} \tag{15}$$

This completes the proof. □

Remark 3.5. Since the inner product $\langle\cdot,\cdot\rangle$ is defined on the real field $\mathbb{R}$, the inner product space $\mathcal{Q}$ is a collection of finite-norm probability measures, where the norm is induced by the inner product. The measure μ in Eq. (12) is user-selectable such that any valid choice of the finite measure μ yields an inner product.

3.3. An alternative norm

The inner product in Definition 3.4 is constructed with the motivation of orthogonal projection. In addition to the norm induced by the inner product, there exist many other norms for this vector space $\mathcal{P}$. For example, the following supremum norm is suitable for construction of a Banach space.

Definition 3.5 (*Subspace* $\mathcal{P}_\infty$). The subspace $\mathcal{P}_\infty$ of the vector space $\mathcal{P}$ is defined as

$$\mathcal{P}_\infty = \left\{p\in\mathcal{P} : \sup_{x\in\Sigma^\star}\log\left(\frac{p(x\tau_{max})}{p(x\tau_{min})}\right)<\infty\right\} \tag{16}$$

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 3.3 (*Supremum norm*). *A function* $\|\cdot\|_\infty : \mathcal{P}_\infty\to[0,\infty)$ *defined as*

$$\|p\|_\infty = \sup_{x\in\Sigma^\star}\log\left(\frac{p(x\tau_{max})}{p(x\tau_{min})}\right) \tag{17}$$

is a norm on the vector space $\mathcal{P}_\infty$.

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

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

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

  and for $k<0$,

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

- Triangular inequality:

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

  Therefore, $\|p_1\oplus p_2\|_\infty\leq(\|p_1\|_\infty+\|p_2\|_\infty)$.

The proof is now complete. □

Remark 3.6. The supremum norm $\|\cdot\|_\infty$ has a simpler mathematical structure than the inner product $\langle\cdot,\cdot\rangle$, primarily because the equivalence relation in Definition 3.3 and the construction of the quotient space are not needed. Also, $\|\cdot\|_\infty$ does not depend on the measure μ, which is suitable for some applications which may not require an inner product.

4. Stochastic regular languages and PFSA

Although probability measures over the σ-algebra $\mathcal{B}_\Sigma$ adequately describe stochastic languages, PFSA representations are more compact and usable for modeling of stochastic regular languages in many applications. However, the mapping $\tilde{\mathbb{H}} : \tilde{\mathcal{G}}\to\mathcal{P}$ in Definition 2.8, which was originally introduced in the prior work [30], is neither an injection nor a surjection (see Remark 2.5). In this section, an isometric isomorphism is constructed between a quotient space of $\tilde{\mathcal{G}}$ and a subspace of a quotient space of $\mathcal{P}$.

Definition 4.1 (*Subspace $\mathcal{Q}_f$*). The set $\mathcal{Q}_f$ is defined to be a subset of the quotient space $\mathcal{Q}$ (see Definition 3.3) having measures with a finite number of Nerode equivalence classes, i.e., $\mathcal{Q}_f \triangleq \{p \in \mathcal{Q} : |\mathbf{N}_p| < \infty\}$.

The following proposition establishes that $\mathcal{Q}_f$ is a vector subspace of $\mathcal{Q}$.

Proposition 4.1. *Let $p_1,p_2 \in \mathcal{Q}_f$ and $x,y \in \Sigma^\star$. Then, following conditions hold*:

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

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

Proof. Let $u_n = \tau_1\tau_2\ldots\tau_n \in \Sigma^\star$ and $p_3 = p_1 \oplus p_2$ where $\tau_i \in \Sigma$. To prove the first identity of the proposition, it will be shown that $p_3(z_1u_n)/p_3(z_1) = p_3(z_2u_n)/p_3(z_2)$ for any $u_n \in \Sigma^\star$. This is achieved by the method of induction:

$$\frac{p_3(z_1u_1)}{p_3(z_1)} = \frac{p_1(z_1u_1)p_2(z_1u_1)}{\sum_{\alpha\in\Sigma} p_1(z_1\alpha)p_2(z_1\alpha)} = \frac{\frac{p_1(z_1u_1)}{p_1(z_1)}\frac{p_2(z_1u_1)}{p_2(z_1)}}{\sum_{\alpha\in\Sigma}\frac{p_1(z_1\alpha)}{p_1(z_1)}\frac{p_2(z_1\alpha)}{p_2(z_1)}}$$
$$= \frac{\frac{p_1(z_2u_1)}{p_1(z_2)}\frac{p_2(z_2u_1)}{p_2(z_2)}}{\sum_{\alpha\in\Sigma}\frac{p_1(z_2\alpha)}{p_1(z_2)}\frac{p_2(z_2\alpha)}{p_2(z_2)}} = \frac{p_3(z_2u_1)}{p_3(z_2)}$$

For the inductive step,

$$\frac{p_3(z_1u_{n+1})}{p_3(z_1)} = \frac{p_3(z_1u_n)}{p_3(z_1)}\frac{p_3(z_1u_{n+1})}{p_3(z_1u_n)}$$
$$= \frac{p_3(z_2u_n)}{p_3(z_2)}\frac{p_3(z_2u_{n+1})}{p_3(z_2u_n)} = \frac{p_3(z_2u_{n+1})}{p_3(z_2)}$$

The second identity of the proposition can be derived in the same way. □

Definition 4.2 (*PFSA equivalence*). Following Definition 3.3, two PFSA $\tilde{G}$ and G are said to be equivalent if $\tilde{\mathbb{H}}(\tilde{G}) \sim \tilde{\mathbb{H}}(G)$. The equivalence class of G is denoted as: $\Xi(G) \triangleq \{\tilde{G} \in \mathcal{G} : \tilde{\mathbb{H}}(\tilde{G}) \sim \tilde{\mathbb{H}}(G)\}$. Defining a quotient space $\mathcal{G} \triangleq \tilde{\mathcal{G}}/\Xi$, the associated quotient map is obtained as

$$\mathbb{H} : \mathcal{G} \longrightarrow \mathcal{Q}_f \tag{18}$$

The quotient map $\mathbb{H}$ in Eq. (18) is injective by this quotient space construction and it is also surjective by the application of probabilistic Nerode equivalence (see Remark 2.5). Hence, $\mathbb{H}$ is a bijection and the associated inverse map is denoted by $\mathbb{F}$, i.e., $\mathbb{F} \triangleq \mathbb{H}^{-1}$.

Remark 4.1. Given a PFSA $G = (Q,\Sigma,\delta,q_0,\tilde{\pi}) \in \mathcal{G}$, each state $q \in Q$ is a Nerode equivalence class $S \in \mathbf{N}_{\mathbb{H}(G)}$, where $\mathbf{N}_{\mathbb{H}(G)}$ is the set of all Nerode equivalence classes of the measure $\mathbb{H}(G)$ (see Remark 2.5). By Algorithm 1, it follows that

$$\tilde{\pi}(q,\sigma) = p(\sigma|S) \triangleq \frac{p(x\sigma)}{p(x)}, \quad \forall x \in S \tag{19}$$

So far the vector space $(\mathcal{Q}_f, \oplus, \odot)$ is established. New operations of vector addition and scalar multiplication are introduced on the quotient space $\mathcal{G}$ by use of the bijection $\mathbb{H}$ and its inverse $\mathbb{F}$.

Definition 4.3 (*Vector space $\mathcal{G}$*). Let $G_1,G_2 \in \mathcal{G}$ and $k \in \mathbb{R}$. Then,

- The addition $+ : \mathcal{G} \times \mathcal{G} \to \mathcal{G}$ is defined as a homomorphism
  $G_1 + G_2 = \mathbb{F}(\mathbb{H}(G_1) \oplus \mathbb{H}(G_2))$
- The (scalar) multiplication $\cdot : \mathbb{R} \times \mathcal{G} \to \mathcal{G}$ is defined as a homomorphism
  $k \cdot G_1 = \mathbb{F}(k \odot (\mathbb{H}(G_1))$

Since $\mathbb{F} \triangleq \mathbb{H}^{-1}$, it follows from Definition 4.3 that $\mathbb{H}(G_1 + G_2) = \mathbb{H}(G_1) \oplus \mathbb{H}(G_2)$ and $\mathbb{H}(k \cdot G_1) = k \odot \mathbb{H}(G_1)$. Therefore, the bijection $\mathbb{H}$ is linear and hence $\mathbb{H}$ is an isomorphism between the vector spaces $(\mathcal{Q}_f, \oplus, \odot)$ and $(\mathcal{G}, +, \cdot)$. Similarly, the map $\mathbb{H}$ is used to define the inner product on the space $\mathcal{G}$ in terms of $\langle \cdot, \cdot \rangle$.

Definition 4.4 (*Isometric isomorphism between $\mathcal{G}$ and $\mathcal{Q}_f$*). The inner product $\langle \cdot, \cdot \rangle_A : \mathcal{G} \times \mathcal{G} \to \mathbb{R}$ is defined as

$$\langle G_1, G_2 \rangle_A = \langle \mathbb{H}(G_1), \mathbb{H}(G_2) \rangle \tag{20}$$

Consequently, the quotient map $\mathbb{H} : \mathcal{G} \longrightarrow \mathcal{Q}_f$ in Eq. (18) becomes an isometric isomorphism between the two pre-Hilbert spaces.

The pre-Hilbert space of PFSA may not be complete because a Cauchy sequence of PFSA may converge to a machine with infinite number of states. In many applications, however, PFSA models could be restricted to finite-dimensional subspaces if completeness is a crucial issue (e.g., projection in a Hilbert space setting). Section 7 presents such an example.

In the sequel, it is understood that any operations, defined for probability measures or PFSA, could be translated into the other space using this isomorphism.

Since Definitions 4.3 and 4.4 do not specify an efficient way of computing the algebraic operations in terms of PFSA, the notion of structural similarity is introduced to address this issue.

Definition 4.5 (*Structural similarity*). Two PFSA $G_i \triangleq (Q_i, \Sigma, \delta_i, q_0^i, \tilde{\pi}_i) \in \mathcal{G}$, $i = 1,2$, 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 $\forall \sigma \in \Sigma$.

From the prospective of a graphical representation, if two PFSA of the same structure have the same underlying graph connectivity, then they may differ only in the arc probabilities on the graph as seen in the following proposition.

Proposition 4.2. *Let two PFSA $G_1,G_2 \in \mathcal{G}$ be structurally similar in sense of Definition4.5, i.e., $G_i = (Q,\Sigma,\delta,q_0,\tilde{\pi}_i)$,*

$i \in \{1,2\}$. *If* $G \triangleq (Q,\Sigma,\delta,q_0,\tilde{\pi})$, *where*

$$\tilde{\pi}(q,\sigma) = \frac{\tilde{\pi}_1(q,\sigma)\tilde{\pi}_2(q,\sigma)}{\sum_{\alpha\in\Sigma}\tilde{\pi}_1(q,\alpha)\tilde{\pi}_2(q,\alpha)} \tag{21}$$

then $G \in \mathcal{G}$ *and* $G = G_1 + G_2$.

Proof. Let $p_i = \mathbb{H}(G_i)$, $i \in \{1,2\}$. Since G_1 and G_2 are structurally similar, it follows that

$$\frac{p_i(x\sigma)}{p_i(x)} = \tilde{\pi}_i(\delta^\star(q_0,x),\sigma) = \tilde{\pi}_i(q,\sigma), \quad i \in \{1,2\}$$

for all strings x in state $q \in Q$ and all $\sigma \in \Sigma$. By Definitions 3.1 and 2.8, it follows that

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

This implies that $p_1 \oplus p_2 = \mathbb{H}(G_1 + G_2) = \mathbb{H}(G)$ and the proof is complete. □

If G_1 and G_2 have different structures, then synchronous composition can be used to perform the structural transform first. Following Propositions 4.2 and 2.1, the addition operation on any two PFSA G_1 and G_2 is performed as

$$G_1 + G_2 = (G_1 \otimes G_2) + (G_2 \otimes G_1) \tag{22}$$

Proposition 4.3. *Given a PFSA* $G \triangleq (Q,\Sigma,\delta,q_0,\tilde{\pi}) \in \mathcal{G}$ *and* $k \in \mathbb{R}$, *if* $\tilde{G} \triangleq (Q,\Sigma,\delta,q_0,\tilde{\pi}')$, *where*

$$\tilde{\pi}'(q,\sigma) = \frac{(\tilde{\pi}(q,\sigma))^k}{\sum_{\alpha\in\Sigma}(\tilde{\pi}(q,\alpha))^k} \tag{23}$$

then $\tilde{G} = k \cdot G$.

Proof. By Proposition 4.1, the scalar multiplication by k does not change the structure of G and therefore the state transition function δ and the initial state q_0 also remain unchanged. Denoting $p = \mathbb{H}(G)$, it follows from Eq. (5) that

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

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

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

Therefore, $k \odot p = \mathbb{H}(\tilde{G})$. □

The following result is obtained by using Algorithm 1 for computing the inner product $\langle \cdot,\cdot \rangle_A$ in Definition 4.4 as described below.

Proposition 4.4. *Let* $G_i = (Q,\Sigma,\delta,q_0,\tilde{\pi}_i) \in \mathcal{G}$, $i=1,2$. *The following inner product is computed as*

$$\langle G_1,G_2 \rangle_A = \frac{1}{2}\sum_{\sigma_i,\sigma_j\in\Sigma}\sum_{q\in Q}\log\frac{\tilde{\pi}_1(q,\sigma_i)}{\tilde{\pi}_1(q,\sigma_j)}\log\frac{\tilde{\pi}_2(q,\sigma_i)}{\tilde{\pi}_2(q,\sigma_j)}\mu(q) \tag{24}$$

The computation of the measure $\mu(q)$ is addressed later in Section 6. Note that, if G_1 and G_2 do not have the same structure, the synchronous composition should be used to compute the inner product via

$$\langle G_1,G_2 \rangle_A = \langle G_1 \otimes G_2, G_2 \otimes G_1 \rangle_A \tag{25}$$

The corresponding norm is defined by this inner product as

$$\|G\|_A = \sqrt{\langle G,G \rangle_A} \tag{26}$$

5. Interpretation of algebraic operations

This section interprets the significance of the algebraic operations in the vector space of the probability measures and presents an analogy of the norm $\|\cdot\|$ to the entropy rate in the setting of information theory [24]. In Definition 3.1, the $\oplus$ operation of vector addition is performed via elementwise multiplication of the morph matrix. In Definition 3.2, the $\odot$ operation of scalar multiplication is computed elementwise by taking the power of the morph matrix. These properties are largely similar to those of the logarithm function. Therefore, by taking logarithms, these operations could be made analogous to the usual vector addition and scalar multiplication in the Euclidean space; however, they are not exactly the same due to the additional step of normalization. This analogy suggests the potential use of this technique for sensor data analysis. For example, let a sensor signal be contaminated with a multiplicative noise in the Euclidean space. By taking logarithm of the sensor reading and by appropriate scaling, a linear representation could be obtained in the form of additive noise in the $\mathcal{G}$-domain (see Definition 4.2).

The zero vector $\underline{e}$, called *symbolic white noise*, in the vector space corresponds to the uniform distribution on $\mathcal{B}_\Sigma$ and is perfectly encoded by the PFSA $E \in \mathcal{G}$ that is expressed as

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

where $\delta(q,q) = q$ and $\tilde{\pi}(q,\sigma) = 1/|\Sigma|, \forall\sigma \in \Sigma$. 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 model for *symbolic white noise* in a dynamical system, because no additional information is provided through vector addition of E to any PFSA.

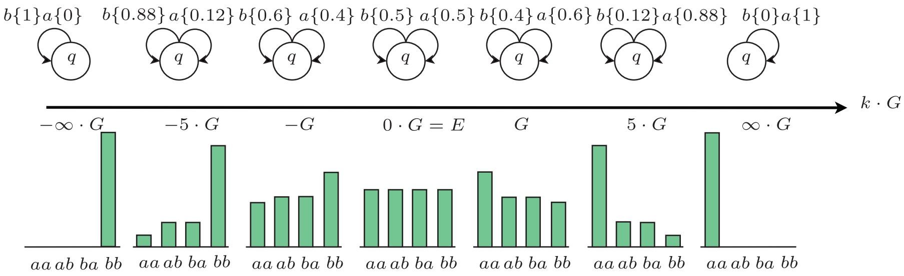


Fig. 3. Scalar multiplication of the one-state PFSA G.

Multiplication by a scalar k relates to reshaping the probability distribution p over the alphabet $\mathcal{B}_\Sigma$. The effects of $k \in \mathbb{R}$ on an i.i.d. process G are illustrated in Fig. 3, where $\Sigma = \{a,b\}$. As k is increased from 0 to $+\infty$, the string with the highest probability (e.g., the string bb in the extreme left bottom and the string aa in the extreme right bottom of Fig. 3) stands out in the histogram and the distribution p approaches the delta distribution. Similarly, as k is decreased from $+\infty$ to 0, the distribution p becomes more and more flat and finally reaches $\underline{e}$ that is the uniform distribution at $k=0$. This behavior of a positive scalar k is analogous to the inverse temperature β in the setting of Statistical Mechanics [32]. In the latter case, $\beta = 0$ yields the uniform distribution over the energy states while the distribution tends to a delta distribution as β approaches infinity. In contrast to a positive k, a negative k favors the least probable string in G (e.g., the string bb in Fig. 3). It inverts the distribution p in the sense that a string originally with large probability now has small probability of occurring and vice versa. Similarly, another delta distribution is achieved as $k \to -\infty$. The constructed vector space $\mathcal{G}$ is related to the Euclidean space as explained below.

To interpret the possible meaning of the norm $\|\cdot\|_A$ in Eq. (26), the entropy rate of a PFSA G is given in the setting of Information Theory [24] as

$$h(G) = -\sum_{q \in Q} \wp(q) \left[\sum_{\sigma \in \Sigma} \tilde{\pi}(q,\sigma) \log \tilde{\pi}(q,\sigma) \right] \tag{27}$$

while, in the present formulation, the norm of the PFSA G is induced by the inner product in Eq. (24) as

$$\|G\|_A \triangleq \sqrt{\sum_{q \in Q} \left[\frac{\mu(q)}{2} \sum_{\sigma_i,\sigma_j \in \Sigma} (\log \tilde{\pi}(q,\sigma_i) - \log \tilde{\pi}(q,\sigma_j))^2 \right]} \tag{28}$$

Eqs. (27) and (28) have structural similarity in the sense that both are represented as a weighted sum over the states. However, two major differences are as follows:

1. For each state $q \in Q$, a weight $\mu(q)$ is used in Eq. (28) instead of the stationary probability $\wp(q)$ in Eq. (27).
2. The root mean square (rms) difference of logarithm of the probabilities of a pair of symbols conditioned on each state is used instead of the expectation of logarithm of a symbol's conditional probability. This rms value in the norm is a consequence of the inner product.

In contrast to the entropy rate, which is a measure of the uncertainty, the ideal deterministic symbolic system should have the maximum norm while the completely random process should have a zero norm. For 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 symbol a and the symbol b be θ and $(1-\theta)$, respectively, with $\theta \in (0,1)$.

Fig. 4(a) compares $(1-h(G))$ (solid line) and $(2/\pi)\tan^{-1}(\|G\|_A)$ (dashed line), where the entropy rate $h(G)$ ranges in $[0,1]$ and the range of the norm $\|G\|_A$ is normalized from $[0,\infty)$ to $[0,1]$ by using the arc tangent function. It is observed that the profiles for $(1-h(G))$ and $(2/\pi)\tan^{-1}(\|G\|_A)$ are qualitatively similar. Hence, by drawing an analogy, it is possible to interpret the norm in Eq. (28) 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 $\tilde{P}$ and P, respectively. Then, a diversity between $\tilde{G}$ and G is defined as the Kullback–Leibler (K–L) divergence [24] of $\tilde{P}$ and P:

$$\mathbb{D}(\tilde{G}\|G) \triangleq \sum_i \tilde{P}(i) \log \frac{\tilde{P}(i)}{P(i)} \tag{29}$$

Setting the PFSA $\tilde{G}$ to the *symbolic white noise* E, Fig. 4(b) compares a distance $\rho(E,G) \triangleq \|E-G\|_A$ (dash-dot line), the K–L divergence $\mathbb{D}(E\|G)$ (solid line), and the K–L divergence $\mathbb{D}(G\|E)$ (dashed line) versus the probability parameter θ. It is seen that these three curves are qualitatively very similar as all of them approach infinity when θ approaches 0 or 1 and achieve the minimum at 0 if $\theta = 0.5$. An advantage of the proposed measure is that ρ is a metric but K–L divergence is not.

With the inner product defined in Eq. (24), the correlation $\gamma : \mathcal{G} \times \mathcal{G} \to [-1,1]$ between two PFSA is defined in terms of the normalized inner product as

$$\gamma(G_1,G_2) = \frac{\langle G_1, G_2 \rangle_A}{\|G_1\|_A \cdot \|G_2\|_A}. \tag{30}$$

If the vectors G_1 and G_2 are perfectly negatively correlated, i.e., $G_1 = -G_2$, then their vector addition is exactly the zero vector $\underline{e}$ (i.e., the *symbolic white noise*) that represents the uniform distribution.

6. Computation of the measure μ

Computation in Eq. (12) depends on the choice of the measure μ. The measure of all strings in the equivalence

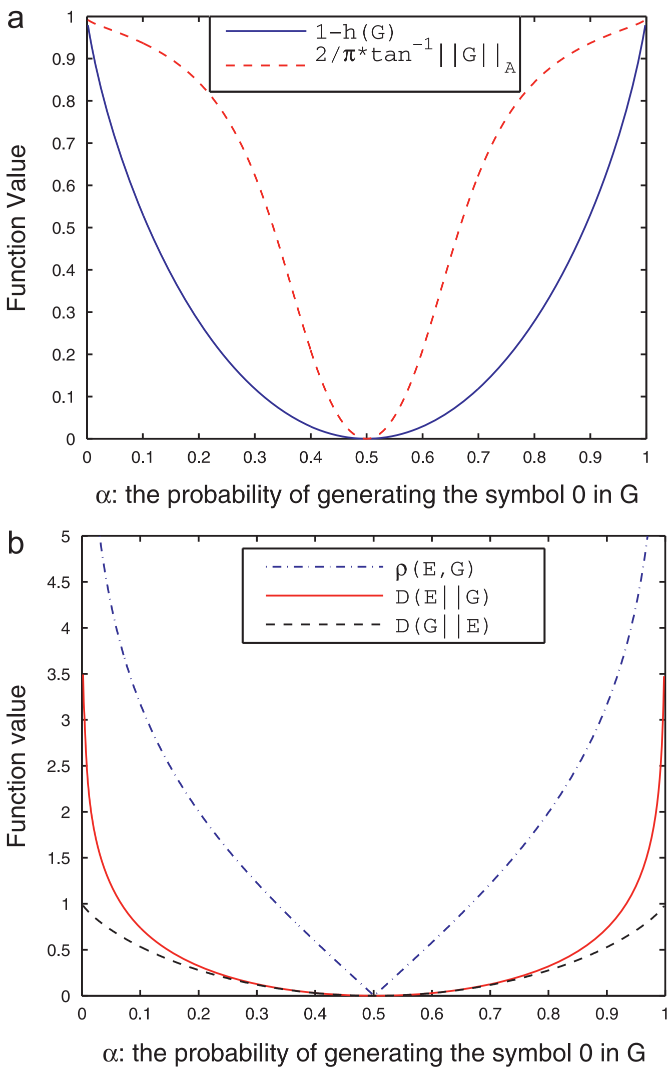


Fig. 4. Metric comparison: norm and information-theoretic metrics. (a) $1-h(G)$ and $\|G\|_A$ of an i.i.d. process G, (b) comparison of $\rho(E,G_2)$, $\mathbb{D}(E\|G_2)$, and $\mathbb{D}(G_2\|E)$ versus parameter α with E being the *Symbolic white noise*.

class represented by a state $q \in Q$ of a PFSA is obtained as $\mu(q) = \sum_{x \in q} \mu(\{x\})$. In general, for a countable summation over the strings, the convergence is not guaranteed. This section presents a few common choices of μ and explains how to compute them.

Let $0 < \theta < 1$ and let μ_1 and μ_2 be two measures defined as follows:

$$\mu_1(\{x\}) = (1-\theta) \cdot \left(\frac{\theta}{|\Sigma|}\right)^{|x|} \tag{31}$$

$$\mu_2(\{x\}) = (1-\theta)^2 \cdot \frac{|x|\theta^{|x|-1}}{|\Sigma|^{|x|}} \tag{32}$$

Since both measures μ_1 and μ_2 depend only on the length of the symbol string, it implies that strings of the same length are identical under a given measure. Accordingly, a smaller measure is assigned to a longer string, because the probability of its occurrence is small. Both measures μ_1 and μ_2 decay exponentially as the string length $|x|$ grows and the rate is specified by the parameter θ. As the parameter θ is increased, a larger number of longer strings contributes to the inner product. However, a major difference between the measures μ_1 and μ_2 lies in the fact that $\mu_1(\epsilon) = (1-\theta) \neq 0$ and $\mu_2(\epsilon) = 0$, where ϵ is the null string with $|\epsilon| = 0$. Since the transition from the initial state q_0 to itself is represented by ϵ, it would be logical to assign a non-zero measure $\mu_1(\epsilon)$ if q_0 is significant in the PFSA model; alternatively, assigning $\mu_2(\epsilon) = 0$ puts no significance to q_0. The following definitions are introduced to compute the measures μ_1 and μ_2 for each state of a given PFSA.

Definition 6.1 (*Dependence of measure on string length*). Let the map $m_n : 2^{\Sigma^\star} \to [0,1]$ be defined as

$$m_n(L) \triangleq \frac{|\{x \in L : |x| = n\}|}{|\Sigma|^n} \quad \forall L \subseteq \Sigma^\star \tag{33}$$

Remark 6.1. The function $m_n(L)$ in Definition 6.1 is the ratio of the number of strings of length n in the set L to the total number of strings of length n in $\Sigma^\star$. It represents the size of the set L in terms of strings of length n.

Definition 6.2 (*Uniformizer of PFSA*). Given a PFSA $G=(Q,\Sigma,\delta,q_0,\tilde{\pi})$, the PFSA G' is called the uniformizer of G if $G'=(Q,\Sigma,\delta,q_0,\tilde{\pi}')$, where $\tilde{\pi}'(q,\sigma)=1/|\Sigma|,\forall q\in Q$, $\forall\sigma\in\Sigma$.

The uniformizer of a PFSA G is denoted by $\mathbb{U}(G)$, which simply modifies the original probability morph function to a uniform distribution over the symbols at each state. Note that $\mathbb{U}(G)$ retains the graph connectivity of G.

Proposition 6.1 (*State transition matrix for Uniformizer*). *Let* $G=(Q,\Sigma,\delta,q_0,\tilde{\pi}_G)$ *be a PFSA and*

$$\boldsymbol{m}_n \triangleq \boldsymbol{m}_0(\Pi^{\mathbb{U}(G)})^n \tag{34}$$

where $\boldsymbol{m}_n=[m_n(q_1),m_n(q_2),\dots,m_n(q_{|Q|})]$. *Then,* $\Pi^{\mathbb{U}(G)}$ *is the state transition matrix for the uniformizer* $\mathbb{U}(G)$, *and*

$$\boldsymbol{m}_0(q)=\begin{cases}1 & \text{if } q=q_0\\ 0 & \text{if } q\neq q_0\end{cases}$$

Proof. For any $q_i\in Q$ and $n\in\mathbb{N}$, it follows that

$$|\Sigma|^{n+1}m_{n+1}(q_i)=|\{x\in q_i:|x|=n+1\}|=|\Sigma|^n\sum_{\delta(q_j,\sigma)=q_i}m_n(q_j)$$

Then,

$$m_{n+1}(q_i)=\frac{1}{|\Sigma|}\sum_{\delta(q_j,\sigma)=q_i}m_n(q_j)=\sum_{\delta(q_j,\sigma)=q_i}\tilde{\pi}^{\mathbb{U}(G)}(q_j,\sigma)m_n(q_j) \tag{35}$$

Following Definition 2.2, a matrix representation of Eq. (35) is obtained as

$$\boldsymbol{m}_{n+1}=\boldsymbol{m}_n\Pi^{\mathbb{U}(G)} \tag{36}$$

from which Eq. (34) follows. This completes the proof. □

Let $f_\theta(q)=\sum_{i=0}^\infty m_i(q)\cdot\theta^i$ for $\theta\in(0,1)$. Given a PFSA $G=(Q,\Sigma,\delta,q_0,\tilde{\pi})\in\mathcal{G}$, it follows that if $\boldsymbol{f}_\theta\triangleq[f_\theta(q_1), f_\theta(q_2),\dots,f_\theta(q_{|Q|})]$, then

$$\boldsymbol{f}_\theta=\sum_{i=0}^\infty\boldsymbol{m}_i\cdot\theta^i \tag{37}$$

and an application of Eq. (34) yields

$$\boldsymbol{f}_\theta=\boldsymbol{m}_0\sum_{i=0}^\infty(\theta\cdot\Pi^{\mathbb{U}(G)})^i=\boldsymbol{m}_0(I-\theta\cdot\Pi^{\mathbb{U}(G)})^{-1} \tag{38}$$

The last step is valid since $\|\theta\cdot\Pi^{\mathbb{U}(G)}\|_\infty<1$.

Proposition 6.2. *Let G be a PFSA. Then, the following two measures are valid.* $\boldsymbol{\mu}_i\triangleq[\mu_i(q_1),\mu_i(q_2),\dots,\mu_i(q_{|Q|})]$ *where* $i\in\{1,2\}$. *Then,*

(1) $\boldsymbol{\mu}_1=\boldsymbol{m}_0(1-\theta)(I-\theta\cdot\Pi^{\mathbb{U}(G)})^{-1}$;
(2) $\boldsymbol{\mu}_2=(1-\theta)^2\boldsymbol{m}_0(I-\theta\cdot\Pi^{\mathbb{U}(G)})^{-1}\Pi^{\mathbb{U}(G)}\cdot(I-\theta\cdot\Pi^{\mathbb{U}(G)})^{-1}$.

Proof.

(1) For any $q\in Q$, we have

$$\mu_1(q)\triangleq\sum_{x\in q}\mu_1(\{x\})=\sum_{k=0}^\infty(m_k(q)|\Sigma|^k)(1-\theta)\cdot\left(\frac{\theta}{|\Sigma|}\right)^k=(1-\theta)\sum_{k=0}^\infty m_k(q)\theta^k=(1-\theta)f_\theta(q) \tag{39}$$

It follows from Eq. (38) that

$$\boldsymbol{\mu}_1=\boldsymbol{m}_0(1-\theta)(I-\theta\cdot\Pi^{\mathbb{U}(G)})^{-1} \tag{40}$$

(2) Similarly for μ_2, it follows that

$$\mu_2(q)=\sum_{k=1}^\infty(m_k(q)|\Sigma|^k)(1-\theta)^2\cdot\frac{|x|\cdot\theta^{|x|-1}}{|\Sigma|^k}=(1-\theta)^2\sum_{k=1}^\infty m_k(q)\cdot k\cdot\theta^{k-1}=(1-\theta)^2\frac{df_\theta(q)}{d\theta} \tag{41}$$

$$\frac{d\boldsymbol{f}_\theta}{d\theta}=\frac{1}{\theta^2}\sum_{k=1}^\infty\boldsymbol{m}_k\cdot k\cdot\theta^{k+1} \tag{42}$$

Thus,

$$\boldsymbol{\mu}_2=(1-\theta)^2\frac{d\boldsymbol{f}_\theta}{d\theta} \tag{43}$$

Since the convergence regions of $d\boldsymbol{f}_\theta/d\theta$ and $\boldsymbol{f}_\theta$ are the same, μ_2 converges. The fact that $dA^{-1}/dt=-A^{-1}(dA/dt)A^{-1}$ for an invertible matrix A (that is dependent on a parameter t) leads to the following result:

$$\boldsymbol{\mu}_2=(1-\theta)^2\boldsymbol{m}_0(I-\theta\cdot\Pi^{\mathbb{U}(G)})^{-1}\Pi^{\mathbb{U}(G)}(I-\theta\cdot\Pi^{\mathbb{U}(G)})^{-1} \quad \square \tag{44}$$

This section is concluded with a theorem that addresses PFSA-based modeling of stochastic regular languages.

Theorem 6.1 (*Approximation of stochastic languages*). *The subspace* $\mathcal{Q}_f$ *is dense in the inner product space* $(\mathcal{Q},\oplus,\odot,\langle\cdot,\cdot\rangle)$.

Proof. Let us consider any $p\in\mathcal{Q}$. By Remark 3.5, it follows that

$$\langle p,p\rangle=\frac{1}{2}\sum_{\sigma_i,\sigma_j\in\Sigma}\sum_{x\in\Sigma^\star}\left(\log\frac{p(x\sigma_i)}{p(x\sigma_j)}\right)^2\mu(\{x\})<\infty$$

This implies that the infinite tail of the above sum must converge to zero. That is, $\forall\varepsilon>0$, there exists $N(\varepsilon)\in\mathbb{N}$ such that

$$S_N\triangleq\frac{1}{2}\sum_{\sigma_i,\sigma_j\in\Sigma}\sum_{|x|\geq N}\left(\log\frac{p(x\sigma_i)}{p(x\sigma_j)}\right)^2\mu(\{x\})<\varepsilon$$

We now define another measure p' such that

$$p'(x)=\begin{cases}p(x) & \text{if } |x|\leq N\\ p(pref_N(x))\cdot\frac{1}{|\Sigma|^{|x|-N}} & \text{if } |x|>N\end{cases}$$

where $pref_N(x)$ means the prefix of length N of the string x. The number of the Nerode equivalence class of p' is at most $\sum_{i=0}^{N}|\Sigma|^{i}+1$ since any strings of length more than N are equivalent to each other and therefore $p' \in \mathcal{Q}$. The difference between p and p' is given as

$$(p \oplus (-p'))(x) = \begin{cases} \underline{e}(x) & \text{if } |x| \le N \\ \dfrac{p(x)}{|\Sigma|^{N} \cdot p(pref_N(x))} & \text{if } |x| > N \end{cases}$$

and this implies

$$\|p \oplus (-p')\|^2 = S_N < \varepsilon$$

This proves that $\mathcal{Q}_f$ is dense in $\mathcal{Q}$. □

The implication of Theorem 6.1 is that any stochastic language with a finite norm can be approximated as closely as desired by a PFSA model. In other words, it is always possible to arbitrarily reduce the modeling error by increasing the number of states of the PFSA model. Therefore, by restricting the symbolic system model to PFSA, there is no significant loss of modeling power in the sense of the metric defined on the vector space.

7. Examples for concept validation

The section presents three examples. The first and second examples, respectively, illustrate the numerical steps in the computation of an inner product and an orthogonal projection from a Hilbert space of PFSA onto a (closed) Markov subspace. Orthogonal projection provides optimization in the Hilbert space setting, which is useful for diverse applications in signal representation and model identification. The third example is based on experimental data and demonstrates the efficacy of the proposed PFSA-based tool for model order reduction.

The space $\mathcal{G}$ of PFSA is not complete because a Cauchy sequence of PFSA with an increasingly number of states will have the limit point that is not a finite state machine (e.g., similar to the space of polynomials). However, in practice, a finite-dimensional subspace could be adequate for feature extraction from symbolic sequences. For example, finite-order D-Markov machines have been used for anomaly detection in polycrystalline alloys [33] and in electronic systems [14]. Since all finite-dimensional vector spaces over the real field $\mathbb{R}$ are guaranteed to be complete [27], any closed finite subspace of $\mathcal{G}$ is a well-defined Hilbert space. Therefore, an inner product defined on a finite-dimensional subspace of the space $\mathcal{G}$ admits orthogonal projections on closed subspaces with guaranteed existence and uniqueness.

Let $\mathbb{P}_{\mathcal{G}_2}: \mathcal{G}_1 \to \mathcal{G}_2$ denote the orthogonal projection from a finite-dimensional subspace $\mathcal{G}_1$ of the space $\mathcal{G}$ onto another smaller closed subspace of $\mathcal{G}_2 \subset \mathcal{G}_1$. If $\{V_i\}_{i=1}^{n}$ is an orthonormal basis for the space $\mathcal{G}_2$, where $n = \dim(\mathcal{G}_2) \in \mathbb{N}$, then it follows that

$$\mathbb{P}_{\mathcal{G}_2}(G) = \sum_{i=1}^{n} \langle G, V_i \rangle_A \cdot G \tag{45}$$

The error due to projection onto the smaller dimensional space $\mathcal{G}_2$ is obtained as $\|G - \mathbb{P}_{\mathcal{G}_2}(G)\|_A$.

Finite-dimensional subspaces spanned by D-Markov machines [14] (that belong to a class of shifts of finite type [25]) have been used for system identification and anomaly detection in diverse applications [14,33]. Let the D-Markov subspace with depth d be denoted as $\mathcal{D}_d$, where d is a positive integer. The rationale for this choice is that D-Markov machines have a clearly defined physical meaning for each of their states and the D-Markov algorithm is computationally efficient [19]. The objective here is to project the original system model onto a D-Markov machine subspace for enhancement of computational efficiency without significantly compromising the modeling accuracy. In this formulation, D-Markov machines are referred to those with positive morph matrices. The underlying procedure is illustrated in the following two examples.

Example 7.1 (*Numerical computation*). Let the alphabet for the symbolic system be $\Sigma = \{a,b\}$. A set of orthonormal basis vectors needs to be specified first. Since all D-Markov machines with a specified depth d have a common structure [14], it is convenient to select those PFSA of the same structure and having different morph matrices as a basis. As an example, Fig. 5 shows two vectors in the $\mathcal{D}_1$ space, where e_1 and e_2 have the same Markov structure with $d=1$. They have two states $q_0 = \{\star a\}$ (i.e., all strings ending with the symbol a) and $q_1 = \{\star b\}$ (i.e., all strings ending with symbol b). The inner product of e_1 and e_2 is computed via Eq. (24) by using the base 2 logarithm as

$$\begin{aligned}\langle e_1, e_2 \rangle_A &= \log\frac{1/3}{2/3}\log\frac{2/3}{1/3}\mu(q_0) + \log\frac{1/3}{2/3}\log\frac{1/3}{2/3}\mu(q_1) \\ &= \mu(q_1) - \mu(q_0)\end{aligned} \tag{46}$$

Following Eq. (32), let the measure μ be chosen as μ_2 with the parameter $\theta = 1/2$. Then, Proposition 6.2 is applied to yield

$$\boldsymbol{\mu}_2 = \begin{bmatrix}\mu(q_0) \\ \mu(q_1)\end{bmatrix} = [1 \;\; 0](1-\theta)^2 (I - \theta \cdot \Pi^{\mathbb{U}(G)})^{-1} \cdot \Pi^{\mathbb{U}(G)} (I - \theta \cdot \Pi^{\mathbb{U}(G)})^{-1} \tag{47}$$

with $\Pi^{\mathbb{U}(G)} = \begin{bmatrix}1/2 & 1/2 \\ 1/2 & 1/2\end{bmatrix}$. Since $\boldsymbol{\mu}_2 = \begin{bmatrix}1/2 \\ 1/2\end{bmatrix}$ for $\theta = 1/2$, it follows that $\langle e_1, e_2 \rangle_A = 0$. With this choice of the measure μ, it is also verified that $\|e_1\|_A = \|e_2\|_A = 1$. This implies that e_1 and e_2 form an orthonormal basis for the space $\mathcal{D}_1$. In general, the dimension of the space $\mathcal{D}_d$ is $K \le |\Sigma|^d (|\Sigma| - 1)$. Therefore, an orthonormal basis can be obtained by applying the Gram–Schmidt procedure on a linearly independent set of K vectors.

Example 7.2 (*Orthogonal projection*). Let the alphabet for the symbolic system be $\Sigma = \{a,b\}$. Fig. 6 presents two PFSA, G_1 and G_2, which are not D-Markov machines, and their projections onto the D-Markov space $\mathcal{D}_1$ are $\mathbb{P}_{\mathcal{D}_1}(G_1)$ and $\mathbb{P}_{\mathcal{D}_1}(G_2)$, respectively. These projections are obtained from Eq. (45) in terms of the orthonormal bases e_1 and e_2 (see

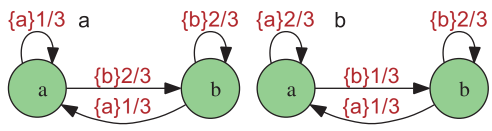


Fig. 5. An orthonormal basis for $\mathcal{D}_1$ ($q_0 = \{\star a\}$ and $q_1 = \{\star b\}$). (a) e_1, (b) e_2.

Fig. 5). Two plots in Fig. 7 display the projection errors of PFSA G_1 (solid line) and PFSA G_2 (dashed line) onto the projection chain $\mathcal{D}_1 \subset \mathcal{D}_2 \subset \cdots \subset \mathcal{D}_8$, respectively. It is observed that, for both G_1 and G_2, the projection errors decrease as the depth d of the D-Markov machine $\mathcal{D}_d$ becomes larger. Following Theorem 6.1, it is expected that as the depth d of $\mathcal{D}_d$ increases, the projection errors for both G_1 and G_2 would monotonically decrease.

To numerically interpret the meaning of the projection, a symbol sequence of length 10,000 is generated by simulating the PFSA G_1 and then the D-Markov algorithm is applied on the symbol sequence to obtain a D-Markov machine with depth $d=1$. The resulting output is shown in Fig. 8, which is very close to the analytically derived projection $\mathbb{P}_{\mathcal{D}_1}(G_1)$ in Fig. 6(c). That is, a low-order model captured by the D-Markov algorithm from the simulated symbolic sequence is very close to the optimal projection point in the proposed Hilbert space setting.

Example 7.3 (*Fatigue damage detection in polycrystalline alloys*). This example addresses the construction of a semantic model and the associated model order reduction based on time series of ultrasonic signals, collected from an experimental apparatus for fatigue damage detection in polycrystalline alloys [33].

Fig. 9(a) shows a picture of the experimental apparatus that is built upon a special-purpose uniaxial fatigue damage testing machine. The apparatus is instrumented with ultrasonic flaw detectors and an optical traveling microscope; the details of the operating procedure of the fatigue test apparatus and its instrumentation and control system are reported in [33]. Tests have been conducted using center-notched 7075-T6 aluminum specimens (see Fig. 9(b)) under a periodically varying load, where the maximum and minimum (tensile) loads were kept constant at 87 MPa and 4.85 MPa at 12.5 Hz frequency. Each specimen is 3 mm thick and 50 mm wide, and has a slot of

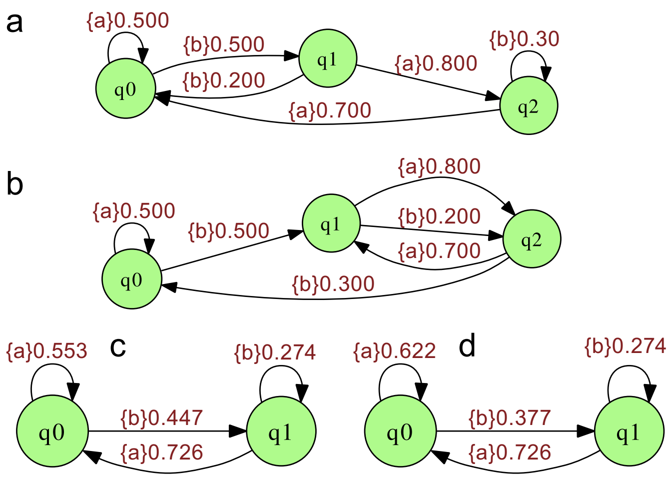


Fig. 6. Projection of PFSA G_1 and G_2 on $\mathcal{D}_1$. (a) G_1, (b) G_2, (c) $\mathbb{P}_{\mathcal{D}_1}(G_1)$, (d) $\mathbb{P}_{\mathcal{D}_1}(G_2)$.

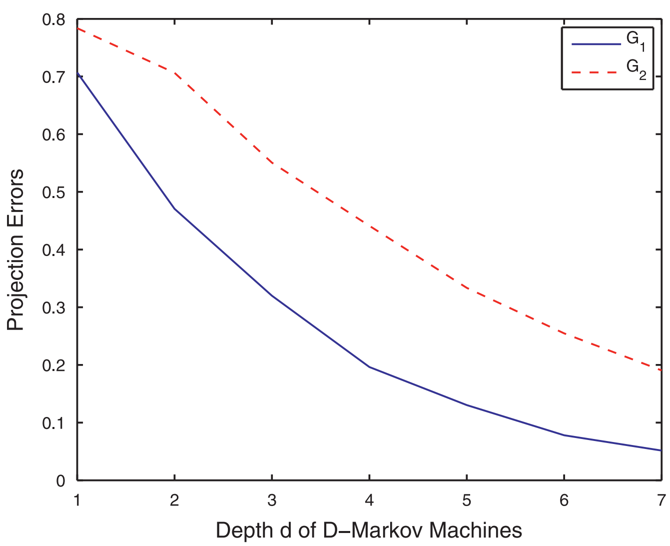


Fig. 7. Projection errors of G_1 and G_2 on D-Markov subspaces.

1.58 mm × 4.5 mm at the center. The central notch increases the stress concentration factor that ensures crack initiation and propagation at the notch ends [33]. The ultrasonic sensing device is triggered at a frequency of 5 MHz at each peak of the cyclic load. The time epochs, at which data are collected, are chosen to be 1000 load cycles (i.e., ~80 s) apart. At the beginning of each time epoch, the ultrasonic data points have been collected for 50 load cycles (i.e., ~4 s) which produce a time series of 15,000 data points, i.e., 300 data points around each load peak. It is assumed that no significant changes occur in the fatigue crack behavior of the test specimen during the tenure of data acquisition at a given time epoch. The nominal condition at the time epoch τ_0 is chosen to be 1.0

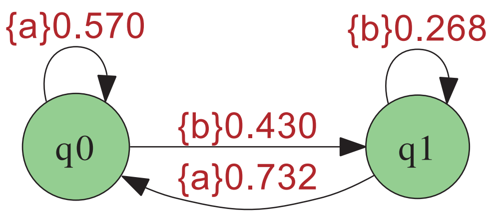


Fig. 8. Results of $\mathcal{D}_1$ algorithm based on simulated data from G_1.

kilocycles to ensure that the electro-hydraulic system of the test apparatus had come to a steady state and that no significant damage occurs till that point, i.e., zero damage at the time epoch τ_0. The fatigue damage at subsequent time epochs, $\tau_1, \tau_2, \ldots \tau_k \ldots$, are then calculated with respect to the nominal condition at τ_0. The set consists of data collected at 56 consecutive epochs.

The time-series data set at the time epoch τ_0 is first converted into a symbol sequence based on the maximum entropy partitioning (MEP) [34] for a given symbol alphabet size $|\Sigma| = 6$ and this partitioning is retained for all subsequent epochs, $\tau_1, \tau_2, \ldots$. Note that each partition segment is associated with a unique symbol in the alphabet and each symbol sequence characterizes the evolving fatigue damage and is modeled via a PFSA with Q states. Then, by the MEP property, the stationary state probability vector $\boldsymbol{p}_0$ of the resulting probabilistic finite state automaton (PFSA) model at the epoch τ_0 is uniformly distributed, i.e., $\boldsymbol{p}_0 = (1/|Q|)\boldsymbol{e}$, where $\boldsymbol{e}$ is the $|Q|$-dimensional row vector of all ones. Starting from the initial value of zero at the epoch τ_0, the fatigue damage at an epoch τ_k is expressed in terms of the respective (scalar) damage divergence defined as

$$m_k = d(\boldsymbol{p}_k, \boldsymbol{p}_0) \tag{48}$$

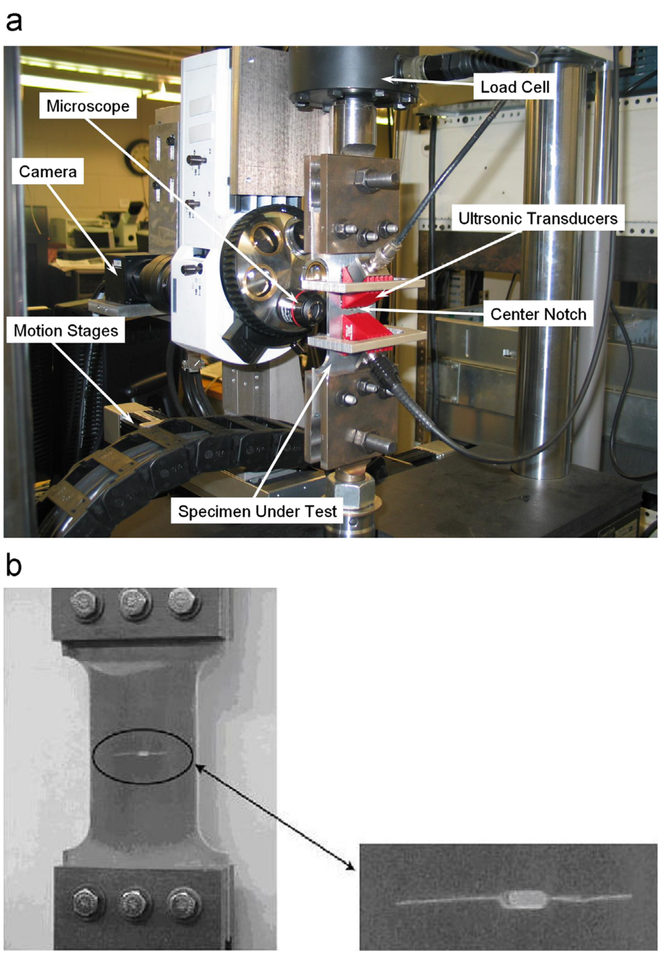


Fig. 9. Computer-instrumented apparatus and a 7075-T6 aluminum alloy specimen. (a) Experimental apparatus, (b) damaged specimen.

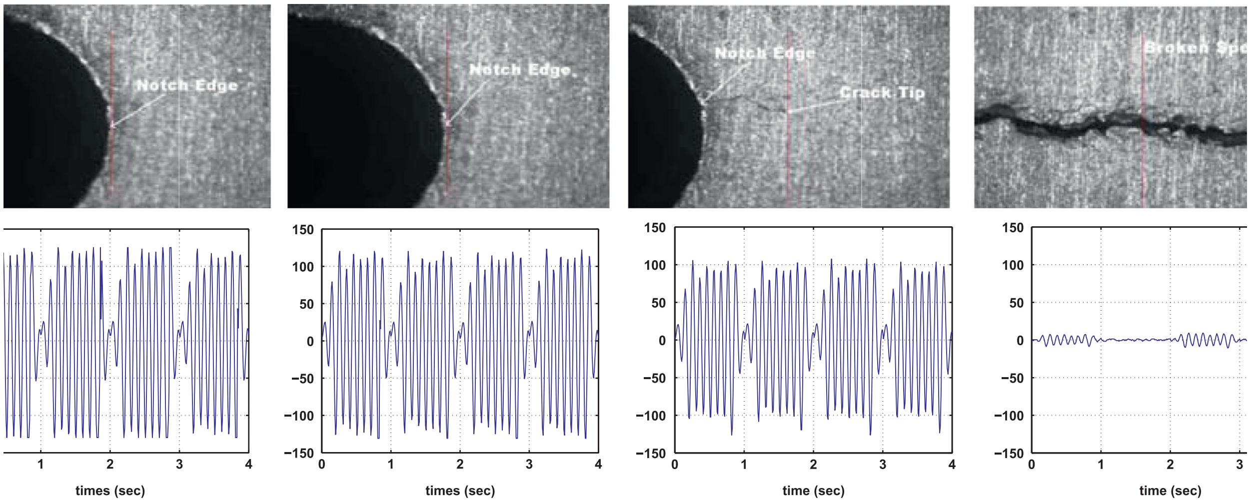


Fig. 10. Optical microscope images and ultrasonic time series of surface crack evolution in a 7075-T6 aluminum alloy test specimen. (a) Nominal: 1 kcycles, (b) internal damage: 23 kcycles, (c) surface crack: 34 kcycles, (d) full crack: 47 kcycles.

where $d(\bullet,\bullet)$ is an appropriate metric. In this study, $d(\bullet,\bullet)$ is chosen to be the standard Euclidean distance.

The two-dimensional (optical microscope) images of the specimen surface and the corresponding profiles of ultrasonic sensor outputs are, respectively, presented in the top row and the bottom row of Fig. 10 at different time epochs, approximately at 1, 23, 34 and 47 kilocycles, to visually display the gradual evolution of fatigue damage from the pre-selected reference condition. Fig. 10(a) and (b) shows that no surface crack becomes visible until ~ 23 kilocycles. Fig. 10(c) shows the appearance of a crack on the image of the specimen surface in the vicinity of 34 kilocycles while there is a small change in the profile of the ultrasonic signal. Eventually the amplitude of the ultrasonic signal dramatically decreases as seen in Fig. 10(d) when the surface crack is fully developed at 47 kilocycles.

At the nominal condition of the time epoch τ_0 and subsequent epochs τ_k with $k=1,2,\ldots$, respective PFSA models are constructed from the collected data sets. Each of these PFSA models has 20 states and is not restricted to be a D-Markov machine [14]. The 20-state (non-D-Markov) PFSA models are projected onto the subspaces of (lower dimensional) D-Markov machines with different values of $|Q|$ for model order reduction.

The D-Markov machine with $D=0$ generates the space of the single-state PFSA, because the number of PFSA states $|Q|=|\Sigma|^D=1$. Projection onto this space is not of interest as it yields a grossly oversimplified model. However, the D-Markov machine with $D=1$ implies that the space consists of PFSA with $|Q|=|\Sigma|$, i.e., $|Q|=6$. Furthermore, the measure in the construction of the inner product is chosen to be μ_2 (see Eq. (32)) with $\theta=\frac{1}{2}$. The six-state PFSA is found to be adequate for representation of damage divergence (see Eq. (48)) in terms of its state probability distribution, as seen in Fig. 11 that shows a comparison of a pair of damage evolution profiles: the plot in solid line corresponds to the original model of 20-state PFSA and the plot in dashed line corresponds to the projected model of six-state PFSA. These two profiles of damage divergence are observed to be very close to each other. It is apparent from Fig. 11 that not only the orthogonal projection reduces the complexity of the semantic model of the dynamical system (i.e., the process of fatigue damage evolution) practically without compromising the performance of damage detection, but also the reduced order model removes the small ripples in the damage divergence profile produced by the 20-state model.

8. Conclusions and future work

With the objective of signal representation and modeling of interacting dynamical systems, this paper develops a vector space model for a class of probabilistic finite state automata (PFSA) that are constructed based on finite-length symbol sequences derived from time series of physical signals. The construction procedure is formulated in a measure-theoretic setting, where 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 user-configurable inner products on the respective vector spaces. Numerical examples are presented to illustrate the computational steps of the proposed method and to explain the operation of orthogonal projection from Hilbert spaces of general PFSA onto closed Markov subspaces that belong to a class of shifts of finite type [25]. The concepts of vector space model and model order reduction by orthogonal projection are validated on the experimental data. The orthogonal projection technique in the Hilbert space of PFSA is potentially useful for signal representation, modeling and analysis (e.g., model identification, model order reduction, and system performance analysis) of physical systems in a computationally efficient manner. In this context, some of the research topics that are envisioned to enhance the theory and applications of symbolic system modeling, presented in this paper, are delineated below.

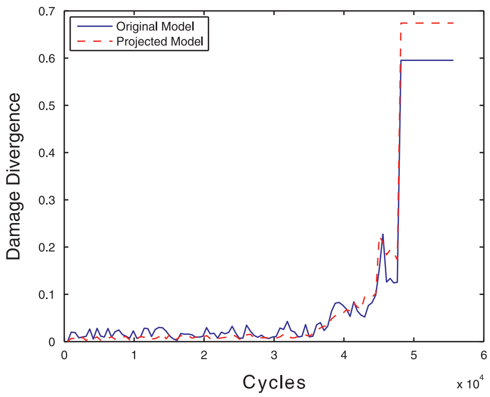


Fig. 11. Evolution of damage divergence.

(i) *Signal representation*: Systematic procedures for construction of symbolic systems need to be integrated with those for construction of Hilbert spaces of PFSA for representation of physical signals in different contexts.
(ii) *Pattern classification*: Systematic procedures for construction of PFSA models need to be developed to generate feature vectors for pattern classification in physical systems.
(iii) *Choice of the measure μ for construction of the inner product*: A systematic procedure for selection of an application-dependent μ and its effects on the inner product needs further investigation.
(iv) *Performance and computational complexity of the symbolic model*: State merging and state splitting algorithms need to be investigated toward this end.

Acknowledgments

The authors thankfully acknowledge the benefits of discussion with Dr. Ishanu Chattopadhyay.

References

[1] R. Badii, A. Politi, Complexity: Hierarchical Structures and Scaling in Physics, Cambridge University Press, Cambridge, UK, 1997.
[2] G. Pola, P. Tabuada, Symbolic models for nonlinear control systems: alternating approximate bisimulations, SIAM Journal of Control and Optimization 48 (2) (2009) 719–733.
[3] X. Jin, S. Gupta, K. Mukherjee, A. Ray, Wavelet-based feature extraction using probabilistic finite state automata for pattern classification, Pattern Recognition 44 (7) (2011) 1343–1356.
[4] S. Gupta, A. Ray, Statistical mechanics of complex systems for pattern identification, Journal of Statistical Physics 134 (2) (2009) 337–364.
[5] E. Vidal, F. Thollard, C. de la Higuera, F. Casacuberta, R. Carrasco, Probabilistic finite-state machines—part I and part II, IEEE Transactions on Pattern Analysis and Machine Intelligence 27 (7) (2005) 1013–1039.
[6] K. Deng, P. Mehta, S. Meyn, Optimal Kullback–Leibler aggregation via spectral theory of Markov chains, IEEE Transactions on Automatic Control 56 (12) (2011) 2793–2808.
[7] L. Rabiner, A tutorial on hidden Markov models and selected applications in speech processing, Proceedings of the IEEE 77 (2) (1989) 257–286.
[8] M. Vidyasagar, The complete realization problem for hidden Markov models: a survey and some new results, Mathematics of Control, Signals, and Systems 23 (December) (2011) 1–65.
[9] P. Lanchantin, W. Pieczynski, Unsupervised restoration of hidden nonstationary Markov chains using evidential priors, IEEE Transactions on Signal Processing 53 (8) (2005) 3091–3098.
[10] R. Carrasco, J. Oncina, Learning stochastic regular grammars by means of a state merging method, in: Proceedings of the Second International Colloquium Grammatical Inference and Applications (ICGI94), 1994, pp. 139–152.
[11] L. Saul, F. Pereira, Aggregate and mixed-order Markov models for statistical language processing, Empirical Methods in Natural Language Processing (1997) 81–89.
[12] D. Ron, Y. Singer, N. Tishby, On the learnability and usage of acyclic probabilistic finite automata, Journal of Computer and System Science 56 (2) (1998) 133–152.
[13] 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, Arlington, VA, USA, AUAI Press, 2004, pp. 504–511.
[14] A. Ray, Symbolic dynamic analysis of complex systems for anomaly detection, Signal Processing 84 (7) (2004) 1115–1130.
[15] G. Mallapragada, A. Ray, X. Jin, Symbolic dynamic filtering and language measure for behavior identification of mobile robots, IEEE Transactions on System, Man, and Cybernetics—Part B: Cybernetics 42 (June (3)) (2012) 647–659.
[16] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, NY, USA, 2006.
[17] S. Iyenger, P. Varshney, T. Damarla, A parametric copula-based framework for hypothesis testing using heterogeneous data, IEEE Transactions on Signal Processing 59 (May (5)) (2011) 2308–2319.
[18] A. Sunderesan, P. Varshney, Location estimation of a random signal source based on correlated sensor observations, IEEE Transactions on Signal Processing 59 (February (2)) (2011) 787–799.
[19] C. Rao, A. Ray, S. Sarkar, M. Yasar, Review and comparative evaluation of symbolic dynamic filtering for detection of anomaly patterns, Signal, Image, and Video Processing 3 (2009) 101–114.
[20] A. Ray, Signed real measure of regular languages for discrete event supervisory control, International Journal of Control 78 (12) (2005) 949–967.
[21] 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.
[22] Y. Wen, A. Ray, Vector space formulation of probabilistic finite state automata, Journal of Computer and System Sciences 78 (July) (2012) 1127–1141.
[23] P. Adenis, Y. Wen, A. Ray, An inner product space on irreducible and synchronizable probabilistic finite state automata, Mathematics of Control, Signals, and Systems 23 (January (1)) (2012) 281–310.
[24] T. Cover, J. Thomas, Elements of Information Theory, 2nd ed. Wiley-Interscience, New York, NY, USA, 2006.
[25] D. Lind, B. Marcus, An Introduction to Symbolic Dynamics and Coding, Cambridge University Press, Cambridge, UK, 1995.
[26] M. Sipser, Introduction to the Theory of Computation, PWS Publishing, Boston, MA, USA, 1997.
[27] W. Rudin, Real and Complex Analysis, 3rd ed. McGraw-Hill book Co., New York, NY, USA, 1987.
[28] T. Ferguson, A Bayesian analysis of some nonparametric problems, The Annals of Statistics 1 (2) (1973) 209–230.
[29] Y. Wen, K. Mukherjee, A. Ray, Adaptive pattern classification for symbolic dynamic systems, Signal Processing 93 (January) (2013) 252–260.
[30] I. Chattopadhyay, A. Ray, Structural transformations of probabilistic finite state machines, International Journal of Control 81 (5) (2008) 820–835.
[31] L. Ljung, System Identification: Theory for the User, 2nd ed. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1999.
[32] R. Pathria, Statistical Mechanics, 2nd ed. Butterworth-Heinemann, Oxford, UK, 1996.
[33] S. Gupta, A. Ray, E. Keller, Symbolic time series analysis of ultrasonic data for early detection of fatigue damage, Mechanical Systems and Signal Processing 21 (2) (2007) 866–884.
[34] V. Rajagopalan, A. Ray, Symbolic time series analysis via wavelet-based partitioning, Signal Processing 11 (November) (2006) 3309–3320.