International Journal of Control
Vol. 81, No. 5, May 2008, 820–835

Taylor & Francis
Taylor & Francis Group

Structural transformations of probabilistic finite state machines

I. CHATTOPADHYAY and A. RAY*

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

(*Received 12 June 2007; in final form 25 September 2007*)

Probabilistic finite state machines have recently emerged as a viable tool for modelling and analysis of complex non-linear dynamical systems. This paper rigorously establishes such models as finite encodings of probability measure spaces defined over symbol strings. The well known Nerode equivalence relation is generalized in the probabilistic setting and pertinent results on existence and uniqueness of minimal representations of probabilistic finite state machines are presented. The binary operations of probabilistic synchronous composition and projective composition, which have applications in symbolic model-based supervisory control and in symbolic pattern recognition problems, are introduced. The results are elucidated with numerical examples and are validated on experimental data for statistical pattern classification in a laboratory environment.

1. Introduction and motivation

Probabilistic finite state machines have recently emerged as a modelling paradigm for constructing causal models of complex dynamics. The general inapplicability of classical identification algorithms in complex non-linear systems has led to development of several techniques for construction of probabilistic representations of dynamical evolution from observed system behaviour. The essential feature of a majority of such reported approaches is partial or complete departure from the classical continuous-domain modelling towards a formal language theoretic and hence symbolic paradigm (Ray 2004, Shalizi and Shalizi 2004). The continuous range of a sufficiently long observed data set is discretized and tagged with labels to obtain a symbolic sequence (Ray 2004), which is subsequently used to compute a language-theoretic finite state probabilistic predictor via recursive model update algorithms. Symbolization essentially discretizes the continuous state space and gives rise to probabilistic dynamics from the underlying deterministic process, as illustrated in figure 1.

Among various reported symbolic reconstruction algorithms, causal-state splitting reconstruction (CSSR) (Shalizi and Shalizi 2004) computes optimal representations (e.g., ε-machines) and is reported to yield the minimal representation consistent with accurate prediction. In contrast, the D-Markov construction (Ray 2004) produces a sub-optimal model, but it has a significant computational advantage and has been shown to be better suited for online detection of small parametric anomalies in dynamic behaviour of physical processes (Rajagopalan and Ray 2006).

This paper addresses the issue of structural manipulation of such inferred probabilistic models of system dynamics. The ability to transform and manipulate the automaton structure is critical for design of supervisory control algorithms for symbolic models and real-time pattern recognition from symbol sequences. Specific issues are delineated in the sequel.

1.1 *Applications of symbolic model-based control*

The natural setting for developing control algorithms for symbolic models is that of probabilistic languages. The notion of probabilistic languages in the context of studying qualitative stochastic behaviour of discrete-event systems first appeared in Garg (1992a, b), where the concept of p-languages ('p' implying probabilistic) is introduced and an algebra is developed to model probabilistic languages based on concurrency.

*Corresponding author. Email: axr2@psu.edu

International Journal of Control
ISSN 0020–7179 print/ISSN 1366–5820 online © 2008 Taylor & Francis
http://www.tandf.co.uk/journals
DOI: 10.1080/00207170701704746

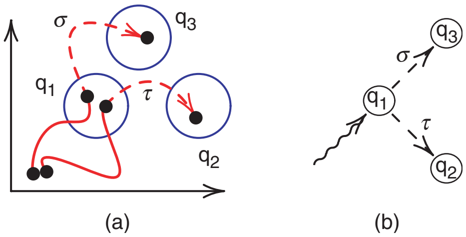

Figure 1. Emergence of probabilistic dynamics from the underlying deterministic system due to discretization as symbolic states q_1, q_2, q_3; and σ, τ are symbolic events.

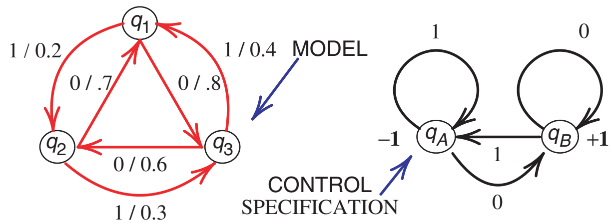

Figure 2. Symbolic model and control specification.

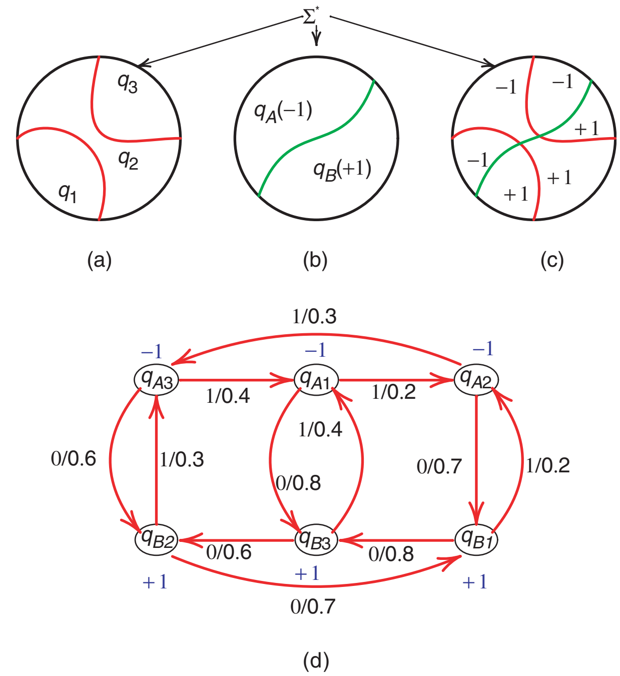

Figure 3. Imposing control specification by probabilistic synchronous composition of automata.

A multitude of control algorithms for *p*-language-theoretic models have been reported. Earlier approaches Lawford and Wonham (1993) and Kumar and Garg (2001) attempt a direct generalization of Ramadge and Wonham's supervisory control theory (Ramadge and Wonham 1987) for deterministic languages and proves to be somewhat cumbersome in practice. A significantly simpler approach is suggested in Ray (2005), Chattopadhyay (2006) and Chattopadhyay and Ray (2007), where supervisory control laws are synthesized by elementwise maximization of a language measure vector (Ray 2005, Chattopadhyay and Ray 2006) to ensure that the generated event strings cause the supervised plant to visit the "good" states while attempting to avoid the "bad" states optimally in a probabilistic sense. The notion of "good" and "bad" is induced by specifying scalar weights on the model states, with relatively more negative weights indicating less desirable states. Unlike the previous approaches, the measure-theoretic approach does not require a "specification automaton"; however, a specification weight is assigned to each state of the finite state machine. (Note: These states are different from the states obtained via symbolic reconstruction of observed physical data.)

Figure 2 illustrates the underlying concept. The symbolic model shown on the left which has three states q_1, q_2, q_3, while the control objective is specified by weights $+1$ and -1 on states q_A, q_B of the two-state automaton on the right.

Recalling that every finite state automaton induces a right invariant partition on the set of all possible finite length strings, the above situation is illustrated by figures 3(a) and (b). The operation of probabilistic synchronous composition, defined later in this paper, resolves the problem by considering the product partition in figure 3(c). Then, the given model is ztransformed into the one shown in figure 3(d), on which the optimization algorithm reported in Chattopadhyay and Ray (2007) can be directly applied to yield the optimal supervision policy.

1.2 *Applications of symbolic pattern recognition*

As mentioned earlier, the symbolic reconstruction algorithms (Ray 2004, Shalizi and Shalizi 2004) generate probabilistic finite state models from observed time series. However, in a pattern classification problem, one may be only interested in a given class of possible future evolutions. For example, as illustrated in figure 4, while the systems $G_1, G_2, \ldots, G_k$ yield different symbolic models $A_1, A_2, \ldots, A_k$, we may be only interested in matching a given template, i.e., knowing how similar the systems are as far as strings with even number of 0s is concerned (note: $q_A = \{\text{strings with even number of 0s}\}$).

The operation of projective composition, defined in this paper, allows transformation of each model A_i to the structure of the template while preserving the distribution over the strings of interest, and is of critical importance in symbolic pattern classification problems. As shown in the sequel, the model order of the machines

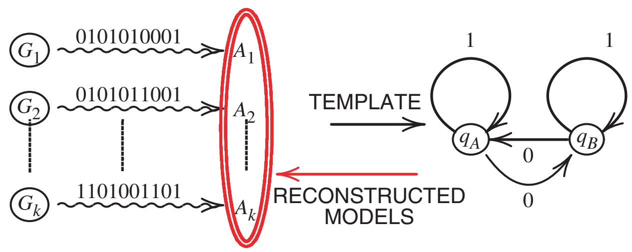

Figure 4. Symbolic template matching problem.

A_i is not particularly important; hence projective composition accomplishes model order reduction within a quantifiable error.

1.3 *Organization of the paper*

The paper is organized in seven sections including the present one. Section 2 presents preliminary concepts and pertinent results that are necessary for subsequent development. Section 3 introduces the concept of probabilistic finite state automata as finite encodings of probability measure spaces. The concept of Nerode equivalence is generalized to probabilistic automata and the key results on existence and uniqueness of minimal representations are established. Section 4 presents metrics on the space of probability measures on symbolic strings which is shown to induce pseudometrics on the space of probabilistic finite state automata. Along this line, the concept of probabilistic synchronous composition is introduced and the results are elucidated with a simple example. Section 5 defines projective composition and invariance of projected distributions is established. A numerical example is provided for clarity of exposition. Section 6 demonstrates applicability of the developed method to a pattern classification problem on experimental data. The paper is summarized and concluded in §7 with recommendations for future research.

2. Preliminary notions

A deterministic finite state automaton (*DFSA*) is defined (Hopcroft *et al*. 2001) as a quintuple $G_i = (Q, \Sigma, \delta, q_i, Q_m)$, where Q is the finite set of states, and $q_i \in Q$ is the initial state; Σ is the (finite) alphabet of events. The Kleene closure of Σ, denoted as Σ^*, is the set of all finite-length strings of events including the empty string ε; the set of all finite-length strings of events excluding the empty string ε is denoted as Σ^+ and the set of all *strictly* infinite-length strings of events is denoted as Σ^ω. A subset of Σ^ω is called an ω-language on the alphabet Σ and a subset of Σ^* is called a *-language. If the meaning is clear from context, we refer to a set of strings simply as a language. The function $\delta: Q \times \Sigma \to Q$ represents the state transition map and $\delta^*: Q \times \Sigma^* \to Q$ is the reflexive and transitive closure (Hopcroft *et al.* 2001) of δ and $Q_m \subseteq Q$ is the set of marked (i.e. accepting) states. For given functions f and g, we denote the composition as $f \circ g$.

Definition 1: The classical Nerode equivalence N (Hopcroft *et al*. 2001) on Σ^* with respect to a given language L is defined as:

$$\forall x, y \in \Sigma^*, \quad (x \mathcal{N} y \Leftrightarrow (\forall u \in \Sigma^* (xu \in L) \Longleftrightarrow (yu \in L))). \tag{1}$$

A language $L \subseteq \Sigma^*$ is regular if and only if the corresponding Nerode equivalence is of finite index (Hopcroft *et al*. 2001).

Probabilistic finite state automata (PFSA) considered in this paper are built upon deterministic finite state automata (DFSA) with a specified event generating function. The formal definition is stated next.

Definition 2 (PFSA)**:** A probabilistic finite state automata (PFSA) is a quintuple $P_i \triangleq (Q, \Sigma, \delta, q_i, \tilde{\pi})$ where the quadruple (Q, Σ, δ, q_i) is a DFSA with unspecified marked states and the mapping $\tilde{\pi}: Q \times \Sigma \to [0, 1]$ satisfies the following condition:

$$\forall q_j \in Q, \quad \sum_{\sigma \in \Sigma} \tilde{\pi}(q_j, \sigma) = 1. \tag{2}$$

In the sequel, $\tilde{\pi}$ is denoted as the event generating function. For a PFSA P_l, cardinality of the set of states is denoted as $\text{NUMSTATES}(P_l)$.

Definition 3: For every PFSA $P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi})$, there is an associated stochastic matrix $\prod \in \mathbb{R}^{\text{NUMSTATES}(P_i) \times \text{NUMSTATES}(P_i)}$ called the state transition probability matrix, which is defined as follows:

$$\prod_{jk} = \sum_{\sigma: \delta(q_j, \sigma) = q_k} \tilde{\pi}(q_j, \sigma). \tag{3}$$

We further note that for every stochastic matrix Π, there exists at least one row-vector $\wp$ such that

$$\wp \Pi = \wp, \quad \text{where} \quad \forall j \quad \wp_j \geqq 0 \quad \text{and} \quad \sum_{j=1}^{\text{NUMSTATES}(P_i)} \wp_j = 1, \tag{4}$$

where $\wp$ is a stable long term distribution over the PFSA states. If Π is irreducible, then $\wp$ is unique. Otherwise, there may exist more than one possible solution to equation (4), one for each eigenvector corresponding to unity eigenvalue. However, if the initial state is specified (as it is in this paper), then $\wp$ is always unique. Several efficient algorithms have been reported in Kemeny and Snell (1960), Harrod and

Plemmons (1984) and Stewart (1999) for computation of $\wp$.

Key definitions and results from measure theory that are used here are recalled.

Definition 4 (σ-Algebra)**:** A collection $\mathfrak{M}$ of subsets of a non-empty set X is said to be a σ-algebra (Rudin 1988) in X if $2M$ has the following properties:

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

Theorem 1: *If $\mathscr{F}$ is any collection of subsets of X, there exists a smallest σ-algebra $\mathfrak{M}$ in X such that $\mathscr{F} \subseteq \mathfrak{M}^*$.*

Proof: See (Rudin 1988, Theorem 1.10) □

Definition 5 (Measure)**:** A finite (non-negative) measure is a countably additive function μ, defined on a σ-algebra $\mathfrak{M}$, whose range is $[0, K]$ for some $K \in \mathbb{R}$. Countable additivity means that if $\{A_i\}$ is a disjoint countable collection of members of $\mathfrak{M}$, then

$$\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \tag{5}$$

Theorem 2: *If μ is a (non-negative) measure on a σ-algebra $\mathfrak{M}$, then*

(1) $\mu(\emptyset) = 0$
(2) (*Monotonicity*) $A \subseteq B \Rightarrow \mu(A) \leq \mu(B)$ *if* $A, B \in \mathfrak{M}$.

Proof: See (Rudin 1988, Theorem 1.19). □

Definition 6: A probability measure on a non-empty set with a specified σ-algebra $\mathfrak{M}$ is a finite non-negative measure on $\mathfrak{M}$. Although not required by the theory, a probability measure is defined to have the unit interval [0, 1] as its range.

Definition 7: A probability measure space is a triple $(X, \mathfrak{M}, \mathfrak{p})$ where X is the underlying set, $\mathfrak{M}$ is the σ-algebra in X and $\mathfrak{p}$ is a finite non-negative measure on $\mathfrak{M}$.

3. Properties of probabilistic finite state automata

For any $\tau \in \Sigma^*$, the language $\tau\Sigma^\omega$ has an important physical interpretation pertaining to systems modeled as probabilistic language generators (figure 5). A string $\tau \in \Sigma^*$ can be interpreted as a symbol sequence that has been already generated, and any string in Σ^ω qualifies as a possible future evolution. Thus, the language $\tau\Sigma^\omega$ is conceptually associated with the current dynamical state of the modelled system.

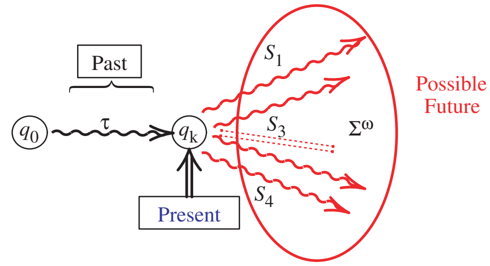

Figure 5. Interpretation of the language $\tau\Sigma^\omega$ pertaining to dynamical evolution of a language generator.

Definition 8: Given an alphabet Σ, the set $\mathfrak{B}_\Sigma \triangleq 2^{\Sigma^*}\Sigma^\omega$ is defined to be the σ-algebra generated by the set $\{L : L = \tau\Sigma^\omega \text{ where } \tau \in \Sigma^*\}$, i.e., the smallest σ-algebra on the set Σ^ω which contains the set $\{L : L = \tau\Sigma^\omega \text{ where } \tau \in \Sigma^*\}$.

Remark 1: Cardinality of $\mathfrak{B}_\Sigma$ is $\aleph_1$ because both $2^{\Sigma*}$ and Σ^ω have cardinality $\aleph_1$.

The following relations in the probability measure space $(\Sigma^\omega, \mathfrak{B}_\Sigma, \mathfrak{p})$ are consequences of Definition 8.

- $\mathfrak{p}(\Sigma^\omega) = \mathfrak{p}(\varepsilon\Sigma^\omega) = 1$
- $\forall x, u \in \Sigma^*$, $xu\Sigma^\omega \subseteqq x\Sigma^\omega$ and hence $\mathfrak{p}(xu\Sigma^\omega) \leqq \mathfrak{p}(\Sigma^\omega)$

Notation 1: For brevity, the probability $\mathfrak{p}(\tau\Sigma^\omega)$ is denoted as $\mathfrak{p}(\tau)\forall\tau \in \Sigma^*$ in the sequel.

Next the notion of probabilistic Nerode equivalence $\mathcal{N}_\mathfrak{p}$ is introduced on Σ^* for representing the measure space $(\Sigma^\omega, \mathfrak{B}_\Sigma, \mathfrak{p})$ in the form of a PFSA. In this context, the following logical formulae are introduced.

Definition 9: For $x, y \in \Sigma^*$,

$$\mathbb{U}_1(x, y) \triangleq (\mathfrak{p}(x) = 0 \wedge \mathfrak{p}(y) = 0) \tag{6a}$$

$$\mathbb{U}_2(x, y) \triangleq (\mathfrak{p}(x) \neq 0 \wedge \mathfrak{p}(y) \neq 0) \wedge \left(\forall u \in \Sigma^*\left(\frac{\mathfrak{p}(xu)}{\mathfrak{p}(x)} = \frac{\mathfrak{p}(yu)}{\mathfrak{p}(y)}\right)\right). \tag{6b}$$

Theorem 3 (probabilistic nerode equivalence)**:** *Given an alphabet Σ, every measure space $(\Sigma^\omega, \mathfrak{B}_\Sigma, \mathfrak{p})$ induces a right-invariant equivalence relation $\mathcal{N}_\mathfrak{p}$ on Σ^* defined as*

$$\forall x, y \in \Sigma^*, \left(x\mathcal{N}_\mathfrak{p}y \Leftrightarrow \mathbb{U}_1(x, y) \vee \mathbb{U}_2(x, y)\right). \tag{7}$$

Proof: Reflexivity and symmetry properties of the relation $\mathcal{N}_\mathfrak{p}$ follow from Definition 9. Let $x, y, z \in \Sigma^*$ be distinct and arbitrary strings such that $x\mathcal{N}_\mathfrak{p}y$ and $y\mathcal{N}_\mathfrak{p}z$. Then, transitivity property of $\mathcal{N}_\mathfrak{p}$ follows from equation (7) and Definition 9. Hence, $\mathcal{N}_\mathfrak{p}$ is an equivalence relation.

To establish right-invariance (Hopcroft *et al.* 2001) of $\mathcal{N}_\mathfrak{p}$, it suffices to show that

$$\forall x, y \in \Sigma^*, \left(x\mathcal{N}_\mathfrak{p}y \Leftrightarrow \forall u \in \Sigma^*, \left(xu\mathcal{N}_\mathfrak{p}yu\right)\right). \tag{8}$$

Let x, y, u be arbitrary strings in Σ^* such that $x\mathcal{N}_\mathfrak{p}y$. If $\mathfrak{p}(x)=0$, $\mathfrak{p}(y)=0$ from equation (7). Then, it follows from the monotonicity property of the measure (see Theorem 2) that $\mathfrak{p}(xu)=0$, which implies the truth of $\mathbb{U}_1(xu,yu)$ and hence the truth of $xu\mathcal{N}_\mathfrak{p}yu$. If $\mathfrak{p}(x)\neq 0$, then $(x\mathcal{N}_\mathfrak{p}y)\wedge(\mathfrak{p}(x)\neq 0)$ implies $\mathfrak{p}(y)\neq 0$. Hence,

$$\frac{\mathfrak{p}(xu\tau)}{\mathfrak{p}(xu)}=\frac{\mathfrak{p}(xu\tau)}{\mathfrak{p}(x)}=\frac{\mathfrak{p}(x)}{\mathfrak{p}(xu)}. \tag{9}$$

If $\mathfrak{p}(x)=\mathfrak{p}(y)$, then $x\mathcal{N}_\mathfrak{p}y$ implies $\mathfrak{p}(xu)=\mathfrak{p}(yu)$ and also $\forall\tau\in\Sigma^*(\mathfrak{p}(xu\tau)=\mathfrak{p}(yu\tau))$. Similarly, if $\mathfrak{p}(x)\neq\mathfrak{p}(y)$, then $x\mathcal{N}_\mathfrak{p}y$ implies $\mathfrak{p}(xu)\neq\mathfrak{p}(yu)$ and also $\forall\tau\in\Sigma^*\times(\mathfrak{p}(xu\tau)\neq\mathfrak{p}(yu\tau))$. Hence, $\forall\tau\in\Sigma^*((\mathfrak{p}(xu)=\mathfrak{p}(yu))\Leftrightarrow(\mathfrak{p}(xu\tau)=\mathfrak{p}(yu\tau)))$. □

Definition 10 (perfect encoding)**:** Given an alphabet Σ, PFSA $P_i=(Q,\Sigma,\delta,q_i,\tilde{\pi})$ is defined to be a perfect encoding of the measure space $(\Sigma^\omega,\mathfrak{B}_\Sigma,\mathfrak{p})$ if $\forall\tau\in\Sigma^+$ and $\tau=\sigma_1\sigma_2\ldots\sigma_r$,

$$\mathfrak{p}(\tau)=\tilde{\pi}(q_i,\sigma_1)\prod_{k=1}^{r-1}\tilde{\pi}(\delta^*(q_i,\sigma_1\ldots\sigma_k),\sigma_{k+1}). \tag{10}$$

Remark 2: The implications of Definition 10 are as follows: The encoding introduced is perfect in the sense that the measure $\mathfrak{p}$ can be reconstructed without error from the specification of P_i.

Theorem 4: *A PFSA is a perfect encoding if and only if the corresponding probabilistic Nerode equivalence $\mathcal{N}_\mathfrak{p}$ is of finite index.*

Proof (Left to Right)**:** Let Q be the finite set of equivalence classes of the relation $\mathcal{N}_\mathfrak{p}$ of the PFSA $P_i=(Q,\Sigma,\delta,q_i,\tilde{\pi})$ that is constructed as follows:

(1) Since $\mathcal{N}_\mathfrak{p}$ is an equivalence relation on Σ^*, there exists a unique $q_i\in Q$ such that $\varepsilon\in q_i$. The initial state of P_i is set to q_i.
(2) If $x\in q_j$ and $x\sigma\in q_k$, then $\delta(q_j,\sigma)=q_k$
(3) $\tilde{\pi}(q_j,\sigma)=(\mathfrak{p}(x\sigma)/\mathfrak{p}(x))$ where $x\in q_j$.

First we verify that the steps 2 and 3 are consistent in the sense that δ and $\tilde{\pi}$ are well-defined.

Probabilistic nerode equivalence (see Theorem 3) implies that if $x,y\in\Sigma^*$, then $((x\in q_j)\wedge(x\sigma\in q_k)\wedge(y\in q_j))\Rightarrow(y\sigma\in q_k)$. Therefore, the constructed δ is well-defined. Similarly, since $(x,y\in q_j)\Rightarrow(\mathfrak{p}(x)=\mathfrak{p}(y))\wedge(\mathfrak{p}(x\sigma)=\mathfrak{p}(y\sigma))$, the constructed $\tilde{\pi}$ is also well-defined. Therefore, the steps 2 and 3 are consistent. For $\tau=\sigma_1\sigma_2\ldots\sigma_r\in\Sigma^+$, it follows that

$$\mathfrak{p}(\tau)=p(\sigma_1)\prod_{r=2}^{R}\frac{\mathfrak{p}(\sigma_1\ldots\sigma_r)}{\mathfrak{p}(\sigma_1\ldots\sigma_{r-1})}$$
$$=\tilde{\pi}(q_i,\sigma_1)\prod_{r=1}^{R-1}\tilde{\pi}(\delta^*(q_i,\sigma_1\ldots\sigma_r),\sigma_{r+1}).$$

Hence, the criterion for perfect encoding (see Definition 10) is satisfied.

Right to Left: Let the PFSA $P_i=(Q,\Sigma,\delta,q_i,\tilde{\pi})$ be a perfect encoding; and let the probabilistic Nerode equivalence $\mathcal{N}_\mathfrak{p}$ be of infinite index. Then, there exists a set of strings $\mathscr{H}\subseteq\Sigma^*$, having the same cardinality as Σ^*, such that each element of $\mathscr{H}$ belongs to a distinct $\mathcal{N}_\mathfrak{p}$-equivalence class. That is, $\forall h_j,h_k\in\mathscr{H}$ such that $j\neq k$, we have $h_j\mathcal{N}_\mathfrak{p}h_k$. Since $\mathfrak{p}(h_j)=\mathfrak{p}(h_k)=0$ implies $h_j\mathcal{N}_\mathfrak{p}h_k$, there can exist at most one element $h_0\in\mathscr{H}$ such that $\mathfrak{p}(h_0)=0$. That is, $\mathfrak{p}(h_j)\neq 0\ \forall h_j\in\mathscr{H}-\{h_0\}$.

For the PFSA $P_i=(Q,\Sigma,\delta,q_i,\tilde{\pi})$, where Q is the finite set of states, there exists $q_\ell\in Q$ and $h_j,h_k\in\mathscr{H}$ such that $\delta^*(q_i,h_j)=\delta^*(q_i,h_k)=q_\ell$. Let $\tau\in\Sigma^+$ and $\tau=\sigma_1\sigma_2\cdots\sigma_r$. Since P_i ia a perfect encoding, it follows from Definition 10 that

$$\mathfrak{p}(h_j\tau)=\mathfrak{p}(h_j)\tilde{\pi}(q_\ell,\sigma_1)\prod_{m=1}^{r-1}\tilde{\pi}(\delta^*(q_\ell,\sigma_1\cdots\sigma_m),\sigma_{m+1})$$
$$\mathfrak{p}(h_k\tau)=\mathfrak{p}(h_k)\tilde{\pi}(q_\ell,\sigma_1)\prod_{m=1}^{r-1}\tilde{\pi}(\delta^*(q_\ell,\sigma_1\cdots\sigma_m),\sigma_{m+1}).$$

Now, it follows that

$$\big(\mathfrak{p}(h_j)\neq 0\wedge\mathfrak{p}(h_k)\neq 0\big)\bigwedge\left(\frac{\mathfrak{p}(h_j\tau)}{\mathfrak{p}(h_j)}=\frac{\mathfrak{p}(h_k\tau))}{\mathfrak{p}(h_k)}\right)$$
$$\Rightarrow\mathbb{U}_2(h_j,h_k)\Rightarrow h_j\mathcal{N}_\mathfrak{p}h_k$$

which contradicts the initial assertion that $h_j\mathcal{N}_\mathfrak{p}h_k\ \forall h_j,h_k\in\mathscr{H}$. This completes the proof. □

The construction in the first part of Theorem 4 is stated in the form of Algorithm 1.

Algorithm 1: Construction of PFSA from the probability measure space $(\Sigma^\omega,\mathfrak{B}_\Sigma,\ \mathfrak{p})$

input: $(\Sigma^\omega,\mathfrak{B}_\Sigma,\ \mathfrak{p})$ such that $\mathcal{N}_\mathfrak{p}$ is of finite index
output: P_i
1 **begin**
2 Let $Q=\{q_j: j\in\mathcal{J}\subsetneqq\mathbb{N}\}$ be the set of equivalence classes of the relation $\mathcal{N}_\mathfrak{p}$;
3 Set the initial state of P_i as q_i such that ε belongs to the equivalence class q_i; If $x\in q_j$ and $x\sigma\in q_k$, then set $\delta(q_j,\sigma)=q_k$;
4 $\tilde{\pi}(q_j,\sigma)=\mathfrak{p}(x\sigma)/\mathfrak{p}(x)$ where $x\in q_j$;
5 **end**

Corollary 1 (to Theorem 4)**:** *A PFSA $P_i=(Q,\Sigma,\delta,q_i,\tilde{\pi})$ induces a probability measure $\mathfrak{p}$ on the σ-algebra $\mathfrak{B}_\Sigma$ and the corresponding probabilistic Nerode equivalence is of finite index.*

Proof: Let a probability measure $\mathfrak{p}$ be constructed on the σ-algebra $\mathfrak{B}_\Sigma$ as follows:

$$\forall \tau \in \Sigma^+, \left(\mathfrak{p}(\tau) = \tilde{\pi}(q_i, \sigma_1) \prod_{k=1}^{r-1} \tilde{\pi}(\delta^*(q_i, \sigma_1 \cdots \sigma_k), \sigma_{k+1}) \right).$$

It follows from Definition 10 that P_i perfectly encodes the measure $\mathfrak{p}$ and Theorem 4 implies that the corresponding $\mathcal{N}_\mathfrak{p}$ is of finite index. □

On account of Corollary 1, we can map any given PFSA to a measure space $(\Sigma^\omega, \mathfrak{B}_\Sigma, \mathfrak{p})$.

Definition 11: Let $\mathscr{P}$ be the space of all probability measures on $\mathfrak{B}_\Sigma$ and $\mathscr{A}$ be the space of all possible PFSA $P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi})$.

- The map $\mathbb{H}: \mathscr{A} \to \mathscr{P}$ is defined as $\mathbb{H}(P_i) = \mathfrak{p}$ such that

$$\forall \tau \in \Sigma^+, \left(\mathfrak{p}(\tau) = \tilde{\pi}(q_i, \sigma_1) \prod_{k=1}^{r-1} \tilde{\pi}(\delta^*(q_i, \sigma_1 \cdots \sigma_k), \sigma_{k+1}) \right)$$
$$\text{where } \tau = \sigma_1\sigma_2 \cdots \sigma_r. \tag{11}$$

- The map $\mathbb{H}_{-1} : \mathscr{P} \to \mathscr{A}$ is defined as

$$\mathbb{H}_{-1}(\mathfrak{p}) = \begin{cases} P_i \text{ given by Algo.1} & \text{if } \mathcal{N}_\mathfrak{p} \text{ is of finite index} \\ \text{Undefined} & \text{otherwise.} \end{cases} \tag{12}$$

Lemma 1: *P_i is a perfect encoding for $\mathbb{H}(P_i)$.*

Proof: The proof follows from Definition 10 and Definition 11. □

Next we show that, similar to classical finite state machines, an arbitrary PFSA can be uniquely minimized. However, the sense in which the minimization is achieved is somewhat different. To this end, we introduce the notion of reachable states in a PFSA and define isomorphism of two PFSA.

Definition 12 (reachable states)**:** Given a PFSA $P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi})$, the set of reachable states $\mathrm{RCH}(P_i) \subseteqq Q$ is defined as:

$$\tilde{q} \in \mathrm{RCH}(P_i) \Rightarrow \exists \tau = \sigma_1 \cdots \sigma_R \in \Sigma^*$$
$$\text{such that } (\delta^*(q_i, \tau) = \tilde{q})$$
$$\wedge \left(\tilde{\pi}(q_i, \sigma_i) \prod_{r=1}^{R-1} \tilde{\pi}(\delta^*(q_i, \sigma_1 \cdots \sigma_r), \sigma_{r+1}) > 0 \right).$$

Remark 3: The strict positivity condition in Definition 12 ensures that every state in the set of reachable states can actually be attained with a strictly non-zero probability. In other words, for every state $q_j \in \mathrm{RCH}(P_i)$, there exists at least one string ω, initiating from q_i and eventually terminating on state q_j, such that the generation probability of ω is strictly positive.

Definition 13 (Isomorphism)**:** Two PFSA $P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi})$ and $P'_{i'} = (Q', \Sigma, \delta', q'_{i'}, \tilde{\pi}')$ are defined to be isomorphic if there exists a bijective map $\eta: \mathrm{RCH}(P_i) \to \mathrm{RCH}(P'_{i'})$ such that

$$\tilde{\pi}(q_j, \sigma) \neq 0 \Rightarrow \left(\tilde{\pi}'(\eta(q_j), \sigma) = \tilde{\pi}(q_j, \sigma) \right)$$
$$\bigwedge \left(\delta'(\eta(q_j), \sigma_k) = \eta(\delta(q_j, \sigma_k)) \right).$$

Remark 4: The notion of isomorphism stated in Definition 13 generalizes graph isomorphism to PFSA by considering only the states that can be reached with non-zero probability and transitions that have a non-zero probability of occurrence.

Theorem 5 (Minimization of PFSA)**:** *For a PFSA $P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi})$, $\mathbb{H}_{-1} \circ \mathbb{H}(P_i)$ is the unique minimal realization of P_i in the sense that the following conditions are satisfied:*

(1) *The PFSA $\mathbb{H}_{-1} \circ \mathbb{H}(P_i)$ perfectly encodes the probability measure $\mathbb{H}(P_i)$.*
(2) *For a PFSA $P'_{i'}$ that perfectly encodes $\mathbb{H}(P_i)$, the inequality $\mathrm{CARD}(\mathrm{RCH}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i))) \leqq \mathrm{CARD}(\mathrm{RCH}(P'_{i'}))$ holds.*
(3) *The equality, $\mathrm{CARD}(\mathrm{RCH}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i))) \leqq \mathrm{CARD}(\mathrm{RCH}(P'_{i'}))$, implies isomorphism of P_i and $P'_{i'}$ in the sense of Definition 13.*

Proof:

(1) The proof follows from the construction in Theorem 4.
(2) Let $P'_{i'} = (Q', \Sigma, \delta', q'_{i'}, \tilde{\pi}')$ be an arbitrary PFSA that perfectly encodes the probability measure $\mathbb{H}(P_i)$. Let us construct a PFSA $P^\dagger_{i'} = (Q' \cup \{q_d\}, \Sigma, \delta^\dagger, q_{i'}, \tilde{\pi}^\dagger)$, where q_d is a new state not in Q', as follows:

$$\forall q'_j \in Q', \sigma \in \Sigma,$$
$$\delta^\dagger(q'_j \sigma_k) = \begin{cases} q_d & \text{if } \tilde{\pi}'(q'_j \sigma_k) = 0 \\ \delta'(q'_j \sigma_k) & \text{otherwise} \end{cases} \tag{13a}$$

$$\forall \sigma \in \Sigma, \ \delta^\dagger(q_d, \sigma_k) = q_d \tag{13b}$$

$$\forall q'_j \in Q', \ \forall \sigma \in \Sigma, \ \tilde{\pi}^\dagger(q'_j \sigma_k) = \tilde{\pi}'(q'_j \sigma_k). \tag{13c}$$

It is seen that $P^\dagger_{i'}$ perfectly encodes $\mathbb{H}(P_i)$ as well, which follows from Definition 10 and equation (13c). It is claimed that

$$\mathrm{CARD}(\mathrm{RCH}(P^\dagger_{i'})) = \mathrm{CARD}(\mathrm{RCH}(P'_{i'})) \tag{14}$$

based on the following rationale.

Let $q'_j \in \mathrm{RCH}(P'_{i'})$. Following Definition 12, there exists a string $\tau \in \Sigma^*$ such that $\delta'^*(q_{i'}, x) = q'_j$ and $\tilde{\pi}(q_{i'}, \sigma_1) \prod_{r=1}^{R-1} \tilde{\pi}'(\delta'^*(q_{i'}, \sigma_1 \cdots \sigma_r), \sigma_{r+1}) > 0$. It follows from Equation (13c) that $\tilde{\pi}^\dagger(q_{i'}, \sigma_1) \prod_{r=1}^{R-1} \tilde{\pi}^\dagger \times (\delta^{\dagger *}(q_{i'}, \sigma_1 \cdots \sigma_r), \sigma_{r+1}) > 0$ and hence we conclude using equation (13a) that $\delta^{\dagger *}(q_{i'}, x) = q_j \neq q_d$ which then implies that $q'_j \in \text{Rch}(P^{\dagger}_{i'})$. Hence we have $\text{Card}(\text{Rch}(P'_{i'})) \leqq \text{Card}(\text{Rch}(P^{\dagger}_{i'})$. By a similar argument, we have $\text{Card}(\text{Rch}(P^{\dagger}_{i'}) \leqq \text{Card}(\text{Rch}(P'_{i'}))$ and hence $\text{Card}(\text{Rch}(P^{\dagger}_{i'}) \leqq \text{Card}(\text{Rch}(P'_{i'}))$.

Next, we claim

$$\forall x, y \in \Sigma^*\left(\left(\delta^{\dagger}(q_{i'}, x) = \delta^{\dagger}(q_{i'}, y)\right) \Rightarrow x\mathcal{N}_{\mathbb{H}(P_i)}y\right) \quad (15)$$

based on the following rationale.

Let $x, y \in \Sigma^*$ s.t. $(\delta^{\dagger}(q_{i'}, x) = \delta^{\dagger}(q_{i'}, y))$. It follows from equations (13a–c) that

$$\begin{cases} (\mathbb{H}(P_i)(x) = 0 \wedge \mathbb{H}(P_i)(y) = 0) & \text{if } \delta^{\dagger}(q_{i'}, x) = q_d \\ (\mathbb{H}(P_i)(x) \neq 0 \wedge \mathbb{H}(P_i)(y) \neq 0) & \text{otherwise.} \end{cases} \quad (16)$$

Now, if $(\mathbb{H}(P_i)(x) = 0 \wedge \mathbb{H}(P_i)(y) = 0)$, then it follows from equation (6b) that $x\mathcal{N}_{\mathbb{H}(P_i)}y$. On the other hand, if $(\mathbb{H}(P_i)(x) \neq 0 \wedge \mathbb{H}(P_i)(y) \neq 0)$, then equation (6a) yields

$$\forall u = \sigma_1 \cdots \sigma_R \in \Sigma^*, \ \frac{\mathbb{H}(P_i)(xu)}{\mathbb{H}(P_i)(x)} = \frac{\mathbb{H}(P_i)(yu)}{\mathbb{H}(P_i)(y)}$$
$$= \tilde{\pi}^{\dagger}(q_{i'}, \sigma_1) \prod_{r=1}^{R-1} \tilde{\pi}^{\dagger}\left(\delta^{\dagger}(q_{i'}, \sigma_1 \cdots \sigma_r), \sigma_{r+1}\right) \Rightarrow x\mathcal{N}_{\mathbb{H}(P_i)}y.$$

We define a map $\zeta : \text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i)) \rightarrow \text{Rch}(P^{\dagger}_{i'})$ as follows: Let $q^{\#} \in \text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i))$ and let $\varepsilon(q^{\#})$ be the equivalence class of the relation $\mathcal{N}_{\mathbb{H}(P_i)}$ represented by $q^{\#}$. Let $x = \sigma_1 \cdots \sigma_R \in \varepsilon(q^{\#})$.

$$\mathbb{H}(P_i)(x) > 0 \text{ (See Definition 12)}$$
$$\Rightarrow \tilde{\pi}^{\dagger}(q_{i'}, \sigma_1) \prod_{r=1}^{R-1} \tilde{\pi}^{\dagger}\left(\delta^{\dagger *}(q_{i'}, \sigma_1 \cdots \sigma_r), \sigma_{r+1}\right) > 0$$
$$\text{(Since } P^{\dagger}_{i'} \text{ perfectly encodes } \mathbb{H}(P_i))$$
$$\Rightarrow \delta^{\dagger *}(q_{i'}, x) \in \text{Rch}(P^{\dagger}_{i'}).$$

Let $\zeta(q^{\#}) = \delta^{\dagger *}(q_{i'}, x)$. Note that $\zeta(q^{\#})$ depends on the choice of x. Let $q^{\#}_1, q^{\#}_2 \in \text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i))$ such that $\zeta(q^{\#}_1) = \zeta(q^{\#}_2)$. If x_1, x_2 are the corresponding strings chosen to define $\zeta(q^{\#}_1), \zeta(q^{\#}_2)$, we have $\delta^{\dagger *}(q_{i'}, x_1) = \delta^{\dagger *}(q_{i'}, x_2)$ which implies $x_1\mathcal{N}_{\mathbb{H}(P_i)}x_2$, i.e., $q^{\#}_1 = q^{\#}_2$. Hence we conclude ζ is injective which, in turn, implies

$$\text{Card}(\text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i))) \leqq \text{Card}(\text{Rch}(P^{\dagger}_{i'})). \quad (17)$$

Finally, from equations (14) and (17), it follows that

$$\text{Card}(\text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i))) \leqq \text{Card}(\text{Rch}(P'_{i'})). \quad (18)$$

Let $P'_{i'} = (Q', \Sigma, \delta', q_{i'}, \tilde{\pi}')$ be an arbitrary PFSA that perfectly encodes $\mathbb{H}(P_i)$ such that

$$\text{Card}(\text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i))) \leqq \text{Card}(\text{Rch}(P'_{i'})). \quad \text{(C1)}$$

Let the PFSA $\mathbb{H}_{-1} \circ \mathbb{H}(P_i)$ be denoted as $(Q^{\#}, \Sigma, \delta^{\#}, q^{\#}_{i^{\#}}, \tilde{\pi}^{\#})$. Let $\mathscr{E}(q^{\#}_j)$ denote the equivalence class of $\mathcal{N}_{\mathbb{H}(P_i)}$ that $q^{\#}$ represents. We define a map $\phi\colon \text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i)) \rightarrow 2^{\text{Rch}(P'_{i'})}$ as follows:

$$\phi(q^{\#}_j) = \left\{q'_j \in Q' \middle| \exists x \in \mathscr{E}(q^{\#}_j) \text{ s.t. } \delta'^{*}(q_{i'}, x) = q'_j\right\}. \quad (19)$$

We claim

$$\forall q^{\#}_j, q^{\#}_k \in Q^{\#}\left(\left(q^{\#}_j \neq q^{\#}_k\right) \Rightarrow \left(\phi(q^{\#}_j) \bigcap \phi(q^{\#}_k) = \emptyset\right)\right). \quad \text{(C2)}$$

Let $q'_\ell \in \phi(q^{\#}_j) \cap \phi(q^{\#}_k)$. Hence there exists $x_j \in \mathscr{E}(q^{\#}_j)$, $x_k \in \mathscr{E}(q^{\#}_k)$ such that

$$q'_\ell = \delta'^{*}(q_{i'}, x_j) = \delta'^{*}(q_{i'}, x_k) \quad (20)$$

that

$$x_j\mathcal{N}_{\mathbb{H}(P_i)}x_k \Rightarrow \exists u \in \Sigma^*\left(\frac{\mathbb{H}(P_i)(x_j u)}{\mathbb{H}(P_i)(x_j)} \neq \frac{\mathbb{H}(P_i)(x_k u)}{\mathbb{H}(P_i)(x_k)}\right) \quad (21)$$

but, $P'_{i'}$ perfectly encodes $\mathbb{H}(P_i)$ implying

$$\forall u = \sigma_1 \cdots \sigma_R \in \Sigma^*\left(\frac{\mathbb{H}(P_i)(x_j u)}{\mathbb{H}(P_i)(x_j)} \neq \frac{\mathbb{H}(P_i)(x_k u)}{\mathbb{H}(P_i)(x_k)}\right.$$
$$\left. = \tilde{\pi}'(q'_\ell, \sigma_1) \prod_{r=1}^{R-1} \tilde{\pi}'(\delta'^{*}(q_{i'}, \sigma_1 \cdots \sigma_r), \sigma_{r+1}\right)$$

which contradicts equation (21).

Next we claim that

$$\forall q^{\#}_j \in \text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i)) \ \text{Card}(\phi(q^{\#}_j)) = 1 \quad \text{(C3)}$$

Let $x_1, x_2 \in \mathcal{E}(q^{\#}_j)$ such that

$$\left.\begin{array}{l} \delta'^{*}(q_i, x_1) = q'_j \\ \delta'^{*}(q_i, x_2) = q'_k \end{array}\right\} \quad \text{with } q'_j \neq q'_k. \quad (22)$$

Therefore,

$$\text{Card}(\phi(q^{\#}_j)) > 1$$
$$\Longrightarrow \sum_{q^{\#}_k \in \text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i))'} \text{Card}(\phi(q^{\#}_k))$$
$$> \text{Card}(\text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i)))$$
$$\Longrightarrow \text{Card}(\text{Rch}(P'_{i'})) > \text{Card}(\text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i)))$$

which contradicts **C1** thus proving **C3**.

On account of **C2** and **C3**, let us define a bijective map $\tilde{\phi}\colon \text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i)) \rightarrow \text{Rch}(P'_{i'})$ as $\tilde{\phi}(q^{\#}_j) = \delta'^{*}(q_i, x)$, $x \in \mathscr{E}(q^{\#}_j)$. Then,

$$\left.\begin{array}{l} \forall \sigma_k \in \Sigma, \forall q^{\#}_j \in \text{Rch}(\mathbb{H}_{-1} \circ \mathbb{H}(P_i))), x \in \mathscr{E}(q^{\#}_j) \\ \tilde{\pi}^{\#}(q^{\#}_j, \sigma_k) = \dfrac{\mathbb{H}(P_i)(x\sigma_k)}{\mathbb{H}(P_i)(x)} = \tilde{\pi}^{\#}(\tilde{\phi}(q^{\#}_j), \sigma_k) \end{array}\right\} \quad (23)$$

$$\tilde{\phi}(\delta^{\#}(q_j^{\#}, \sigma_k)) = \delta'^{*}(q_{i'}, x\sigma_k)$$
$$= \delta'(\delta'^{*}(q_{i'}, \sigma_k), \sigma_k) = \delta'(\tilde{\phi}(q_j^{\#}), \sigma_k) \quad (24)$$

which implies that $\mathbb{H}_{-1} \circ \mathbb{H}(P_i)$ and $P'_{i'}$, are isomorphic in the sense of Definition 13. This completes the proof. □

Theorem 6: *For a PFSA $P_i = (\mathscr{E}, \Sigma, \delta, E_i, \tilde{\pi})$, the function $\tilde{\pi} : Q \times \Sigma \to Q$ can be extended to $\tilde{\pi} : Q \times \Sigma^* \to Q$ as*

$$\forall q_j \in Q, \quad \tau \in \Sigma^*,$$
$$\sigma \in \Sigma, \begin{cases} \tilde{\pi}(q_j, \varepsilon) = 1 \\ \tilde{\pi}(q_j, \sigma\tau) = \tilde{\pi}(q_j, \sigma)\tilde{\pi}(q_j, \sigma)\tilde{\pi}(\delta(q_j, \sigma), \tau). \end{cases} \quad (25)$$

Proof: Let $\nu = \mathbb{H}(P_i)$. We note that that P_i perfectly encodes $\mathfrak{p}$ (see Lemma 1). It follows from Theorem 4 that

$$\forall q_j \in Q, \tilde{\pi}(q_j, \varepsilon) = \frac{\mathfrak{p}(x\varepsilon)}{\mathfrak{p}(x)} = \frac{\mathfrak{p}(x)}{\mathfrak{p}(x)} = 1 \quad \text{where } \delta^*(q_i, x) = q_j.$$

Similarly, for a string $\sigma\tau$ initiating from state q_j, where $\sigma \in \Sigma, \tau \in \Sigma^*$, we have

$$\tilde{\pi}(q_j, \sigma\tau) = \frac{\mathfrak{p}(x\sigma\tau)}{\mathfrak{p}(x)} = \frac{\mathfrak{p}(x\sigma)}{\mathfrak{p}(x)} \times \frac{\mathfrak{p}(x\sigma\tau)}{\mathfrak{p}(x\sigma)}. \quad (26)$$

We note that $\mathfrak{p}(x\sigma)/\mathfrak{p}(x) = \tilde{\pi}(q_j, \sigma)$. Also, $\delta^*(q_i, x) = q_j$ implies $\delta(q_j, \sigma) = \delta^*(q_i, x\sigma)$. Therefore, $\mathfrak{p}(x\sigma\tau)/\mathfrak{p}(x\sigma) = \tilde{\pi}(\delta(q_j, \sigma), \tau)$ and hence

$$\tilde{\pi}(q_j, \sigma\tau) = \tilde{\pi}(q_j, \sigma)\tilde{\pi}(\delta(q_j, \sigma), \tau).$$

This completes the proof. □

Theorem 7: *For a measure space $(\Sigma^\omega, \mathfrak{B}_\Sigma, \mathfrak{p})$,*

$$\mathbb{H} \circ \mathbb{H}_{-1}(\mathfrak{p}) = \mathfrak{p} \quad (27)$$

i.e., $\mathbb{H} \circ \mathbb{H}_{-1}$ is the identity map from $\mathscr{P}$ onto itself.

Proof: Let $\mathbb{H}_{-1}(\mathfrak{p}) = P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi})$. We note P_i perfectly encodes $\mathfrak{p}$ (See Lemma 3.1). Let $\mathbb{H}(P_i) = \mathfrak{p}'$. We claim

$$\forall x \in \Sigma^*, \mathfrak{p}(x) = \mathfrak{p}'(x).$$

The result is immediate for $|x| = 0$, i.e., $x = \epsilon$. For $|x| \geq 1$, we proceed by the method of induction. For $|x| = 1$, we note

$$\forall \sigma \in \Sigma, \mathfrak{p}'(\sigma) = \tilde{\pi}(q_i, \sigma) = \mathfrak{p}(\sigma) \text{ (perfect Encoding).}$$

Next let us assume that $\forall x \in \Sigma^*$, s.t. $|x| = r \in \mathbb{N}$, $\mathfrak{p}'(x) = \mathfrak{p}(x)$. Since $\forall x \in \Sigma^*$ with $|x| = r \in \mathbb{N}$, it follows that

$$\mathfrak{p}'(x\sigma) = \mathfrak{p}'(x)\tilde{\pi}(q_j, \sigma) \text{ where } \delta^*(q_i, x) = q_i$$
$$= \mathfrak{p}(x)\tilde{\pi}(q_i, \sigma) = \mathfrak{p}(x\sigma).$$

This completes the proof. □

4. Metrization of the space $\mathscr{P}$ of probability measures on $\mathfrak{B}_\Sigma$

Metrization of $\mathscr{P}$ is important for differentiating physical processes modeled as dynamical systems evolving probabilistically on discrete state spaces of finite cardinality. In this section, we introduce two metric families, each of which captures a different aspect of such dynamical behaviour and can be combined to form physically meaningful and useful metrics for system analysis and design.

Definition 14: Given two probability measures $\mathfrak{p}_1, \mathfrak{p}_2$ on the σ-algebra $\mathfrak{B}_\Sigma$ and a parameter $s \in [1, \infty]$, the function $d_s : \mathscr{P} \times \mathscr{P} \to [0, 1]$ is defined as follows:

$$d_s(\mathfrak{p}_1, \mathfrak{p}_2) = \sup_{x \in \Sigma^*} \left(\sum_{j=1}^{|\Sigma|} \left| \frac{\mathfrak{p}_1(x\sigma_j)}{\mathfrak{p}_1(x)} - \frac{\mathfrak{p}_2(x\sigma_j)}{\mathfrak{p}_2(x)} \right|^s \right)^{1/s}$$
$$\forall s \in [1, \infty) \quad (28a)$$

$$d_\infty(\mathfrak{p}_1, \mathfrak{p}_2) = \sup_{x \in \Sigma^*} \max_{\sigma \in \Sigma} \left| \frac{\mathfrak{p}_1(x\sigma_j)}{\mathfrak{p}_1(x)} - \frac{\mathfrak{p}_2(x\sigma_j)}{\mathfrak{p}_2(x)} \right|. \quad (28b)$$

Theorem 8: *The space $\mathscr{P}$ of all probability measures on $\mathfrak{B}_\Sigma$ is d_s-metrizable for $s \in [1, \infty]$.*

Proof: Strict positivity and symmetry properties of a metric follow directly from Definition 14. Validity of the remaining property of triangular inequality follows by application of Minkowski inequality (Rudin 1988). □

Definition 15: Let $\mathcal{M}$ be a right invariant equivalence relation on Σ^* and the ith equivalence class of $\mathcal{M}$ be denoted as $\mathcal{M}^i$, $i \in I$, where I is an arbitrary index set. Let $\mathfrak{p}$ be a probability measure on the σ-algebra $\mathfrak{B}_\Sigma$ inducing the probabilistic Nerode equivalence $\mathcal{N}_\mathfrak{p}$ on Σ^* with the jth equivalence class of $\mathcal{N}_\mathfrak{p}$ denoted as $\mathcal{N}_\mathfrak{p}^j$, $j \in \mathcal{J}$, where $\mathcal{J}$ is an index set distinct from $\mathcal{I}$. Then, the map $\Omega_\mathcal{M} : \mathscr{P} \to [0,1]^{\text{CARD}(\mathcal{I})} \times [0,1]^{\text{CARD}(\mathcal{J})}$ is defined as

$$\Omega_\mathcal{M}(\mathfrak{p})\big|_{ij} = \sum_{x \in \mathcal{M}^i \cap \mathcal{N}_\mathfrak{p}^j} \mathfrak{p}(x).$$

Definition 16: Let $\mathfrak{p}_1, \mathfrak{p}_2$ be two probability measures on the σ-algebra $\mathfrak{B}_\Sigma$. Then, the function $d_F : \mathscr{P} \times \mathscr{P} \to [0, 1]$ is defined as follows:

$$d_F(\mathfrak{p}_1, \mathfrak{p}_2) = \left\| \Omega_{\mathcal{N}_{\mathfrak{p}_2}}(\mathfrak{p}_1) - \Omega_{\mathcal{N}_{\mathfrak{p}_2}}(\mathfrak{p}_2) \right\|_F, \quad (29)$$

where $\|\Theta\|_F = \sqrt{\text{Trace}[\Theta^\text{H}\Theta]}$ is the Frobenius norm of the operator Θ, and Θ^H is the Hermitian of Θ.

Definition 16 implies that if $\mathcal{I}$ and $\mathcal{J}$ are the index sets corresponding to $\mathcal{N}_{\mathfrak{p}_1}$ and $\mathcal{N}_{\mathfrak{p}_2}$ respectively, then $\Omega_{\mathcal{N}_{\mathfrak{p}_1}}(\mathfrak{p}_2) \in [0,1]^{\text{CARD}(\mathcal{I})} \times [0,1]^{\text{CARD}(\mathcal{J})}$ and $\Omega_{\mathcal{N}_{\mathfrak{p}2}}(\mathfrak{p}_1) \in [0,1]^{\text{CARD}(\mathcal{J})} \times [0,1]^{\text{CARD}(\mathcal{I})}$.

Theorem 9: *The function d_F is a pseudometric on the space $\mathscr{P}$ of probability measures.*

Proof: The Frobenius norm on a probability space satisfies the metric properties except strict positivity because of the almost sure property of a probability measure. □

Theorem 10: *For $\forall \alpha \in [0,1)$ and $\forall s \in [1,\infty)$, the parametized function $\mu_{\alpha,s} \triangleq \alpha d_F + (1-\alpha)d_s$ is a metric on $\mathscr{P}$.*

Proof: Following Theorems 8 and 9, d_s is a metric for $s \in [1,\infty]$ and d_F is a pseudometric on $\mathscr{P}$. Non-negativity, finiteness, symmetry and sub-additivity of $\mu_{\alpha,s}$ follow from the respective properties of d_F and d_s. Strict positivity of $\mu_{\alpha,s}$ on $\alpha \in [0,1)$ is established below

$$\mu_{\alpha,s}(\mathfrak{p}_1,\mathfrak{p}_2) = 0 \Rightarrow (1-\alpha)d_s(\mathfrak{p}_1,\mathfrak{p}_2) = 0 \Rightarrow \mathfrak{p}_1 = \mathfrak{p}_2. \tag{30}$$

□

Remark 5: If two physical processes are modelled as discrete-event dynamical systems, then the respective probabilistic language generators can be associated with probability measures $\mathfrak{p}_1$ and $\mathfrak{p}_2$. The metric $d_s(\mathfrak{p}_1,\mathfrak{p}_2)$ is related to the production of single symbols as arbitrary strings and hence captures the difference in short term dynamic evolution. In contrast, the pseudometric d_F is related to generation of all possible strings and therefore captures the difference in long term behavior of the physical processes. The metric $\mu_{\alpha,s}$ thus captures the both short-term and long-term behaviour with respective relative weights of $1-\alpha$ and α.

Definition 17: The metric $\mu_{\alpha,s}$ on $\mathscr{P}$ for $\alpha \in [0,1)$, $s \in [1,\infty]$, induces a function $\nu_{\alpha,s}$ on $\mathscr{A} \times \mathscr{A}$ as follows:

$$\forall P_i, P'_{i'} \in \mathscr{A}, \nu_{\alpha,s}(P_i, P'_{i'}) = \mu_{\alpha,s}(\mathbb{H}(P_i), \mathbb{H}(P'_{i'})). \tag{31}$$

Corollary 2 (to Theorem 10)**:** *The function $\nu_{\alpha,s}$ in Definition 17 for $\alpha \in [0,1)$ and $s \in [1,\infty]$ is a pseudometric on $\mathscr{A}$. Specifically, the following condition holds*:

$$\nu_{\alpha,s}(P_i, \mathbb{H}_{-1} \circ \mathbb{H}(P_i)) = 0. \tag{32}$$

Proof: Following Theorem 7,

$$\begin{aligned}\nu_{\alpha,s}(P_i, \mathbb{H}_{-1} \circ \mathbb{H}(P_i)) &= \mu_{\alpha,s}(\mathbb{H}(P_i),\ \mathbb{H} \circ \mathbb{H}_{-1} \circ \mathbb{H}(P_i)) \\ &= \mu_{\alpha,s}(\mathbb{H}(P_i), (\mathbb{H} \circ \mathbb{H}_{-1}) \circ \mathbb{H}(P_i)) \\ &= \mu_{\alpha,s}(\mathbb{H}(P_i), \mathbb{H}(P_i)) = 0.\end{aligned}$$

□

Remark 6: Corollary 2 can be physically interpreted to imply that the metric family $\nu_{\alpha,s}$ does not differentiate between different realizations of the same probability measure. Thus when comparing two probabilistic finite state machines, we need not concern ourselves with whether the machines are represented in their minimal realizations; the distance between two non-minimal realizations of the same PFSA is always zero. However this implies that $\nu_{\alpha,s}$ only qualifies as a pseudo-metric on $\mathscr{A}$.

4.1 *Explicit computation of the pseudometric v for PFSA*

The pseudometric $\nu_{\alpha,s}$ is computed explicitly for pairs of PFSA over the same alphabet. Before proceeding to the general case, $\nu_{\alpha,s}$ is computed for the special case, where the pair of PFSA have identical state sets, initial states and transition maps.

Lemma 2: *Given two PFSA $P_i^1 = (Q,\Sigma,\delta,q_i,\tilde{\pi}^1)$, $P_i^2 = (Q,\Sigma,\delta,q_i,\tilde{\pi}^2)$, and $\sigma \in \Sigma$, the steps for computation of $\nu_{0,s}(P_i^1,P_i^2)$ are*

$$\begin{aligned}&\textit{Set } \Delta(q_j) = \tilde{\pi}^1(q_j,\sigma) - \tilde{\pi}^2(q_j,\sigma) \\ &\textit{Then, } \nu_{0,s}(P_i^1,P_i^2) = \max_{q_j \in Q} \| \Delta(q_j) \|_s.\end{aligned}$$

Proof: Let $\mathbb{H}(P_i^2)(x\sigma)/\mathbb{H}(P_i^2)(x)$ denote a $|\sum|$-dimensional vector-valued function, where $\sigma \in \Sigma$. For proof of the lemma, it suffices to show that the following relation holds:

$$\sup_{x \in \Sigma^*} d_s\left(\frac{\mathbb{H}(P_i^1)(x\sigma)}{\mathbb{H}(P_i^1)(x)}, \frac{\mathbb{H}(P_i^2)(x\sigma)}{\mathbb{H}(P_i^2)(x)}\right) = \max_{q_i \in Q} \| \Delta(q_j) \|_s.$$

Since P_i^1 perfectly encodes $\mathbb{H}(P_i^1)$, it follows that $(\forall x, y \in \Sigma^*, \delta(q_i,x) = \delta(q_i,y))$ implies

$$\frac{\mathbb{H}(P_i^1)(x\sigma_k)}{\mathbb{H}(P_i^1)(x)} = \tilde{\pi}^1(q_j,\sigma_k) = \frac{\mathbb{H}(P_i^1)(y\sigma)}{\mathbb{H}(P_i^1)(y)},$$

where $\delta(q_i,\ x) = q_j$. Similar argument holds for $\mathbb{H}(P_i^2)$. Hence, it follows that for computing $\nu_{0,s}(P_i^1,P_i^2)$, only one string needs to be considered for each state $q_j \in Q$. That is,

$$\begin{aligned}\nu_{0,s}(P_i^1,P_i^2) &= \max_{x:\delta(q_i,x)=q_j} \left\| \frac{\mathbb{H}(P_i^1)(x\sigma_k)}{\mathbb{H}(P_i^1)(x)} - \frac{\mathbb{H}(P_i^2)(x\sigma_k)}{\mathbb{H}(P_i^2)(x)} \right\|_s \\ &= \max_{x:\delta(q_i,x)=q_j} \| \tilde{\pi}^1(q_j,\sigma_k) - \tilde{\pi}^2(q_j,\sigma_k) \|_s \\ &= \max_{q_j \in Q} \| \Delta(q_j) \|_s.\end{aligned}$$

Lemma 3: *Let $\wp_1, \wp_2$ be the stable probability distributions for PFSA $P_i^1 = (Q,\Sigma,\delta,q_i,\tilde{\pi}^1)$ and $P_i^2 = (Q,\Sigma,\delta,q_i,\tilde{\pi}^2)$ respectively. Then,*

$$\lim_{\alpha \to 1} \nu_{\alpha,s}(P_i^1,P_i^2) = d_2(\wp_1,\wp_2)$$

Proof: Since P_i^1 and P_i^2 have the same initial state and state transition maps,

$$\mathcal{N}^j_{\mathbb{H}(P_i^1)} \bigcap \mathcal{N}^k_{\mathbb{H}(P_i^2)} = \begin{cases} \oslash & \text{if } j \neq k \\ \mathcal{N}^j_{\mathbb{H}(P_i^1)} = \mathcal{N}^k_{\mathbb{H}(P_i^2)} & \text{otherwise} \end{cases}$$

where $\mathcal{N}^j_{\mathbb{H}(P_i^1)}$ and $\mathcal{N}^k_{\mathbb{H}(P_i^2)}$ are the jth and kth equivalence classes (i.e., states q_j and q_k) for P_i^1, P_i^2, respectively. The result follows from Definition 16 and Corollary 11 and noting that

$$\Omega_{\mathcal{N}_{\mathbb{H}(P_i^2)}}(\mathbb{H}(P_i^1))\Big|_{jk} = \begin{cases} \sum\limits_{x:\delta(q_i,x)=q_j} \mathfrak{p}(x) = \wp_1\big|_j & \text{if } j = k \\ 0 & \text{otherwise} \end{cases}.$$

Theorem 11: *Given two PFSA $P_i^1 = (Q, \Sigma, \delta, q_i, \tilde{\pi}^1)$ and $P_i^2 = (Q, \Sigma, \delta, q_i, \tilde{\pi}^2)$, the pseudometric $\nu_{\alpha,s}(P_i^1, P_i^2)$ can be computed explicitly for $\alpha \in [0, 1)$ and $s \in [1, \infty]$ as*

$$\nu_{\alpha,s}(P_i^1, P_i^2) = \alpha \lim_{\alpha \to 1} \nu_{\alpha,s}(P_i^1, P_i^2) + (1-\alpha)\nu_{0,s}(P_i^1, P_i^2). \tag{33}$$

Proof: The result follows from Theorem 10 and Corollary 2. □

The algorithm for computation of the the pseudometric ν is presented below.

To extend the approach presented in Lemma 2 to arbitrary pairs of PFSA, we need to define the synchronous composition of a pair of PFSA.

Algorithm 2: Computation of $\nu_{\alpha,s}(P_i, P'_{i'})$

input: $P_i, P'_{i'}, s, \alpha$
output: $\nu_{\alpha,s}(P_i, P'_{i'})$
begin
  Compute $P_{12} = (\bar{Q}, \Sigma, \bar{\delta}, \tilde{\pi}_{12}) = P_i \otimes P'_{i'}$;
  Compute $P_{21} = (\bar{Q}, \Sigma, \bar{\delta}, \tilde{\pi}_{21}) = P'_{i'} \otimes P_i$;
  for $j = 1$ **to** CARD($\bar{Q}$) **do**
    $|\Delta(J) = \|\tilde{\pi}_{12}(q_j, \sigma_k) - \tilde{\pi}_{21}(q_j, \sigma_k)\|s$;
  endfor
  $\nu_{0,s}(P_i, P'_{i'}) = \max_j \Delta(j)$;
  Compute $\wp_{12}$; /∗ State prob. for P_{12} (Def. 2.3) ∗/
  Compute $\wp_{21}$; /∗ State prob. for P_{21} (Def. 2.3) ∗/
  Compute $d = \|\wp_{12} - \wp_{21}\|_2$;
  Compute $\nu_{\alpha,s}(P_i, P'_{i'}) = \alpha d + (1-\alpha)\nu_{0,s}(P_i, P'_{i'})$;
end

Definition 18: The binary operation of synchronous composition of PFSA, denoted as $\otimes : \mathscr{P} \times \mathscr{P} \to \mathscr{P}$, is defined as follows:

$$\text{Let} \begin{cases} P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi}) \\ G_{i'} = (Q', \Sigma, \delta', q'_{i'}, \tilde{\pi}'). \end{cases}$$

Then, $P_i \otimes G_{i'} = (Q \times Q', \Sigma, \delta^{\otimes}, (q_i, q'_{i'}), \tilde{\pi}^{\otimes})$, where

$$\delta^{\otimes}((q_j, q'_{k'}), \sigma) = \begin{cases} \Big(\delta(q_j, \sigma), \delta'(q'_{j'}, \sigma)\Big), & \text{if } \delta(q_j, \sigma) \text{ and } \delta'(q'_{j'}, \sigma) \text{ are defined} \\ \text{Undefined} & \text{otherwise} \end{cases} \tag{34}$$

$$\tilde{\pi}^{\otimes}((q_j, q'_{k'}), \sigma) = \tilde{\pi}(q_j, \sigma). \tag{35}$$

Remark 7: Synchronous composition for **PFSA** is not commutative, i.e., for an arbitrary pair P_i and $G_{i'}$,

$$P_i \otimes G_{i'} \neq G_{i'} \otimes P_i. \tag{36}$$

Synchronous composition of PFSA is associative, i.e.,

$$\forall P^1_{i1}, P^2_{i2}, P^3_{i3} \in \mathscr{P}, \left(P^1_{i1} \otimes P^2_{i2}\right) \otimes P^3_{i3} = P^1_{i1} \otimes \left(P^2_{i2} \otimes P^3_{i3}\right)$$

Theorem 11: *For a pair of PFSA P_i and $G_{i'}$ over the same alphabet,*

$$\mathbb{H}(P_i) = \mathbb{H}(P_i \otimes G_{i'}). \tag{37}$$

Proof: Let $\mathfrak{p} = \mathbb{H}(P_i)$ and $\mathfrak{p}' = \mathbb{H}(P_i \otimes G_{i'})$. It suffices to show that

$$\forall x \in \Sigma^*, \mathfrak{p}(x) = \mathfrak{p}'(x). \tag{C4}$$

For $|x| = 0$, i.e., $x = \epsilon$, the result is immediate. For $|x| \geqq 1$, we use the method of induction. Since P_i perfectly encodes $\mathbb{H}(P_i)$,

$$\forall \sigma \in \Sigma, \mathfrak{p}(\sigma) = \tilde{\pi}(q_i, \sigma) = \tilde{\pi}^{\otimes}((q_i, q'_{i'}), \sigma) = \mathfrak{p}'(\sigma).$$

Hence **C4** is true for $|x| \leqq 1$.

With the induction hypothesis

$$\forall x \in \Sigma^*, s.t., |x| = r \in \mathbb{N}, \mathfrak{p}(x) = \mathfrak{p}'(x) \tag{38}$$

we proceed with an arbitrary $\sigma \in \Sigma$ to yield

$$\begin{aligned} \mathfrak{p}(x\sigma) &= \mathfrak{p}(x)\tilde{\pi}(q_j, \sigma) \text{ where } \delta(q_i, x) = q_j \\ &= \mathfrak{p}'(x)\tilde{\pi}^{\otimes}((q_j, q'_{j'}), \sigma) \text{ where } \delta^{\otimes *}((q_i, q'_{i'}), x) \\ &= (q_j, q'_{j'}) = \mathfrak{p}'(x\sigma). \end{aligned}$$

This completes the proof. □

Theorem 12: *Given a pair of PFSA P_i, $P'_{i'}$ and an arbitrary parameter $s \in [1, \infty]$, Algorithm 2 computes $\nu_{\alpha,s}(P_i, P'_{i'})$ for $\alpha \in [0, 1)$, $s \in [1, \infty]$.*

Proof: By Theorem 11, $\forall \alpha \in [0, 1)$, $s \in [1, \infty]$,

$$\nu_{\alpha,s}(P_i, P'_{i'}) = \nu_{\alpha,s}(P_i \otimes P'_{i'}, P'_{i'} \otimes P_i). \tag{39}$$

Since $P_i \otimes P'_{i'}$ and $P'_{i'} \otimes P_i$ have the same state sets, initial states and transition maps (see Definition 18), correctness of Algorithm 2 follows from Lemmas 2 and 3. □

Example 1: The theoretical results of § 4 are illustrated with a numerical example. The following PFSA are considered:

$$P^1_{q_1} = (\{q_1, q_2, q_3\}, \{0, 1\}, \delta^1, q_1, \tilde{\pi}^1) \quad (40)$$

$$P^2_{q_A} = (\{q_A, q_B\}, \{0, 1\}, \delta^2, q_A, \tilde{\pi}^2) \quad (41)$$

as shown in figures 6 and 7, respectively, and figure 8 illustrates the computed compositions $P^1_{q_1} \otimes P^2_{q_A}$ (above) and $P^2_{q_A} \otimes P^1_{q_1}$ (below).

Following Algorithm 2, we have

$$\Delta_s = \begin{bmatrix} \|0.4 - 0.9 & 0.6 - 0.1\|_s \\ \|0.2 - 0.9 & 0.8 - 0.1\|_s \\ \|0.3 - 0.9 & 0.7 - 0.1\|_s \\ \|0.3 - 0.7 & 0.7 - 0.3\|_s \\ \|0.4 - 0.7 & 0.6 - 0.3\|_s \\ \|0.2 - 0.7 & 0.8 - 0.3\|_s \end{bmatrix}. \quad (42)$$

As an illustration, we set $s = \infty$. Hence,

$$\Delta_\infty = [0.5 \quad 0.7 \quad 0.6 \quad 0.4 \quad 0.3 \quad 0.5]^T \quad (43)$$

$$\Rightarrow \nu_{0,\infty}(P^1_{q_1}, P^2_{q_A}) = \max(\Delta_\infty) = 0.7. \quad (44)$$

The final state probabilities are computed to be

$$\wp_{P^1 \otimes P^2} = [0.15 \quad 0.07 \quad 0.09 \quad 0.2 \quad 0.22 \quad 0.27] \quad (45)$$

$$\wp_{P^2 \otimes P^1} = [0.29 \quad 0.29 \quad 0.29 \quad 0.043 \quad 0.043 \quad 0.043]. \quad (46)$$

For $\alpha = 0.5$, we have

$$\nu_{0.5,\infty}\left(P^1_{q_1}, P^2_{q_A}\right) = 0.5 d_2\left(\wp_{P^1 \otimes P^2}, \wp_{P^2 \otimes P^1}\right) + 0.5 \times 0.7 = 0.4599.$$

The pseudonorm $\nu_{0,\infty}(P^1_{q_1}, P^2_{q_A}) = 0.7$ is interpreted as follows. There exists a string $x \in \Sigma$ and an event $\sigma \in \Sigma$ such that probability of occurrence of σ, given that x has already occurred, is 70% more in one system compared to the other. Also, the occurrence probability of any event, given an arbitrary string has already occurred, is different by no more than 70% for the two systems. The composition $P^1_{q_1} \otimes P^2_{q_A}$ shown in the upper part of figure 8 is an encoding of the measure $\mathbb{H}(P^1_i)$ and hence is a non-minimal realization of P^1_i, while the composition $P^2_{q_A} \otimes P^1_{q_1}$ shown in the lower part of figure 8 encodes $\mathbb{H}(P^2_{i'})$ and therefore is a non-minimal realization of $P^2_{i'}$. Although the structures of the two compositions are identical in a graph-theoretic sense (i.e. there is a graph isomorphism between the compositions), they represent very different probability distributions on $\mathfrak{B}_\Sigma$.

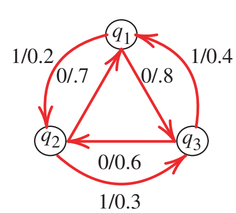

Figure 6. PFSA P^1.

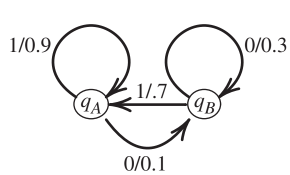

Figure 7. PFSA P^2.

5. Model order reduction for PFSA

This section investigates the possibility of encoding an arbitrary probability distribution on $\mathfrak{B}_\Sigma$ by a PFSA with a pre-specified graph structure. As expected, such encodings will not always be perfect. However, we will show that the error can be rigorously computed and hence is useful for very close approximation of large PFSA models by smaller models.

Definition 19: The binary operation of projective composition $\vec{\otimes} : \mathscr{P} \times \mathscr{P} \to \mathscr{P}$ is defined as follows:

$$\text{Let} \begin{cases} P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi}') \\ G_{i'} = (Q', \Sigma, \delta', q'_{i'}, \tilde{\pi}') \\ G_{i'} \otimes P_i = (Q' \times Q, \Sigma, \delta^{\otimes}, (q'_{i'}, q_i), \tilde{\pi}^{\otimes}). \end{cases}$$

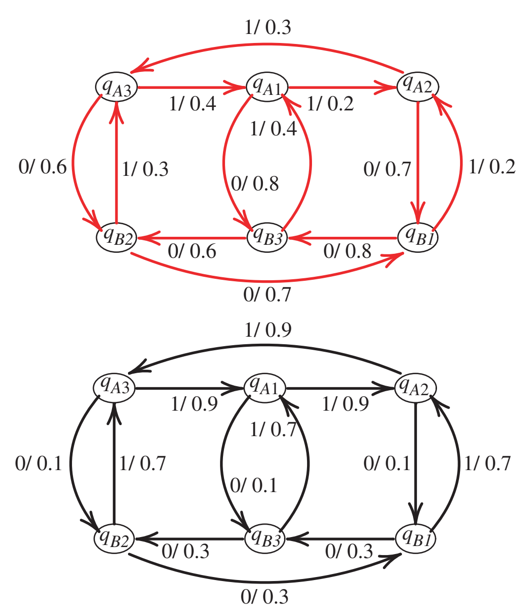

Figure 8. $P^1 \otimes P^2$ (above) and $P^2 \otimes P^1$ (below).

For notational simplicity set $\forall q_j \in Q$ and $\forall q'_{k'} \in Q'$,

$$\vartheta(q'_{k'}, q_j) = \sum_{\substack{x:\delta^{\otimes} * ((q'_{i'}, q_i), x) \\ = (q'_{k'}, q_j)}} (\mathbb{H}(G_{i'})(x)).$$

Then, $G_{i'} \vec{\otimes} P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi}^{\vec{\otimes}})$ s.t.

$$\tilde{\pi}^{\vec{\otimes}}(q_j, \sigma) = \frac{\sum_{q'_{k'} \in Q'} \vartheta(q'_{k'}, q_j) \tilde{\pi}^{\otimes}((q'_{k'}, q_j), \sigma)}{\sum_{q'_{k'}, \in Q'} \vartheta(q'_{k'}, q_j)}. \quad (47)$$

Theorem 13: *For PFSA P_i, G_j, H_k over the same alphabet*,

(1) $P_i \vec{\otimes} (G_j \vec{\otimes} H_k) = P_i \vec{\otimes} H_k$
(2) $(P_i \vec{\otimes} G_j) \vec{\otimes} H_k \neq P_i \vec{\otimes} (G_j \vec{\otimes} H_k)$ *(Non-associative)*
(3) $P_i \vec{\otimes} G_j \neq G_j \vec{\otimes} P_i$ *(Non-commutative)*.

Proof: The results follow from Definition 19. □

We justify the nomenclature 'projective' composition in the following theorem.

Theorem 14: *For arbitrary PFSA P_i and $G_{i'}$ over the same alphabet*,

$$\left(G_{i'} \vec{\otimes} P_i\right) \vec{\otimes} P_i = G_{i'} \vec{\otimes} P_i. \quad (48)$$

Proof: Let $P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi})$. Definition 19 implies that $(G_{i'} \vec{\otimes} P_i) = (Q, \Sigma, \delta, q_i, \tilde{\pi}^{\ddagger}$ for $\tilde{\pi}^{\ddagger}$ computed as specified in equation (47). It further follows from Definition 19, that $(G_{i'} \vec{\otimes} P_i) \vec{\otimes} P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi}^{\vec{\otimes}})$, i.e., $(G_{i'} \vec{\otimes} P_i) \vec{\otimes} P_i$ and $G_{i'} \vec{\otimes} P_i$ have the same state set, initial state and state transition maps. Thus, it suffices to show that

$$\forall q_j \in Q, \sigma \Sigma, \tilde{\pi}^{\ddagger}(q_j, \sigma) = \tilde{\pi}^{\vec{\otimes}}(q_j, \sigma). \quad (49)$$

Considering the probabilistic synchronous composition $(G_{i'} \vec{\otimes} P_i) \otimes P_i = (Q \times Q, \Sigma, \delta^{\otimes}, (q_i, q_i), \tilde{\pi}^{\otimes})$ (see Definition 18),

$$\forall x \in \Sigma^*, \ \delta^{\otimes} * ((q_i, q_i), x) = (q_i, q_j), \text{ for some } q_j \in Q$$

It follows that, for $q_k \neq q_j$,

$$\vartheta(q_k, q_j) = \sum_{\substack{x:\delta^{\otimes *}((q_i, q_i), x) \\ = (q_k, q_j)}} \left(\mathbb{H}\left(G_{i'} \vec{\otimes} P_i\right)(x)\right) = 0. \quad (50)$$

Finally we conclude $\forall q_j \in Q$, $\sigma \in \Sigma$,

$$\begin{aligned} \tilde{\pi}^{\vec{\otimes}}(q_j, \sigma) &= \frac{\sum_{q_k \in Q} \vartheta(q_k, q_j) \tilde{\pi}^{\otimes}((q_k, q_j), \sigma)}{\sum_{q_k \in Q} \vartheta(q_k, q_j)} \\ &= \frac{\vartheta(q_j, q_j) \tilde{\pi}^{\otimes}((q_j, q_j), \sigma)}{\vartheta(q_j, q_j)} \\ &= \tilde{\pi}^{\otimes}((q_j, q_j), \sigma) \\ &= \tilde{\pi}^{\ddagger}(q_j, \sigma) \quad \text{(see Definition 18).} \end{aligned} \quad (51)$$

This completes the proof. □

Projective composition preserves the projected distribution which is defined next.

Definition 20 (projected distribution)**:** The projected distribution $\wp \in [0,1]^{\text{NUMSTATES}(P_i)}$ of an arbitrary PFSA $G_{i'}$ with respect to a given PFSA P_i is defined by the map $[\![\cdot]\!] P_i : \mathscr{A} \to [0,1]^{\text{NUMSTATES}(P_i)}$ as follows:

$$[\![G_{i'}]\!] P_i = \wp \in [0,1]^{\text{NUMSTATES}(P_i)},$$

such that if N^j is the jth equivalence class (i.e. the jth state) of P_i,

$$\text{then} \sum_{x \in N^j} \mathbb{H}(G_{i'})(x) = \wp_j.$$

We note $[\![G_{i'}]\!] P_i$ is a probability vector, i.e.,

$$\sum_{j=1}^{\text{NUMSTATES}} [\![G_{i'}]\!] P_i \Big|_j = \sum_{x \in \Sigma^*} \mathbb{H}(G_{i'})(x) = 1. \quad (52)$$

Theorem 15: (Projected Distribution Invariance)**:** *For two arbitrary PFSA P_i and G_i over the same alphabet*,

$$[\![G_{i'}]\!]_{P_i} = [\![G_{i'} \vec{\otimes} P_i]\!]_{P_i}.$$

Proof: Let $P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi})$ and $G_{i'} = (Q', \Sigma, \delta', q'_{i'}, \tilde{\pi}')$. It follows that $G_{i'} \vec{\otimes} P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi}^{\vec{\otimes}})$, where $\tilde{\pi}^{\vec{\otimes}}$ is as computed in Definition 19. Using the same notation as in Definition 19, we have $\forall \sigma \in \Sigma$,

$$\begin{aligned} &\sum_{x:\delta^*(q_i, x) = q_j} \mathbb{H}(G_{i'})(x\sigma) \\ &= \sum_{q'_{k'} \in Q'} \vartheta(q'_{k'}, q_j) \tilde{\pi}'(q'_{k'}, \sigma) \\ &= \sum_{q'_{k'} \in Q'} \vartheta(q'_{k'}, q_j) \left\{ \frac{\sum_{q'_{k'} \in Q'} \vartheta(q'_{k'}, q_j) \tilde{\pi}'(q'_{k'}, \sigma)}{\sum_{q'_{k'} \in Q'} \vartheta(q'_{k'}, q_j)} \right\} \\ &= \left\{ \sum_{q'_{k'} \in Q'} \vartheta(q'_{k'}, q_j) \right\} \tilde{\pi}^{\vec{\otimes}}(q_j, \sigma). \end{aligned} \quad (53)$$

Since $[\![G_{i'}]\!]_{P_i}\big|_j = \sum_{x:\delta^*(q_i, x) = q_j} \mathbb{H}(G_{i'})(x) = \sum_{q'_{k'} \in Q'} \times \vartheta(q'_{k'}, q_j)$, it follows that $\forall \sigma \in \Sigma$,

$$\begin{aligned} &\frac{\sum_{x:\delta^*(q_i, x) = q_j} \mathbb{H}(G_{i'})(x\sigma)}{[\![G_{i'}]\!] P_i \big|_j} \\ &= \tilde{\pi}^{\vec{\otimes}}(q_j, \sigma) \\ &\Rightarrow \sum_{\sigma:\delta(q_j, \sigma) = q\ell} \frac{\sum_{x:\delta^*(q_i, x) = q_j} \mathbb{H}(G_{i'})(x\sigma)}{[\![G_{i'}]\!]_{P_i} \big|_j} \\ &= \sum_{\sigma:\delta(q_j, \sigma) = q\ell} \tilde{\pi}^{\vec{\otimes}}(q_j, \sigma) \\ &\Rightarrow \frac{1}{[\![G_{i'}]\!]_{P_i} \big|_j} \mathbb{H}(G_{i'})(x\Sigma_{j\ell}) = \pi^{\vec{\otimes}}(q_j, q_\ell), \end{aligned} \quad (54)$$

where $\Sigma_{j\ell} \subseteqq \Sigma$ such that $\sigma \in \Sigma_{j\ell} \Rightarrow \delta(q_j, \sigma) = q_\ell$ and $\pi^{\vec{\otimes}}(q_j, q_\ell)$ is the $j\ell$th element of the stochastic state transition matrix $\Pi^{\vec{\otimes}}$ corresponding to the PFSA $G_{i'}\vec{\otimes}P_i$. It follows from (54), that

$$\sum_{q_j \in Q} \mathbb{H}(G_{i'})(x\Sigma_{j\ell}) = \sum_{q_j \in Q} [\![G_{i'}]\!]_{P_i}\big|_j \pi^{\vec{\otimes}}(q_j, q_\ell)$$

$$\Rightarrow [\![G_{i'}]\!]_{P_i}|_\ell = \sum_{q_j \in Q} [\![G_{i'}]\!]_{P_i}\big|_j \pi^{\vec{\otimes}}(q_j, q_\ell). \tag{55}$$

It follows that $[\![G_{i'}]\!]_{P_i}$ satisfies the vector equation

$$[\![G_{i'}]\!]_{P_i} = [\![G_{i'}]\!]_{P_i}\Pi^{\vec{\otimes}}. \tag{56}$$

We note that $[\![G_{i'}\vec{\otimes}P_i]\!]_{P_i}$ is the stable probability distribution of the PFSA $G_{i'}\vec{\otimes}P_i$ and hence, we have

$$[\![G_{i'}\vec{\otimes}P_i]\!]_{P_i} = [\![G_{i'}\vec{\otimes}P_i]\!]_{P_i}\Pi^{\vec{\otimes}}. \tag{57}$$

In general, a stochastic matrix may have more than one eigenvector corresponding to unity eigenvalue (Bapat and Raghavan 1997). However, as per our definition of PFSA (see Definition 2), the initial state is explicitly specified. It follows that the right hand side of (53) assumes that all strings begin from the same state $q_i \in Q$. Hence it follows:

$$[\![G_{i'}]\!]_{P_i} = [\![G_{i'}\vec{\otimes}P_i]\!]_{P_i}. \tag{58}$$

This completes the proof. □

5.1 *Physical significance of projected distribution invariance*

Given a symbolic language theoretic PFSA model for a physical system of interest, one is often concerned with only certain class of possible future evolutions. For example, in the paradigm of deterministic finite state automata (DFSA) (Ramadge and Wonham 1987), the control requirements are expressed in the form of a specification language or a specification automaton. In that setting, it is critical to determine which state of the specification automaton the system is currently visiting. In contrast, for a PFSA, the issue is the probability of certain class of future evolutions. For example, given a large order model of a physical system, it might be necessary to work with a much smaller order PFSA, that has the same long-term behaviour with respect to a specified set of event strings. Although projective composition may incur a representation error in general, the long-term distribution over the states of the projected model is preserved as shown in Theorem 14. The idea is further clarified in the commutative diagram of figure 9.

Probabilistic synchronous composition is an exact representation with no loss of statistical information; but the model order increases due to the product automaton construction. On the other hand, the projective composition has the same number of states as the second argument in $(\bullet)\vec{\otimes}(\bullet)$. Both representations have exactly the same projected distribution with respect to a fixed second argument, thus making $\vec{\otimes}$ an extremely useful tool for model order reduction. Algorithm 3 computes the projected composition of two arbitrary PFSA.

Algorithm 3: Computation of Projected Composition

input: $P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi}), G_{i'} = (Q', \Sigma, \delta', q_{i'}, \tilde{\pi}')$
output: $P_i\vec{\otimes}G_{i'}$

```
begin
  Compute P_i ⊗ G_i' = (Q × Q', Σ, δ^⊗, (q_i, q'_i'), π̃^⊗);
  /* See Definition 4.5 */
  Compute ℘; /* State prob. for P_i ⊗ G_i' Def. 2.3 */
  Set up matrix T s.t. T_jk = ℘((q_j, q'_k));
  Compute π̃^⊗→ = T π̃;
  return P_i ⊗→ G_i' = (Q', Σ, δ', q_i', π̃^⊗→)
end
```

5.2 *Incurred error in projective composition*

Given any two PFSA Pi and $G_{i'}$, the incurred error in projective composition operation $P\vec{\otimes}G_{i'}$ is quantified in the pseudo-metric defined in §4 as follows:

$$\nu_{\alpha,s}(P_i, P_i\vec{\otimes}G_{i'}). \tag{59}$$

Next we establish a sufficient condition for guaranteeing zero incurred error in projective composition.

Theorem 16: *For arbitrary PFSA* $P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi})$ *and* $G_{i'} = (Q', \Sigma, \delta', q'_{i'}, \tilde{\pi}')$ *with corresponding probabilistic Nerode equivalence relations* $\mathcal{N}$ *and* $\mathcal{N}'$, *we have*

$$\mathcal{N} \leqq \mathcal{N}' \Rightarrow \nu_{\alpha,s}(G_{i'}, G_{i'}\vec{\otimes}P_i) = 0.$$

Proof: $\mathcal{N} \leqq \mathcal{N}'$ implies that there exists a possibly non-injective map f: $Q \to Q$ such that

$$\forall x \in \Sigma^\star, \delta^\star(q_i, x) = q_j \in Q \Rightarrow \delta^\star(q'_{i'}, x) = f(q_j) \in Q'.$$

It then follows from Definition 19 that

$$\vartheta(q'_{k'}, q_j) = 0 \quad \text{if } f(q_j) \neq q'_{k'}.$$

Denoting $G_{i'} \otimes P_i = (Q \times Q', \Sigma, \delta^{\otimes}, (q_i, q'_{i'}), \tilde{\pi}^{\otimes})$ and $G_{i'} \overrightarrow{\otimes} P_i = (Q, \Sigma, \delta, q_i, \tilde{\pi}^{\overrightarrow{\otimes}})$, we have fron Definition 19 that

$$\tilde{\pi}^{\overrightarrow{\otimes}}(q_j, \sigma) = \frac{\sum_{q'_{k'} \in Q'} \vartheta(q'_{k'}, q_j)\tilde{\pi}^{\otimes}((q'_{k'}, q_j), \sigma)}{\sum_{q'_{k'} \in Q'} \vartheta(q'_{k'}, q_j)}$$
$$= \tilde{\pi}^{\otimes}((f(q_j), q_j), \sigma) = \tilde{\pi}'(f(q_j), \sigma),$$

where the last step follows from Definition 18. The proof is completed by noting

$$\forall x \in \Sigma^*, \quad \mathbb{H}(G_{i'})(x) = \tilde{\pi}'(q'_{i'}, x) = \tilde{\pi}^{\otimes}((f(q_i), q_i), x)$$
$$= \tilde{\pi}^{\overrightarrow{\otimes}}(q_i, x) = \mathbb{H}(G_{i'} \overrightarrow{\otimes} P_i)(x).$$

□

Example 2: The results of § 5 are illustrated considering the PFSA models described in Example 1. Given the PFSA models $P^1_{q1} = (\{q_1, q_2, q_3\}, \Sigma, \delta^1, q_1, \tilde{\pi}^1)$ and $P^2_{qA} = (\{q_A, q_B\}, \Sigma, \delta^2, q_A, \tilde{\pi}^2)$ (see (40) and (41)), we compute the projected compositions $P^1_{q1} \overrightarrow{\otimes} P^2_{qA} = (\{q_A, q_B\}, \Sigma, \delta^2, q_A, \tilde{\pi}^{12})$ and $P^2_{qA} \overrightarrow{\otimes} P^1_{q1} = (\{q_1, q_2, q_3\}, \Sigma, \delta^1, q_1, \tilde{\pi}^{21})$. The synchronous compositions $P^1_{q1} \otimes P^2_{qA}$ and $P^2_{qA} \otimes P^1_{q1}$ were computed in Example 1 and are shown in figure 8. Denoting the associated stochastic transition matrices for $P^1_{q1} \otimes P^2_{qA}$ and $P^2_{qA} \otimes P^1_{q1}$ as Π^{12} and Π^{21} respectively, we note

$$\Pi^{12} = \begin{bmatrix} 0.2000.8 \\ 00.3.700 \\ .4000.60 \\ 0.2000.8 \\ 00.3.700 \\ .4000.60 \end{bmatrix}, \quad \Pi^{21} = \begin{bmatrix} 0.9000.1 \\ 00.9.100 \\ .9000.10 \\ 0.7000.3 \\ 00.7.300 \\ .7000.30 \end{bmatrix} \begin{matrix} \cdots (q_1, q_A) \\ \cdots (q_2, q_A) \\ \cdots (q_3, q_A) \\ \cdots (q_1, q_B) \\ \cdots (q_2, q_B) \\ \cdots (q_3, q_B). \end{matrix}$$

The stable probability distributions $\wp^{12}$ and $\wp^{21}$ are computed to be

$$\wp^{12} = [0.1458 \quad 0.0695 \quad 0.0864 \quad 0.2017 \quad 0.2186 \quad 0.2780] \tag{60a}$$

$$\wp^{21} = [0.2917 \quad 0.2917 \quad 0.2917 \quad 0.0417 \quad 0.0417 \quad 0.0417]. \tag{60b}$$

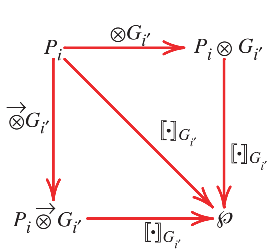

Figure 9. Commutative diagram relating probabilistic composition, projective composition and the original projected distribution.

Using Algorithm 3, we compute the event generating functions $\tilde{\Pi}^{12}$ and $\tilde{\Pi}^{21}$ as

$$\tilde{\Pi}^{12} = \begin{bmatrix} 0.7197 & 0.2803 \\ 0.6891 & 0.3109 \end{bmatrix}, \quad \tilde{\Pi}^{21} = \begin{bmatrix} 0.1250 & 0.8750 \\ 0.1250 & 0.8750 \\ 0.1250 & 0.8750 \end{bmatrix}. \tag{61}$$

We note that the stable distributions for $P^1_{q1} \overrightarrow{\otimes} P^2_{qA}$ and $P^2_{qA} \overrightarrow{\otimes} P^1_{q1}$ are given by

$$\overrightarrow{\wp^{12}} = [0.3017 \quad 0.6983],$$
$$\overrightarrow{\wp^{21}} = [0.3333 \quad 0.3333 \quad 0.3333]. \tag{62}$$

The operations are illustrated in figures 10 and 11 and invariance of the projected distribution is checked as follows:

$$\wp^{12}(1) + \wp^{12}(2) + \wp^{12}(3) = 0.3017 = \overrightarrow{\wp^{12}}(1) \tag{63a}$$

$$\wp^{12}(4) + \wp^{12}(5) + \wp^{12}(6) = 0.6983 = \overrightarrow{\wp^{12}}(2) \tag{63b}$$

$$\wp^{21}(2) + \wp^{21}(5) = 0.333 = \overrightarrow{\wp^{21}}(2) \tag{63c}$$

$$\wp^{21}(3) + \wp^{21}(6) = 0.333 = \overrightarrow{\wp^{21}}(3). \tag{63d}$$

6. An engineering application of pattern recognition

Projective composition is applied to a symbolic pattern identification problem. Continuous-valued data from a laser ranging array in a sensor fusion test bed are fed to a symbolic model reconstruction algorithm (CSSR)

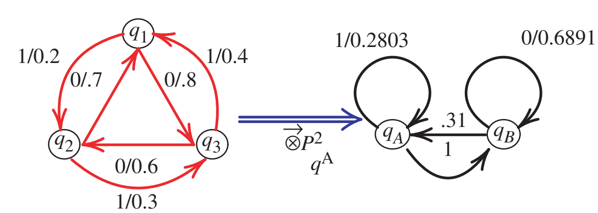

Figure 10. P^1_{q1} projectively composed with P^2_{qA}.

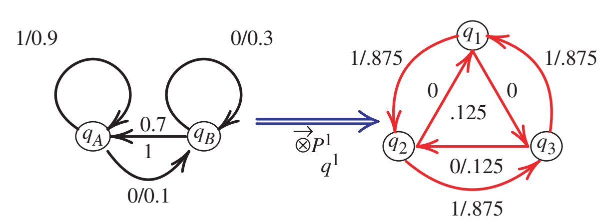

Figure 11. P^2_{qA} projectively composed with P^1_{q1}.

(Shalizi and Shalizi 2004) to yield probabilistic finite state models over a four-letter alphabet. A maximum entropy partitioning scheme (Rajagopalan and Ray 2006) is employed to create the symbolic alphabet on the continuous time series. Figure 12 depicts the results from four different experimental runs. Two of those runs in the top two rows of figure 12 correspond to a human subject moving in the sensor field; the other two runs in the bottom two rows correspond to a robot representing an unmanned ground vehicle (UGV). The symbolic reconstruction algorithm yields PFSA having disparate number of states in each of the above four cases (i.e., two each for the human subject and the robot), with their graph structures being significantly different. The resulting patterns (i.e., state probability vectors) for these PFSA models in each of the four cases are shown on the left side of figure 12. The models are then projectively composed with a 64 state D-Markov machine (Ray 2004) having alphabet size = 4 and depth = 3. The resulting pattern vectors are shown on the right hand column of figure 12. The four rows in figure 12 demonstrate the applicability of projective composition to statistical pattern classification; the state probability vectors of projected models unambiguously identify the respective patterns of a human subject and an UGV.

7. Summary, conclusions and future work

This paper presents a rigorous measure-theoretic approach to probabilistic finite state machines. Key concepts from classical language theory such as the Nerode equivalence relation is generalized to the probabilistic paradigm and the existence and uniqueness of minimal representations for PFSA is established. Two binary operations, namely, probabilistic synchronous composition and projective composition of PFSA are introduced and their properties are investigated. Numerical examples have been provided for clarity of exposition. The applicability of the defined binary operators has been demonstrated on experimental data from a laboratory test bed in a pattern identification and classification problem. This paper lays the framework for three major directions for future research and the associated applications.

- **Probabilistic non-regular languages:** Since projective composition can be used to obtain smaller order models with quantifiable error, the possibility of projectively composing infinite state probabilistic models with finite state machines must be investigated. The extension of the theory developed in this paper to non-regular probabilistic languages would prove invaluable in handling strictly non-Markovian

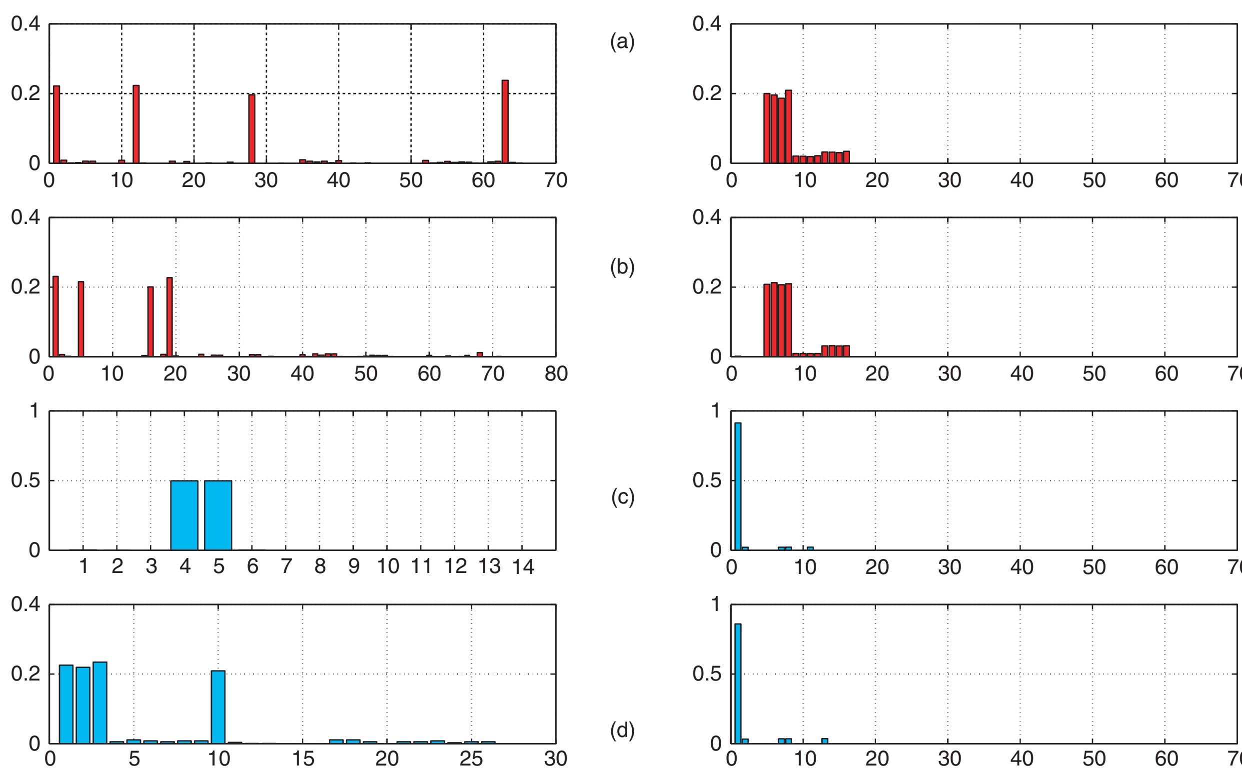

Figure 12. Experimental validation of projective composition in pattern recognition: (a) and (b) correspond to ranging data for a human subject in sensor field; (c) and (d) correspond to an UGV.

models in the symbolic paradigm, especially physical processes that fail to have the semi-Martingle property, e.g., fractional Brownian motion (Decreusefond and Ustunel 1999). Future work will investigate language-theoretic non-regularity as the symbolic analogue to chaotic behavior in the continuous domain.

- **Optimal control:** The reported measure-theoretic approach to optimal supervisor design in PFSA models will be extended in the light of the developments reported in this paper to situations where the control specification is given as weights on the states of DFSA models disparate from the plant under consideration. Such a generalization would allow the fusion of Ramadge and Wonham's constraint based supervision approach (Ramadge and Wonham 1987) with the measure-theoretic approach reported in Chattopadhyay (2006) and Chattopadhyay and Ray (2007). This new control synthesis tool would prove invaluable in the design of event driven controllers in probabilistic robotics.
- **Pattern identification:** Preliminary application in pattern classification has already been demonstrated in §6. Future research will formalize the approach and investigate methodologies for optimally choosing the plant model on which to project the constructed PFSA to yield maximum algorithmic performance. Future investigations will explore applicability of the structural transformations developed in this paper for the fusion, refinement and computation of bounded order symbolic models of observed system behavior in complex dynamical systems.

Acknowledgements

The authors would like to thank Dr. Eric Keller for his valuable contribution in obtaining the experimental results.

This work has been supported in part by the U.S. Army Research Office under Grant Nos. W911NF-06-1-0469 and W911NF-07-1-0376.

References

R. Bapat and T. Raghavan, *Nonnegative Matrices and Applications*, Cambridge University Press, 1997.

I. Chattopadhyay, "Quantitative control of probabilistic discrete event systems," PhD Dissertation, Dept. of Mech. Engg. Pennsylvania State University, http://etda.libraries.psu.edu/theses/approved/WorldWideIndex/ETD-1443, 2006.

I. Chattopadhyay and A. Ray, "Renormalized measure of regular languages", *Int. J. Contr.*, 79, pp. 1107–1117, 2006.

I. Chattopadhyay and A. Ray, "Language-measure-theoretic optimal control of probabilistic finite-state systems", *Int. J. Contr.*, 80, pp. 1271–1290, 2007.

L. Decreusefond and A. Ustunel, "Stochastic analysis of the fractional brownian motion", *Potential Analysis*, 10, pp. 177–214, 1999.

V. Garg, "Probabilistic languages for modeling of DEDs", *Proceedings of* 1992 *IEEE Conference on Information and Sciences*, Princeton, NJ, pp. 198–203, March 1992a.

V. Garg, "An algebraic approach to modeling probabilistic discrete event systems", *Proceedings of* 1992 *IEEE Conference on Decision and Control*, Tucson, AZ, pp. 2348–2353, December 1992b.

W.J. Harrod and R.J. Plemmons, "Comparison of some direct methods for computing the stationary distributions of markov chains,", *SIAM J. Sci. Statist. Comput.*, 5, pp. 453–469, 1984.

J.E. Hopcroft, R. Motwani and J.D. Ullman, *Introduction to Automata Theory, Languages, and Computation*, 2nd ed., Boston, MA, USA: Addison-Wesley, 2001, pp. 45–138.

J.G. Kemeny and J.L. Snell, *Finite Markov Chains*, 2nd ed., New York: Springer, 1960.

R. Kumar and V. Garg, "Control of stochastic discrete event systems modeled by probabilistic languages", *IEEE Trans. Autom. Contr.*, 46, pp. 593–606, 2001.

M. Lawford and W. Wonham, "Supervisory control of probabilistic discrete event systems", *Proceedings of 36th Midwest Symposium on Circuits and Systems*, pp. 327–331, 1993.

V. Rajagopalan and A. Ray, "Symbolic time series analysis via wavelet-based partitioning", *Signal Process.*, 86, pp. 3309–3320, 2006.

P.J. Ramadge and W.M. Wonham, "Supervisory control of a class of discrete event processes", *SIAM J. Contr. Optimiz.*, 25, pp. 206–230, 1987.

A. Ray, "Symbolic dynamic analysis of complex systems for anomaly detection", *Signal Process*, 84, pp. 1115–1130, 2004.

A. Ray, "Signed real measure of regular languages for discrete-event supervisory control", *Int. J. Contr.*, 78, pp. 949–967, 2005.

W. Rudin, *Real and Complex Analysis*, 3rd ed., New York: McGraw Hill, 1988.

C.R. Shalizi and K.L. 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, Virginia, United States, AUAI Press, pp. 504–511, 2004.

W. Stewart, *Computational Probability: Numerical Methods for Computing Stationary Distribution of Finite Irreducible Markov Chains*, New York: Springer, 1999.