Math. Control Signals Syst. (2012) 23:281–310
DOI 10.1007/s00498-012-0075-1

ORIGINAL ARTICLE

An inner product space on irreducible and synchronizable probabilistic finite state automata

Patrick Adenis · Yicheng Wen · Asok Ray

Received: 11 January 2011 / Accepted: 7 January 2012 / Published online: 19 January 2012

Abstract Probabilistic finite state automata (PFSA) have found their applications in diverse systems. This paper presents the construction of an inner-product space structure on a class of PFSA over the real field via an algebraic approach. The vector space is constructed in a stationary setting, which eliminates the need for an initial state in the specification of PFSA. This algebraic model formulation avoids any reference to the related notion of probability measures induced by a PFSA. A formal language-theoretic and symbolic modeling approach is adopted. Specifically, semantic models are constructed in the symbolic domain in an algebraic setting. Applicability of the theoretical formulation has been demonstrated on experimental data for robot motion recognition in a laboratory environment.

Keywords Symbolic dynamics · Probabilistic finite state automata · Vector space

Mathematics Subject Classification (2000) Primary 15A03 · 93A30; Secondary 60J99 · 93E12

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

P. Adenis · Y. Wen · A. Ray (✉)
The Pennsylvania State University, University Park, PA 16802, USA
e-mail: axr2@psu.edu

P. Adenis
e-mail: pxa154@psu.edu

Y. Wen
e-mail: yxw167@psu.edu

Nomenclature

$\mathcal{A}$	is the set of all irreducible FSA [Definition 2.4];
$\bar{\mathcal{A}}$	is the set of all FSA [Definition 2.2];
$\mathcal{A}_s$	is the set of all irreducible and synchronizable FSA [Definition 3.3];
$\mathcal{B}^+$	is the set of all PFSA such that the probability map $\tilde{\pi}$ has only strictly positive entries and the underlying FSA is irreducible [Definition 4.1];
$\mathcal{B}_s^+$	is the subset of synchronizable PFSA in $\mathcal{B}^+$ [Definition 4.1];
$\mathcal{C}^+$	is the quotient set $\mathcal{B}^+/\equiv$ of all PFSA in $\mathcal{B}^+$ [Definition 4.7];
$\mathcal{C}_s^+$	is the quotient set $\mathcal{B}_s^+/\equiv$ of all PFSA in $\mathcal{B}_s^+$ [Definition 4.7];
$\lfloor G \rfloor_K$	is the lift of an FSA G relatively to a PFSA K [Definition 4.8];
$\bar{K}$	is the underlying FSA of the PFSA K [Definition 2.3];
$\widehat{K}$	is the minimal representation of the underlying FSA of the PFSA K [Theorem 4.5];
$\lceil K \rceil$	is the equivalence class of $K \in \mathcal{B}^+$ under the relation $\equiv$ [Definition 4.7];
$\wp$	is the stationary-probability distribution of the states [Sect. 5.3];
Q	is the set of states of the FSA or PFSA [Definition 2.2];
T	is the relabeling map between two FSA [Definition 3.1];
δ	is the state transition map for FSA [Definition 2.2];
δ^*	is the extended state transition map of FSA [Sect. 2];
μ	is the merging map for FSA [Definition 3.2];
π	is the state-to-state transition probability map of PFSA [Sect. 5.3];
$\tilde{\pi}$	is the probability map of a PFSA [Definition 2.3];
$\tilde{\pi}^*$	is the extended probability map of a PFSA [Sect. 2];
Σ	is a finite alphabet of cardinality $\lvert\Sigma\rvert$ [Sect. 2];
Σ^*	is the collection of all *finite-length* words made from Σ [Sect. 2];
ϖ	weight for the inner product $\langle\cdot,\cdot\rangle$ [Subsect. 5.3];
$\mathfrak{S}$	is the state relabeling equivalence for FSA [Definition 3.1];
$\equiv$	is the algebraic equivalence over PFSA [Definition 4.6];
$\lhd$	is the state splitting operation for FSA [Definition 3.2];
$\rhd$	is the state merging operation for FSA [Definition 3.2];
$\asymp$	is the state relabeling operation for PFSA [Definition 4.2];
$\preceq$	is the state merging operation for PFSA [Definition 4.2];
$\vee$	is the join of two FSA [Definition 2.1] [Theorem B.1];
$\curlyvee$	is the join composition of two PFSA [Definition 2.1] [Definition 4.9];
$\oplus$	is the vector addition operation on the space $\mathcal{B}_s^+$ [Definition 5.2];
$\odot$	is the scalar multiplication operation on the space $\mathcal{B}_s^+$ [Definition 5.7];
$+$	is the vector addition operation on the space $\mathcal{C}_s^+$ [Definition 5.4];
$\cdot$	is the scalar multiplication operation on the space $\mathcal{C}_s^+$ [Definition 5.9];
$\langle\langle\bullet,\bullet\rangle\rangle$	is an inner product on the space $\mathcal{B}_s^+$ [Definition 5.12];
$\langle\bullet,\bullet\rangle$	is an inner product on the space $\mathcal{C}_s^+$ [Definition 5.14].

1 Introduction

Symbolic-domain techniques have been developed for probabilistic representation of interacting dynamical systems to compensate for certain inadequacies of classical time-domain and frequency-domain system analysis [1]. The work reported in this paper addresses a formal language-theoretic and symbolic modeling approach instead of the classical continuous domain. Specifically, semantic models are constructed in the symbolic domain, which has wide applications (e.g., pattern classification [2], anomaly detection [3,4], and information fusion in sensor networks [5]) in diverse fields. The basic idea here is that the observed sequence of continuously varying data from the physical system is converted into a symbol sequence over a finite alphabet through a partitioning technique [6]. Then, a finite-state language model is extracted from the symbol sequence to capture the underlying semantics of the process.

Many finite state machine models have been reported in the literature [7,8], such as probabilistic finite state automata (PFSA), hidden Markov models (HMM) [9], stochastic regular grammars [10], Markov chains [11], just to name a few. The rationale of having the PFSA structure of a semantic model is that, in general, PFSA are easier to learn in practice, although PFSA may not always be as powerful as other models like HMM [7]. For example, experimental results [12] show that the usage of a PFSA structure could make learning of a pronunciation model for spoken words to be 10–100 times faster than a corresponding HMM, and yet the overall performance of PFSA may be slightly better. Therefore, this paper focuses on using PFSA as semantic models.

To use PFSA as a feature vector (e.g., for pattern recognition), a mathematical structure on the feature space is needed. However, the theory of how to algebraically manipulate two PFSA has not been explored except for a few cases. The notion of vector space construction for finite state automata (FSA) over the finite field $GF(2)$ was reported in [13]. Along this line, Barfoot and D'Eleuterio [14] 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, i.e., if they have the same alphabet and the number of states together with the same transition maps. A structural manipulation of PFSA models of dynamical systems has been addressed in [15], where the ability to project a PFSA model to an arbitrary structure is critical. The major contribution of this paper is formulation of a general structure of inner product spaces for the analysis of symbolic dynamic systems as delineated below:

1. *An algebraic structure* Formulation of a vector space structure over the real field $\mathbb{R}$ based on a class of irreducible (Definition 2.4) and synchronizable (Definition 3.3) PFSA constructed from finite-length symbol sequences. This formulation does not require usage of the related notion of probabilistic measure [16,17].
2. *A combined topological and algebraic structure* Formulation of an inner-product space formalism on the above vector space to enrich the current theory of PFSA by taking into account disparate automaton structures. In general, this formulation allows construction of a normed space formalism on the vector space of PFSA.

The objective here is to lift two PFSA of different structures to a common equivalent structure, which is accomplished by formalizing the intuitive notions of state merging and state relabeling [18–20]. In particular, the algebraic results (that are derived based on the property of a unique relabeling map) could be generalized to non-synchronizable machines.

The paper is organized in seven sections including this introduction section and two supporting appendices. The basic definitions and necessary concepts are presented in Sect. 2. Sections 3 and 4 develop the algebraic tools on FSA and PFSA, respectively. Using these tools, Sect. 5 constructs an inner product space over PFSA and provides physical interpretations of the underlying algebraic operations. Section 6 presents two illustrative examples, including a laboratory experimentation on pattern recognition of robot motion, to show potential applications of the theoretical results derived in the previous sections. Finally, Sect. 7 concludes this paper with a summary and recommendations for future research. Each of Appendix A and Appendix 8 presents the proof of a pertinent theorem on irreducibility of FSA.

2 Preliminaries

Let Σ denote a fixed (nonempty) finite alphabet, and Σ^* be the collection of all finite words consisting of symbols from Σ. (Note: The vector space is built for a given alphabet Σ.) The following standard definitions are recalled.

Definition 2.1 (*Lattice* [21]) A lattice is defined to be a partially ordered set in which every pair of elements, ℓ_1 and ℓ_2, has both a least upper bound that is called their *join* (denoted as $\ell_1 \vee \ell_2$) and a greatest lower bound that is called their *meet* (denoted as $\ell_1 \wedge \ell_2$).

Definition 2.2 (*Finite state automaton* [20]) A (deterministic) finite state automaton (FSA) G is a tuple (Σ, Q, δ) , where:

- Σ is a (nonempty) finite alphabet with cardinality $|\Sigma|$;
- Q is the (nonempty) finite set of states with cardinality $|Q|$;
- $\delta : Q \times \Sigma \to Q$ is the state transition map.

The set of all FSA is denoted as $\bar{\mathcal{A}}$.

It is noted that Definition 2.2 does not make use of an initial state. Indeed the purpose of this algebraic approach is to work in a stationary setting, where no initial state is provided.

Definition 2.3 (*Probabilistic finite state automaton*) A probabilistic finite state automaton (PFSA) K is a pair $(\bar{K}, \tilde{\pi})$, where:

- $\bar{K}$ is an *underlying FSA* of the PFSA K;
- $\tilde{\pi} : Q \times \Sigma \to [0, 1]$ is the *probability map* that satisfies the condition: $\sum_{\sigma \in \Sigma} \tilde{\pi}(q, \sigma) = 1$ for all $q \in Q$.

The so-called extended maps [20] are defined as: $\delta^* : Q \times \Sigma^* \to Q$ and $\tilde{\pi}^* : Q \times \Sigma^* \to [0, 1]$ for all $w \in \Sigma^*$, all $\sigma \in \Sigma$ and all $q \in Q$, by the recursive relations:

$$\delta^*(q, w\sigma) \triangleq \delta\left[\delta^*(q, w), \sigma\right]$$
$$\tilde{\pi}^*(q, w\sigma) \triangleq \tilde{\pi}^*(q, w) \times \tilde{\pi}\left[\delta^*(q, w), \sigma\right]$$

with $\delta^*(q,\sigma) = \delta(q,\sigma)$ and $\tilde{\pi}^*(q,\sigma) = \tilde{\pi}(q,\sigma)$. Here "$w\sigma$" is to denote the concatenation of the word "w" by the symbol "σ". The above equations represent how the automaton responds to occurrence of a certain block of symbols, i.e, a word $w \in \Sigma^*$ of finite length $|w|$.

Remark 2.1 For all $q \in Q$ and all $w_1, w_2 \in \Sigma^*$, the extended maps satisfy the following conditions:

$$\delta^*(q, w_1 w_2) \triangleq \delta^*\left[\delta^*(q, w_1), w_2\right]$$
$$\tilde{\pi}^*(q, w_1 w_2) \triangleq \tilde{\pi}^*(q, w_1) \times \tilde{\pi}^*\left[\delta^*(q, w_1), w_2\right]$$

Definition 2.4 (*Irreducible FSA*) An FSA G is said to be irreducible if, for all $q_1, q_2 \in Q$, there exists a finite word $w_{1,2} \in \Sigma^*$, such that $q_1 = \delta^*(q_2, w_{1,2})$. The set of all irreducible FSA is denoted as $\mathcal{A} \subset \bar{\mathcal{A}}$.

Definition 2.4 implies that any state can be reached from another state (including the originating state) in a finite number of transitions represented by a word $w_{1,2} \in \Sigma^*$. Graphically, an irreducible FSA is connected. Theorem A.1 in Appendix A presents the partitioning characteristics of irreducible FSA.

Irreducible PFSA are widely used as language models in symbolic systems [19]. In many applications such as pattern classification and anomaly detection, the transitions between the states capture the dynamical evolution of the quasi-stationary system [3], where the initial condition is not important.

Given a finite-length symbol sequence $\mathbb{S}$ over a (finite) alphabet Σ, there exist several PFSA construction algorithms to discover the underlying irreducible PFSA model G of $\mathbb{S}$, such as causal-state splitting reconstruction (CSSR) [22], D-Markov [3,6], and compression via recursive identification of self-similar semantics (CRISSiS [23]). All these algorithms start with identifying the structure of $G \triangleq (Q, \Sigma, \delta)$. Then, a $|Q| \times |\Sigma|$ count matrix C is initialized to the matrix, each of whose elements is equal to 1. Let N_{ij} denote the number of times that a symbol σ_j is generated from the state q_i upon observing the sequence $\mathbb{S}$. The estimated probability map for the PFSA G is computed as

$$\tilde{\pi}(q_i, \sigma_j) \approx \frac{C_{ij}}{\sum_j C_{ij}} = \frac{1 + N_{ij}}{|\Sigma| + N_{ij}} \tag{2.1}$$

The rationale for initializing each element of C to 1 is that if nothing is observed, then there should be no preference to any particular symbol and it is logical to have $\tilde{\pi}(q,\sigma) = \frac{1}{|\Sigma|}\ \forall q \in Q\ \forall \sigma \in \Sigma$, i.e., the uniform distribution.

The above procedure guarantees that the PFSA, constructed from a finite-length symbol sequence, must have a strictly positive probability map. Note that, under this condition, the state transition map δ in Definition 2.2 is a total function and a PFSA is irreducible if and only if the underlying FSA is.

Definition 2.5 (*Synchronous composition* [15]) Let $G_1 = (\Sigma, Q_1, \delta_1)$ and $G_2 = (\Sigma, Q_2, \delta_2)$ be two FSA. Then, synchronous composition of G_1 and G_2 is defined as $G_1 \otimes G_2 = (\Sigma, Q_1 \times Q_2, \delta)$ such that $\forall (q_1, q_2) \in Q_1 \times Q_2$ and $\forall \sigma \in \Sigma$,

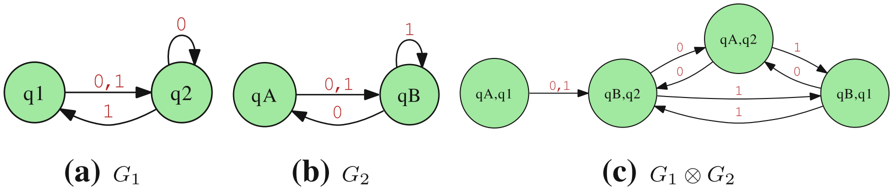


Fig. 1 Synchronous composition: an example

$$\delta\left((q_1, q_2), \sigma\right) \triangleq \left(\delta_1(q_1, \sigma), \delta_2(q_2, \sigma)\right)$$

It is noted that synchronous composition of two irreducible FSA may not always be irreducible. A counter example is shown in Fig. 1, where G_1 and G_2 are both irreducible but their synchronous composition $G_1 \otimes G_2$ is not.

3 Algebraic operations on finite state automata (FSA)

This section focuses on FSA. Algebraic operations are derived on this structure by introducing a partial order over FSA and the notion of synchronizability.

3.1 The *Join* of two FSA

This subsection extends the concept of synchronous composition of two irreducible FSA such that this composition retains the irreducibility. It is achieved by formalizing an equivalence relation and a partial order on $\mathcal{A}$ and by showing the existence of a *join* (i.e., a least upper bound, relatively to this partial order) (see Definition 2.1). In this regard, a few standard definitions of FSA (e.g., state relabeling and state splitting) [18–20] are recalled for completeness of the paper.

Definition 3.1 (*State relabeling equivalence* $\mathfrak{S}$ and relabeling map T in FSA) Two FSA $G_1 = (\Sigma, Q_1, \delta_1)$ and $G_2 = (\Sigma, Q_2, \delta_2)$, both belonging to $\bar{\mathcal{A}}$, are said to be equivalent up to state relabeling, denoted as $G_1 \mathfrak{S} G_2$, if there exists a bijective map $T : Q_2 \to Q_1$, called the *relabeling map* such that $\forall q \in Q_2$ and $\forall \sigma \in \Sigma$,

$$T\left[\delta_2(q, \sigma)\right] = \delta_1\left[T(q), \sigma\right] \tag{3.1}$$

Equation 3.1 implies that the FSA structure is retained by the relabeling operation, i.e., it is possible to 'relabel' the states in the transformation from G_1 to G_2 such that the consistency of transitions is retained.

Definition 3.2 (*State splitting* $\lhd$, *state merging* $\rhd$, *and merging map* μ in FSA) Let $G_1, G_2 \in \bar{\mathcal{A}}$. Then, G_2 is said to be a splitting of G_1 denoted as $G_1 \lhd G_2$, or G_1 a merging of G_2, denoted as $G_2 \rhd G_1$ if there exists a surjective map $\mu : Q_2 \to Q_1$, called the *merging map*, such that $\forall q \in Q_2$ and $\forall \sigma \in \Sigma$,

$$\mu\left[\delta_2(q, \sigma)\right] = \delta_1\left[\mu(q), \sigma\right] \tag{3.2}$$

Remark 3.1 State relabeling and state merging are compatible in the following sense: If $F_2 \mathfrak{S} F_1$ and $G_1 \mathfrak{S} G_2$ with $F_1 \lhd G_1$, then it follows that $F_2 \lhd G_1$ and $F_1 \lhd G_2$.

Remark 3.2 State relabeling $\mathfrak{S}$ is an equivalence relation on $\bar{\mathcal{A}}$, on which the equivalence classes are denoted as $[G]$; and both $\lhd$ and $\rhd$ define respective partial orders over the quotient set $\bar{\mathcal{A}}/\mathfrak{S}$. In particular, $G \leq F$ and $F \leq G$ yield $G \mathfrak{S} F$.

Note that the above equivalence relation $\mathfrak{S}$ is also compatible with the condition of being *irreducible*. Therefore, it is possible to consider the quotient subset $\mathcal{A}/\mathfrak{S}$. Theorem B.1 in Appendix 8 states that any two irreducible FSA admit a join (i.e., a least upper bound) with respect to the state splitting operation $\lhd$, which is critical to ensure the closure property of the vector space to be developed. The following fact is a direct consequence of Theorem B.1:

> If $G_1 \lhd F_1$ and $G_2 \lhd F_2$ then $(G_1 \vee G_2) \lhd (F_1 \vee F_2)$ and also $(G_1 \vee G_2) \vee G_3 = G_1 \vee (G_2 \vee G_3)$. This is true for every partially ordered set where any two elements admit a join.

3.2 Synchronizability

This subsection introduces a specific class of FSA, namely the *synchronizable* FSA [24].

Definition 3.3 (*Synchronizable FSA*) An FSA G is said to be *synchronizable* if there exists a word $w \in \Sigma^*$ such that, $\delta^*(q, w)$ is independent of $q \in Q$. In this case, it is simply written as $\delta^*(w)$, where w is the synchronizing word and $\delta^*(w)$ is the associated synchronizing state. In this context, the set of all irreducible and synchronizable FSA is denoted as $\mathcal{A}_s$.

Synchronizability is compatible with state relabeling, which leads to the quotient subset $\mathcal{A}_s/\mathfrak{S}$.

Theorem 3.4 *Let $G \in \mathcal{A}_s$. Then, the only relabeling map of G onto itself is the identity map. Furthermore, if $F \in \mathcal{A}$ is such that $G \mathfrak{S} F$, then the corresponding relabeling map from G to F is unique.*

Proof Let $T : Q_G \to Q_G$ be a relabeling map of G to itself, and let $w \in \Sigma^*$ be the synchronizing word. Then, for all $q \in Q_G$,

$$\begin{aligned} T\left[\delta_G^*(w)\right] &= T\left[\delta_G^*(q, w)\right] \\ &= \delta_G^*\left[T(q), w\right] \\ &= \delta_G^*(w) \end{aligned}$$

Furthermore, since G is irreducible, for all $q \in Q_G$ there exists $x_q \in \Sigma^*$, which satisfies the following condition: $q = \delta_G^*\left[\delta_G^*(w), x_q\right]$. Then,

$$\begin{aligned} T(q) &= \delta_G^*\left[T(\delta_G^*(w)), x_q\right] \\ &= \delta_G^*\left[\delta_G^*(w), x_q\right] \\ &= q \end{aligned}$$

which implies that $T = \mathrm{Id}_{Q_G}$.

Now, if T_1 and T_2 are two relabeling maps from G to F, then the composed map $T_1^{-1} \circ T_2$ is a relabeling map from G onto itself. So, it follows that $T_1^{-1} \circ T_2 = \mathrm{Id}_{Q_G}$, i.e., $T_1 = T_2$. □

Theorem 3.5 *If* $G_1, G_2 \in \mathcal{A}_s$ *then* $G_1 \vee G_2 \in \mathcal{A}_s$.

Proof If $w_1, w_2 \in \Sigma^*$ are respectively the synchronizing words of G_1 and G_2, then the concatenated word $w_1 w_2$ is a synchronizing word of $G_1 \vee G_2$.

This follows from the fact that, for any two states $q_a, q_b \in Q_{G_1 \vee G_2}$, there exists a merging map μ_i such that $\mu_i(\delta^*_{G_1 \vee G_2}(q_a, w_1 w_2)) = \mu_i(\delta^*_{G_1 \vee G_2}(q_b, w_1 w_2))$, and if $\delta^*_{G_1 \vee G_2}(q_a, w_1 w_2) \neq \delta^*_{G_1 \vee G_2}(q_b, w_1 w_2)$, those two states could be merged together by virtue of Lemma B.2, which would contradict the minimality condition of the join (see proof of theorem B.1 in Appendix 8). □

Remark 3.3 Synchronizability allows perfect state localization after the occurrence of a (finite-length) synchronizing word. A direct analogy is observability in (continuous-time and discrete-time) dynamical systems, where the state of the system can be determined based on a (finite-length) history of the system outputs. Therefore, the assumption of synchronizability is critical in the model construction of symbolic dynamic systems and their performance analysis; it also justifies the rationale for the FSA being independent of an initial state. A few examples are presented below to elucidate how synchronizable PFSA are constructed from symbol sequences.

1. In the D-Markov algorithm [3], the future evolution of the constructed PFSA solely depends on the most recent history of symbol length D. Hence, the D-Markov machines form a strictly proper subset of the set of synchronizable machines.
2. The CSSR [22] and CRISSiS [23] algorithms are capable of identifying up to synchronizable PFSA.

4 Algebraic operations on probabilistic finite state automata (PFSA)

This section generalizes certain results, which were proven for FSA to PFSA, and also introduces a new notion of algebraic equivalence. In this regard, two PFSA are said to be ‘algebraically equivalent’ if they can be obtained, one from the other, by several operations of merging, splitting and relabeling.

Remark 4.1 Any two algebraically equivalent PFSA generate the same probability measure, i.e., they encode the same symbolic dynamics. However, this equivalence relation is purely algebraic and, unlike [16,17], it does not need a probability-measure-theoretic setting.

4.1 Minimal representation

This subsection elaborates the existing concepts of splitting, relabeling, and minimal representation of states [18–20]. In fact, some readers may recognize the PFSA model as a Mealy machine [25] (or a sequential machine [26]) with Σ as the input alphabet and the probability map as the output. However, the output alphabet of a Mealy machine is a finite set while the probability map of PFSA, each of whose elements is a continuum over (0,1) and the sum of probabilities of all symbols emanating from each state is unity. Consequently, the synchronous composition of PFSA is not the same as the normal composition of Mealy machines. However, the minimal representation of PFSA introduced below has a very similar flavor as that of a Mealy machine, which is necessary for defining the algebraic structure of PFSA.

Definition 4.1 (*Spaces $\mathcal{B}^+$ and $\mathcal{B}_s^+$*) The set of all PFSA, whose probability map $\tilde{\pi}$ is strictly positive and whose underlying FSA is irreducible, is denoted as $\mathcal{B}^+$. The subset of all synchronizable PFSA in $\mathcal{B}^+$ is denoted as $\mathcal{B}_s^+$.

Definition 4.2 (*Equivalence $\mathcal{B}^+$of PFSA up to state relabeling*) Following Definition 3.1, two PFSA $K_1, K_2 \in \mathcal{B}^+$ are said to be equivalent up to state relabeling, denoted as $K_1 \asymp K_2$, if $\bar{K}_1 \mathfrak{S} \bar{K}_2$ and $\tilde{\pi}_2(q, \sigma) = \tilde{\pi}_1\left[T(q), \sigma\right]$. Similarly, following Definition 3.2, K_1 is said to be a merging of K_2, denoted as $K_1 \preceq K_2$, if $\bar{K}_1 \lhd \bar{K}_2$ and $\tilde{\pi}_2(q, \sigma) = \tilde{\pi}_1\left[\mu(q), \sigma\right]$.

The above conditions on the probability map $\tilde{\pi}$ imply that relabeling and merging retain the transition probabilities and introduce partial orderings on the quotient sets $(\mathcal{B}^+/\asymp)$ and $(\mathcal{B}^+/\preceq)$, respectively.

Definition 4.3 (*Structural similarity in PFSA*) Two PFSA K and H, belonging to $\mathcal{B}^+$, are said to have the *same structure* if their underlying FSA are $\mathfrak{S}$-equivalent, i.e., $\bar{K} \mathfrak{S} \bar{H}$.

The implication of Definition 4.3 is that K and H have the same underlying FSA, but with potentially different transition probabilities.

Lemma 4.4 *Let $H_1, H_2, K \in \mathcal{B}^+$ such that $H_1 \preceq K$ and $H_2 \preceq K$. Then, there exists $H_3 \in \mathcal{B}^+$ such that $H_3 \preceq H_1$ and $H_3 \preceq H_2$.*

Proof An equivalence relation $\sim$ on Q_K is defined by: $q' \sim q''$ *iff* there exists a finite chain of states $q_1, \ldots, q_{n+1} \in Q_K$ with $q_1 = q', q_{n+1} = q''$ such that: $\forall\, i \in \{1, \ldots, n\}$,

$$\begin{cases} \mu_1(q_i) = \mu_1(q_{i+1}) \\ \qquad\quad \textbf{or} \\ \mu_2(q_i) = \mu_2(q_{i+1}) \end{cases}$$

Let $Q_3 = Q_K/\sim$ be the quotient space, and let $[q]$ be the equivalence class of $q \in Q_K$. The idea is to merge together all the elements being in the same equivalence class.
Claim *One has $\delta_K(q', \sigma) \sim \delta_K(q'', \sigma)$ and $\tilde{\pi}_K(q', \sigma) = \tilde{\pi}_K(q'', \sigma)$ for any $\sigma \in \Sigma$ and any $q \sim q'$. Indeed, if there exist $q_1, q_2 \in Q_K$ and some $i \in \{1, 2\}$ such that $\mu_i(q_1) = \mu_i(q_2)$, then it follows that*

$$\begin{aligned} &\tilde{\pi}_i(\mu_i(q_1), \sigma) = \tilde{\pi}_i(\mu_i(q_2), \sigma) \\ \Rightarrow\ &\tilde{\pi}_K(q_1, \sigma) = \tilde{\pi}_K(q_2, \sigma) \end{aligned}$$

and

$$\begin{aligned} & \delta_i(\mu_i(q_1),\sigma) = \delta_i(\mu_i(q_2),\sigma) \\ \Rightarrow\quad & \mu_i(\delta_K(q_1,\sigma)) = \mu_i(\delta_K(q_2,\sigma)) \\ \Rightarrow\quad & \delta_K(q_1,\sigma) \sim \delta_K(q_2,\sigma) \end{aligned}$$

The claim is then established by induction on the set $\{q_i\}_{i=1,\dots,n+1}$.

To resume the proof of Lemma 4.4, H_3 is defined on Q_3, by setting

$$\begin{aligned} \delta_3([q],\sigma) &\triangleq [\delta_K(q,\sigma)] \\ \tilde{\pi}_3([q],\sigma) &\triangleq \pi_K(q,\sigma) \end{aligned}$$

that are well defined according to the above claim.

Furthermore, for any $q \in Q_K$, for $i \in \{1,2\}$, one has $\mu_i^{-1}(\mu_i(q)) \subseteq [q]$. Therefore, by setting $\tilde{\mu}_i : Q_i \to Q_3$ as $\tilde{\mu}_i(q) \triangleq [\mu_i^{-1}(q)] \forall q \in Q_i$, it follows that $\tilde{\mu}_i$ is surjective because μ_i is surjective, and $\tilde{\mu}_i$ is a merging map.

Indeed, for all $q \in Q_i$ and all $\sigma \in \Sigma$, there exists $q' \in Q_K$ such that $\mu_i(q') = q$. Hence,

$$\begin{aligned} \tilde{\pi}_3(\tilde{\mu}_i(q),\sigma) &= \tilde{\pi}_3\left([q'],\sigma\right) \\ &= \tilde{\pi}_K\left(q',\sigma\right) \\ &= \tilde{\pi}_i(\mu_i(q'),\sigma) \\ &= \tilde{\pi}_i(q,\sigma) \end{aligned}$$

and

$$\begin{aligned} \tilde{\mu}_i\left(\delta_i(q,\sigma)\right) &= \left[\mu_i^{-1}(\delta_i(q,\sigma))\right] \\ &= \left[\delta_K(q',\sigma)\right] \\ &= \delta_3\left([q'],\sigma\right) \\ &= \delta_3\left(\tilde{\mu}_i(q),\sigma\right) \end{aligned}$$

So that $H_3 \preceq H_i$ for $i \in \{1,2\}$. □

Theorem 4.5 (Minimal representation) *Let $K \in \mathcal{B}^+$. There exists a unique (up to state relabeling) irreducible PFSA $\widehat{K} \in \mathcal{B}^+$, called the minimal representation of K such that $\widehat{K} \preceq K$ and $\widehat{K}$ have a minimal number of states. Furthermore, if $H \preceq K$ for any $H \in \mathcal{B}^+$, then $\widehat{K} \asymp \widehat{H}$.*

Proof Let Γ be the set of lower-bounds of K, and $\Gamma' \subset \Gamma$ be the set of lower-bounds having the minimal number of states, say m. Then, for all $H_1, H_2 \in \Gamma'$, applying Lemma 4.4, there exists $H_3 \in \Gamma$ with $|H_3| \leq m$ and $H_3 \preceq H_i$. But the minimality condition imposes to have $|H_3| = m$, that is $H_3 \in \Gamma'$. Thus, $H_1 \asymp H_3 \asymp H_2$. Now, all the lower-bounds of H are lower-bounds of K; therefore, if $H \preceq K$, then one has $\widehat{H} \preceq \widehat{K}$. By applying Lemma 4.4 again to $\widehat{K}$ and $\widehat{H}$, it follows that $\widehat{K} \asymp \widehat{H}$. □

The concept of minimal representation leads to the key notion of *algebraic equivalence* as stated in the following definition.

Definition 4.6 (*Algebraic equivalence of PFSA*) Two PFSA $K, H \in \mathcal{B}^+$ are said to be algebraically equivalent, denoted as $K \equiv H$, if their minimal representations are equal up to state relabeling, i.e., $\widehat{K} \asymp \widehat{H}$.

The above definition implies that two algebraically equivalent PFSA can be obtained from each other by several operations of state splitting and merging.

Definition 4.7 (*Synchronizable minimal representation of PFSA*) Following Definition 4.6, the quotient set $\mathcal{C}^+ \triangleq \mathcal{B}^+ / \equiv$ and the set $\mathcal{C}_s^+$ of PFSA with an irreducible and synchronizable minimal representation is defined as:

$$\mathcal{C}_s^+ \triangleq \mathcal{B}_s^+ / \equiv$$

In other words, the quotient set $\mathcal{C}_s^+$ could be expressed as:

$$\mathcal{C}_s^+ = \left\{ \lceil K \rceil \in \mathcal{C}^+ \mid \overline{\widehat{K}} \in \mathcal{A}_s \right\}$$

where $\lceil K \rceil$ is the equivalence class of $K \in \mathcal{B}^+$ under the relation $\equiv$.

In the sequel, the vector space is built on $\mathcal{C}_s^+$ which is the set of classes of irreducible and synchronizable PFSA.

4.2 Join composition of PFSA

Join composition of PFSA consists of bringing two PFSA to an equivalent common structure, so that the vector addition can be performed. The idea is to compute the join of the underlying FSA of two PFSA, and to label the transitions with the probabilities of one or the other PFSA. This labeling process is what is called the 'lift' of an FSA relative to a PFSA.

Definition 4.8 (*Lift of an FSA relatively to a PFSA*) Let $K \in \mathcal{B}^+$ and $G \in \mathcal{A}$ satisfying with $\bar{K} \lhd G$ (merging mapping μ). The lift of G relatively to K, denoted as $\lfloor G \rfloor_K$, is the PFSA $(G, \tilde{\pi}_G)$, where $\tilde{\pi}_G(q, \sigma) \triangleq \tilde{\pi}_K [\mu(q), \sigma]$.

The lift consists of making an FSA G into a PFSA $\lfloor G \rfloor_K$ by taking the transitions' probabilities of another PFSA K. From the above definition, it follows that $K \preceq \lfloor G \rfloor_K$, and thus $K \equiv \lfloor G \rfloor_K$. More specifically, the lift can be computed for the join of the underlying FSA $\bar{K}_1 \vee \bar{K}_2$ (see Sect. 3.1).

Definition 4.9 (*Join composition*) Let $K, H \in \mathcal{B}^+$. The join composition of K by H, denoted as $K \curlyvee H$, is the lift: $K \curlyvee H \triangleq \lfloor \bar{K} \vee \bar{H} \rfloor_K$.

It follows from the above definition that $K \curlyvee H$ and $H \curlyvee K$ have the same structure although they are not equivalent up to state relabeling; and $K \equiv K \curlyvee H$. Also $\widehat{K} \curlyvee H \preceq K \curlyvee H$ and $K \curlyvee \widehat{H} \preceq K \curlyvee H$.

Remark 4.2 The notions of *join* and *join composition* generalize the idea of synchronous composition for FSA and PFSA, respectively [15]. From the computational perspective, join and join composition are directly obtained from the synchronous composition, where the transient states and the (possibly) superfluous independent cycles are removed; this is what Lemma B.3 in Appendix 8 states.

5 Inner product space structure on the quotient set $\mathcal{C}_s^+$

This section establishes the structure of an inner product space on the quotient set $\mathcal{C}_s^+$ by making use of the algebraic tools developed in the previous sections. Along this line, the following algebraic and topological notions are introduced:

- Vector addition $\oplus$ in Sect. 5.1
- Scalar multiplication $\odot$ in Sect. 5.2
- Inner product in Sect. 5.3

The operations on $\mathcal{C}_s^+$ are established in three steps:

- Defining the operation on PFSA having the same structure in $\mathcal{B}_s^+$
- Extending the above structure to any PFSA using the join composition in Sect. 4.2.
- Showing the operation keeps the algebraic equivalences (compatibility), and extending the definition to the quotient space in $\mathcal{C}_s^+$.

5.1 Abelian group structure

This subsection introduces the notion of vector addition in the space of PFSA.

Definition 5.1 (*Vector addition on the PFSA of similar structure*) Let $K, H \in \mathcal{B}_s^+$ be two synchronizable PFSA having the same FSA structure, i.e., $\bar{K} \mathfrak{S} \bar{H}$. Since they are synchronizable, there exists a unique relabeling map $T : Q_K \to Q_H$ from $\bar{K}$ to $\bar{H}$ (see Sect. 3.2). The addition is defined as $K \oplus H \triangleq (\Sigma, Q_K, \delta_K, \pi_{K\oplus H})$ where

$$\tilde{\pi}_{K\oplus H}(q,\sigma) \triangleq \frac{\tilde{\pi}_K(q,\sigma) \times \tilde{\pi}_H\left[T(q),\sigma\right]}{\sum_{\alpha\in\Sigma} \tilde{\pi}_K(q,\alpha) \times \tilde{\pi}_H\left[T(q),\alpha\right]} \quad \forall q \in Q \ \forall \sigma \in \Sigma \tag{5.1}$$

Remark 5.1 If applied to non-synchronizable machines, Definition 5.1 could have ambiguities. The rationale is that, in general, two equivalent FSA up to state relabeling could admit several corresponding relabeling maps T. For synchronizable machines, existence and uniqueness of such a map T are guaranteed [18] (see Sect. 3.2), and the operation of vector addition relies on the uniqueness of the map T; however, some non-synchronizable machines may also satisfy this condition. For example, Fig. 2 shows two non-synchronizable FSA such that G_1 has a unique relabeling map and G_2 does not due to the perfect symmetry of the graph. Hence, it would be possible to achieve quasi-synchronizability, i.e., to extend this formulation to non-synchronizable machines with the characterization of this unique relabeling map. The underlying

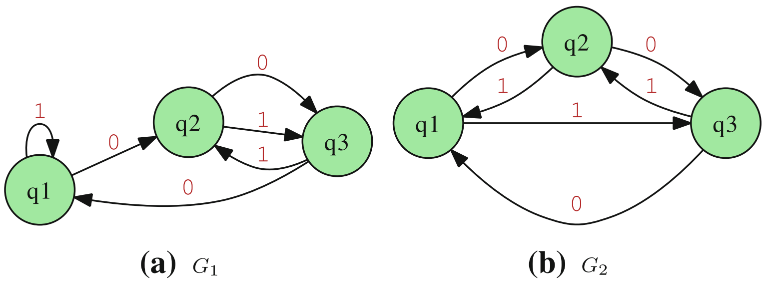


Fig. 2 Examples of non-synchronizable FSA

problem is that the join of two quasi-synchronizable PFSA may not be quasi-synchronizable, itself. Note that such a characterization is difficult to obtain if a measure-theoretic approach [16,17] is adopted rather than the algebraic approach developed in this paper.

Definition 5.2 (*Vector addition on $\mathcal{B}_s^+$*) Let $K, H \in \mathcal{B}_s^+$. Then, the vector addition is defined as:

$$K \oplus H \triangleq (K \curlyvee H) \oplus (H \curlyvee K) \tag{5.2}$$

The above definition is consistent with the foregoing one since the PFSA K and H having the same structure implies that $K \curlyvee H = K$ and $H \curlyvee K = H$.

Proposition 5.3 (Compatibility of $\oplus$ with equivalence classes) *Let $K, H \in \mathcal{B}_s^+$. Then $\widehat{K} \oplus H \preceq K \oplus H$, so that $K \oplus H \equiv \widehat{K} \oplus H$.*

Proof With the same merging map μ, it follows that

$$\widehat{K} \curlyvee H \preceq K \curlyvee H$$
$$H \curlyvee \widehat{K} \preceq H \curlyvee K$$

Since $K \curlyvee H$ and $H \curlyvee K$ have the same structure as well as $\widehat{K} \curlyvee H$ and $H \curlyvee \widehat{K}$, it comes $\mu_1 = \mu_2 = \mu$. It is now claimed that μ is the merging map of $\widehat{K} \oplus H \preceq K \oplus H$. Indeed, μ is surjective and hence it follows that $\forall q \in Q_{K \vee H}\ \forall \sigma \in \Sigma$,

$$\begin{aligned}
\mu\left[\delta_{K \oplus H}(q, \sigma)\right] &= \mu\left[\delta_{K \vee H}(q, \sigma)\right] \\
&= \delta_{\widehat{K} \vee H}(\mu(q), \sigma) \\
&= \delta_{\widehat{K} \oplus H}(\mu(q), \sigma)
\end{aligned}$$

and

$$\tilde{\pi}_{\widehat{K} \oplus H}(\mu(q), \sigma) = \frac{\tilde{\pi}_{\widehat{K} \vee H}(\mu(q), \sigma) \times \tilde{\pi}_{H \vee \widehat{K}}(\mu(q), \sigma)}{\sum_{\alpha \in \Sigma} \tilde{\pi}_{\widehat{K} \vee H}(\mu(q), \alpha) \times \tilde{\pi}_{H \vee \widehat{K}}(\mu(q), \alpha)}$$

$$= \frac{\tilde{\pi}_{K\vee H}(q,\sigma)\tilde{\pi}_{H\vee K}(q,\sigma)}{\sum_{\alpha\in\Sigma}\tilde{\pi}_{K\vee H}(q,\alpha)\tilde{\pi}_{H\vee K}(q,\alpha)}$$
$$= \tilde{\pi}_{K\oplus H}(q,\sigma)$$

□

By symmetry, since $K \oplus H \asymp H \oplus K$ (in this case, the unique relabeling map is T^{-1}), it follows that $K \oplus H \equiv K \oplus \widehat{H}$.

Definition 5.4 (*Vector addition on* $\mathcal{C}_s^+$) Let the binary operator

$$+ : \mathcal{C}_s^+ \times \mathcal{C}_s^+ \to \mathcal{C}_s^+$$

be defined for any $\lceil K \rceil, \lceil H \rceil \in \mathcal{C}_s^+$ by $\lceil K \rceil + \lceil H \rceil \triangleq \lceil \widehat{H} \oplus \widehat{K} \rceil$

This addition operation is well defined because the result does not depend on the specific representation of K and H, as shown in Proposition 5.3.

Remark 5.2 The operation of vector addition of two PFSA is interpreted as follows. Given two PFSA having the same structure that can always be obtained by the join composition (see Sect. 4.2), the vector addition magnifies the probabilities when they are both high, weakens them when they are both very small, and tends to average them in other cases.

Definition 5.5 (*Symbolic white noise*) Let e be the PFSA defined by $Q_e = \{q_e\}$ (only one state), $\delta_e(q_e, \sigma) = q_e$, and $\tilde{\pi}_e(q_e, \sigma) = \frac{1}{|\Sigma|}$. This PFSA is called *symbolic white noise* and the underlying FSA is denoted as $\underline{e}$.

Remark 5.3 The rationale for the PFSA e being called *symbolic white noise* is as follows. The PFSA e encodes a semantic sequence, where any word is completely independent of the past history and all words of the same (finite) length have equal probability of occurrence. Performing the vector addition of this symbolic white noise with another PFSA does not provide any additional information and keeps this PFSA unchanged.

Theorem 5.6 $\left(\mathcal{C}_s^+, +\right)$ *is an Abelian Group with e as the zero element.*

Proof The axioms of an Abelian Group need to be checked.

- *Commutativity* This is obvious according to the definition of the vector addition.
- *Associativity* Let $K_1, K_2, K_3 \in \mathcal{B}_s^+$.
  If K_1, K_2 and K_3 have the same structure[1], then it follows that

$$\tilde{\pi}_{(K_1\oplus K_2)\oplus K_3}(q,\sigma) = \frac{\tilde{\pi}_{K_1\oplus K_2}(q,\sigma)\tilde{\pi}_3(q,\sigma)}{\sum_{\alpha\in\Sigma}\tilde{\pi}_{K_1\oplus K_2}(q,\alpha)\tilde{\pi}_3(q,\alpha)}$$

[1] Since the relabeling map T is unique it is omitted in the following calculations.

$$= \frac{\frac{\tilde{\pi}_1(q,\sigma)\tilde{\pi}_2(q,\sigma)}{\sum_{\beta\in\Sigma}\tilde{\pi}_1(q,\beta)\tilde{\pi}_2(q,\beta)}\tilde{\pi}_3(q,\sigma)}{\sum_{\alpha\in\Sigma}\frac{\tilde{\pi}_1(q,\alpha)\tilde{\pi}_2(q,\alpha)}{\sum_{\beta\in\Sigma}\tilde{\pi}_1(q,\beta)\tilde{\pi}_2(q,\beta)}\tilde{\pi}_3(q,\alpha)}$$
$$= \frac{\tilde{\pi}_1(q,\sigma)\tilde{\pi}_2(q,\sigma)\tilde{\pi}_3(q,\sigma)}{\sum_{\alpha\in\Sigma}\tilde{\pi}_1(q,\alpha)\tilde{\pi}_2(q,\alpha)\tilde{\pi}_3(q,\alpha)}$$

This last expression is symmetrical, and thus $\tilde{\pi}_{(K_1\oplus K_2)\oplus K_3}(q,\sigma) = \tilde{\pi}_{K_1\oplus(K_2\oplus K_3)}(q,\sigma)$.
If K_1, K_2 and K_3 do not have the same structure, then associativity of the join operator $\vee$ over FSA simply needs to be recalled.

$$\bar{K}_1 \vee (\bar{K}_2 \vee \bar{K}_3) = (\bar{K}_1 \vee \bar{K}_2) \vee \bar{K}_3$$

- *Zero element* Given $K \in \mathcal{B}_s^+$, it follows that $\bar{K} \rhd \bar{e}$, and therefore, $\bar{e} \vee \bar{K} = \bar{K}$. This means that $Q_{K\oplus e} = Q_K$ and $\delta_{K\oplus e} = \delta_K$. Furthermore, by definition, $\tilde{\pi}_{K\vee e}(q,\sigma) = \tilde{\pi}_K(q,\sigma)$ and $\tilde{\pi}_{e\vee K}(q,\sigma) = \tilde{\pi}_e(q_e,\sigma) = \frac{1}{|\Sigma|}$, which yields:

$$\tilde{\pi}_{K\vee e}(q,\sigma) = \frac{\tilde{\pi}_K(q,\sigma)\times\frac{1}{|\Sigma|}}{\sum_{\alpha\in\Sigma}\tilde{\pi}_K(q,\alpha)\times\frac{1}{|\Sigma|}} = \tilde{\pi}_K(q,\sigma)$$

- *Inverse* Let $K \in \mathcal{B}_s^+$.
Let $(-K) \triangleq (\Sigma, Q_K, \delta_K, \tilde{\pi}_{-K})$ with

$$\tilde{\pi}_{-K}(q,\sigma) \triangleq \frac{\frac{1}{\tilde{\pi}_K(q,\sigma)}}{\sum_{\alpha\in\Sigma}\frac{1}{\tilde{\pi}_K(q,\alpha)}} \tag{5.3}$$

Then, it follows that $\tilde{\pi}_{K\oplus(-K)}(q,\sigma) = \frac{1}{|\Sigma|}$ for all $q \in Q_K$ and $\sigma \in \Sigma$.
Merging all the states, one finally gets $K \oplus (-K) \equiv e$. □

5.2 Scalar multiplication

This subsection introduces the notion of scalar multiplication in the vector space of PFSA.

Definition 5.7 (*Scalar multiplication on PFSA*) Let the binary operator of scalar multiplication $\odot : \mathbb{R} \times \mathcal{B}_s^+ \to \mathcal{B}_s^+$ be defined for all $\lambda \in \mathbb{R}$ and all $K \in \mathcal{B}_s^+$ by $\lambda \odot K \triangleq (\Sigma, Q_K, \delta_K, \tilde{\pi}_{\lambda\odot K})$ with

$$\tilde{\pi}_{\lambda\odot K}(q,\sigma) \triangleq \frac{(\tilde{\pi}_K(q,\sigma))^\lambda}{\sum_{\alpha\in\Sigma}(\tilde{\pi}_K(q,\alpha))^\lambda}\ \forall q \in Q\ \forall\sigma \in \Sigma \tag{5.4}$$

Proposition 5.8 (Compatibility with equivalence classes) *Let* $\lambda \in \mathbb{R}$ *and* $K \in \mathcal{B}_s^+$. *Then,* $\lambda \odot \widehat{K} \preceq \lambda \odot K$ *so that* $(\lambda \odot K) \equiv (\lambda \odot \widehat{K})$.

Proof The merging map μ associated with $\widehat{K} \preceq K$ is also associated with $(\lambda \odot \widehat{K}) \preceq (\lambda \odot K)$. $\square$

Definition 5.9 (*Scalar multiplication on* $\mathcal{C}_s^+$) Let the binary operator

$$\cdot : \mathbb{R} \times \mathcal{C}_s^+ \to \mathcal{C}_s^+$$

be defined for any $\lambda \in \mathbb{R}$ and any $\lceil K \rceil \in \mathcal{C}_s^+$ by $\lambda \cdot \lceil K \rceil \triangleq \lceil \lambda \odot \widehat{K} \rceil$.

Remark 5.4 Scalar multiplication in the vector space of PFSA retains the structure of the PFSA and simply reshapes its probabilities via the probability map $\tilde{\pi}$. Multiplication by $\lambda \in \mathbb{R}^+$ puts a larger weight to the transition having a higher probability and weakens those having lower probabilities; the effect is opposite if $\lambda < 0$. Therefore, as $\lambda \to +\infty$, the PFSA tends to a delta distribution, where the highest probability in each state tends to 1 while the others tend to 0. Similarly, as $\lambda \to -\infty$, the PFSA also tends to a delta distribution, where the smallest probability in each state tends to 1 while the others tend to 0. Interestingly, in these limits, the PFSA may not be irreducible any more. On the contrary, if $k \to 0$, all the probabilities tend to the uniform distribution, i.e., the PFSA tends to behave as the symbolic white noise. In essence, multiplication of a vector by the scalar 0 yields the zero vector.

Theorem 5.10 (Vector space structure) $\left(\mathcal{C}_s^+, +, \cdot\right)$ *is a vector space over the real field* $\mathbb{R}$.

Proof The missing axioms of the vector space structure need to be checked.

- (*Unitarity*) From the definition, it is clear that $1 \cdot K = K$, for any $K \in \mathcal{B}_s^+$.
- (*Distributivity of* $\cdot$ *on* $+$) Let $K, H \in \mathcal{C}_s^+$ and $\lambda \in \mathbb{R}$.
  If K and H have the same structure, it comes

$$\begin{aligned}
\tilde{\pi}_{\lambda\odot(K\oplus H)}(q,\sigma) &= \frac{(\tilde{\pi}_{K\oplus H}(q,\sigma))^\lambda}{\sum_{\alpha\in\Sigma}(\tilde{\pi}_{K\oplus H}(q,\alpha))^\lambda} \\
&= \frac{\left(\frac{\tilde{\pi}_K(q,\sigma)\times\tilde{\pi}_H(q,\sigma)}{\sum_{\beta\in\Sigma}\tilde{\pi}_K(q,\beta)\times\tilde{\pi}_H(q,\beta)}\right)^\lambda}{\sum_{\alpha\in\Sigma}\left(\frac{\tilde{\pi}_K(q,\alpha)\times\tilde{\pi}_H(q,\alpha)}{\sum_{\beta\in\Sigma}\tilde{\pi}_K(q,\beta)\times\tilde{\pi}_H(q,\beta)}\right)^\lambda} \\
&= \frac{(\tilde{\pi}_K(q,\sigma))^\lambda\times(\tilde{\pi}_H(q,\sigma))^\lambda}{\sum_{\alpha\in\Sigma}(\tilde{\pi}_K(q,\alpha))^\lambda\times(\tilde{\pi}_H(q,\alpha))^\lambda} \\
&= \frac{\left(\frac{\tilde{\pi}_K(q,\sigma)}{\sum_{\beta\in\Sigma}\tilde{\pi}_K(q,\beta)}\right)^\lambda\times\left(\frac{\tilde{\pi}_H(q,\sigma)}{\sum_{\beta\in\Sigma}\tilde{\pi}_H(q,\beta)}\right)^\lambda}{\sum_{\alpha\in\Sigma}\left(\frac{\tilde{\pi}_K(q,\alpha)}{\sum_{\beta\in\Sigma}\tilde{\pi}_K(q,\beta)}\right)^\lambda\times\left(\frac{\tilde{\pi}_H(q,\alpha)}{\sum_{\beta\in\Sigma}\tilde{\pi}_H(q,\beta)}\right)^\lambda} \\
&= \frac{\tilde{\pi}_{\lambda\odot K}(q,\sigma)\times\tilde{\pi}_{\lambda\odot H}(q,\sigma)}{\sum_{\alpha\in\Sigma}\tilde{\pi}_{\lambda\odot K}(q,\alpha)\times\tilde{\pi}_{\lambda\odot H}(q,\alpha)} \\
&= \tilde{\pi}_{\lambda\odot K\oplus\lambda\odot H}(q,\sigma)
\end{aligned}$$

If K and H do not have the same structure, then

$$\begin{aligned}\tilde{\pi}_{\lambda\cdot(K+H)}(q,\sigma) &= \tilde{\pi}_{\lambda\odot(\widehat{K}\oplus\widehat{H})}(q,\sigma)\\ &= \tilde{\pi}_{\widehat{\lambda\cdot K}\oplus\widehat{\lambda\cdot H}}(q,\sigma)\\ &= \tilde{\pi}_{\lambda\cdot K+\lambda\cdot H}(q,\sigma)\end{aligned}$$

- (*Distributivity of + on ·*) Let $K \in \mathcal{B}_s^+$ and $\lambda_1, \lambda_2 \in \mathbb{R}$. Of course, K, $(\lambda_1 \cdot K)$, $(\lambda_2 \cdot K)$ and $(\lambda_1 + \lambda_2) \cdot K$ have the same structure.
  By a very similar calculation, it follows that

$$\tilde{\pi}_{(\lambda_1+\lambda_2)\cdot K}(q,\sigma) = \tilde{\pi}_{\lambda_1\cdot K+\lambda_2\cdot K}(q,\sigma)$$

- (*Compatibility of multiplications*) Let $K \in \mathcal{B}^+$ and $\lambda_1, \lambda_2 \in \mathbb{R}$. It is known that K, $(\lambda_1 \cdot K)$, $(\lambda_2 \cdot (\lambda_1 \cdot K))$ and $(\lambda_1 \times \lambda_2) \cdot K$ have the same structure.
  And again, it is easy to check that:

$$\tilde{\pi}_{(\lambda_1\times\lambda_2)\cdot K}(q,\sigma) = \tilde{\pi}_{\lambda_1\cdot(\lambda_2\cdot K)}(q,\sigma).$$

□

5.3 Inner product

This subsection presents the construction of a general inner-product. The vector addition (see Definitions 5.2 and 5.4) is constructed in three steps. First, the operation of vector addition is defined for PFSA having the same structure; then it is generalized for any two PFSA using the join composition (see Sect. 4.2); and eventually it is generalized to the quotient space showing a compatibility property.

Similar to the usual notion of an inner product in functional analysis [27], it is possible to define a family of inner products having different weights. Here, weight is defined to be a mapping $\varpi_G : Q_G \to (0, +\infty)$ which associates a strictly positive real number to a state of an arbitrary FSA G.

Definition 5.11 (*Compatible weight* ϖ) The family of weight functions $\{\varpi_G : Q_G \to (0, \infty),\ G \in \mathcal{A}_s\}$ is said to be compatible if the following condition holds true:

$$\varpi_F(q) = \sum_{\tilde{q}\in\mu^{-1}(q)} \varpi_G(\tilde{q}) \tag{5.5}$$

whenever $F \lhd G$ with relabeling map μ.

In Definition 5.11, compatibility means that whenever a state is split, the weight of this state is distributed on the split states. Inner products can be defined based on a family of weights $\{\varpi_G\}$.

Definition 5.12 (*Inner product on* $\mathcal{B}_s^+$) Let $K, H \in \mathcal{B}_s^+$ and $\{\varpi_G, G \in \mathcal{A}_s\}$ be a compatible family of weights. If K and H have the same structure, then by synchronizability there exists a unique relabeling map $T : Q_K \to Q_H$ (see Sect. 3.2), and the inner product is defined as:

$$\langle\langle K, H\rangle\rangle \triangleq \sum_{q\in Q_K} \varpi_K(q) \sum_{\alpha,\beta\in\Sigma} \log\left(\frac{\tilde{\pi}_K(q,\alpha)}{\tilde{\pi}_K(q,\beta)}\right) \times \log\left(\frac{\tilde{\pi}_H\,[T(q),\alpha]}{\tilde{\pi}_H\,[T(q),\beta]}\right) \tag{5.6}$$

If K and H do not have the same structure, then the inner product is defined as:

$$\langle\langle K, H\rangle\rangle \triangleq \langle\langle K \curlyvee H, H \curlyvee K\rangle\rangle \tag{5.7}$$

Remark 5.5 There are many ways to define an inner product on a vector space of PFSA. For example, Wen et al. [16,17] have reported the construction of a family of inner products to generate an appropriate metric for the specific problem at hand. However, the procedure outlined in [17] is probability-measure-theoretic and requires dependence on the initial state of the PFSA. A topic of future research is development of a procedure for choice of an inner product that provides an appropriate metric.

The next proposition generalizes Definition 5.12 to the quotient space $\mathcal{C}_s^+$.

Proposition 5.13 *(Compatibility with equivalence classes) For any* $K, H \in \mathcal{B}_s^+$, *one has* $\langle\langle K, H\rangle\rangle = \langle\langle \widehat{K}, H\rangle\rangle$.

Proof One has:

$$\begin{aligned} \widehat{K} \curlyvee H &\preceq K \curlyvee H \qquad (\text{merging map}\,\mu_1) \\ H \curlyvee \widehat{K} &\preceq H \curlyvee K \qquad (\text{merging map}\,\mu_2) \end{aligned}$$

But since $K \curlyvee H$ and $H \curlyvee K$ have the same structure as well as $\widehat{K} \curlyvee H$ and $H \curlyvee \widehat{K}$, it follows $\mu_1 = \mu_2 = \mu$. Then:

$$\begin{aligned} \langle\langle \widehat{K}, H\rangle\rangle &= \sum_{q\in Q_{\widehat{K}\curlyvee H}} \varpi_{\widehat{K}\curlyvee H}(q) \sum_{\alpha,\beta\in\Sigma} \log\left(\frac{\tilde{\pi}_{\widehat{K}\curlyvee H}(q,\alpha)}{\tilde{\pi}_{\widehat{K}\curlyvee H}(q,\beta)}\right) \times \log\left(\frac{\tilde{\pi}_{H\curlyvee\widehat{K}}(q,\alpha)}{\tilde{\pi}_{H\curlyvee\widehat{K}}(q,\beta)}\right) \\ &= \sum_{q\in Q_{\widehat{K}\curlyvee H}} \sum_{\tilde{q}\in\mu^{-1}(q)} \varpi_{K\curlyvee H}(\tilde{q}) \sum_{\alpha,\beta\in\Sigma} \log\left(\frac{\tilde{\pi}_{\widehat{K}\curlyvee H}(\mu(\tilde{q}),\alpha)}{\tilde{\pi}_{\widehat{K}\curlyvee H}(\mu(\tilde{q}),\beta)}\right) \\ &\quad \times \log\left(\frac{\tilde{\pi}_{H\curlyvee\widehat{K}}(\mu(\tilde{q}),\alpha)}{\tilde{\pi}_{H\curlyvee\widehat{K}}(\mu(\tilde{q}),\beta)}\right) \\ &= \sum_{\tilde{q}\in Q_{K\vee H}} \varpi_{K\curlyvee H}(\tilde{q}) \sum_{\alpha,\beta\in\Sigma} \log\left(\frac{\tilde{\pi}_{K\curlyvee H}(\tilde{q},\alpha)}{\tilde{\pi}_{K\curlyvee H}(\tilde{q},\beta)}\right) \times \log\left(\frac{\tilde{\pi}_{H\curlyvee K}(\tilde{q},\alpha)}{\tilde{\pi}_{H\curlyvee K}(\tilde{q},\beta)}\right) \\ &= \langle\langle K, H\rangle\rangle \end{aligned}$$

$\square$

In this above proof, compatibility of the weights is crucial. Indeed it is required to distribute the states by splitting to assure compatibility with algebraic equivalence. Moreover, since the expression of the inner product is symmetric, it follows that

$$\langle\langle K, H\rangle\rangle = \langle\langle K, \widehat{H}\rangle\rangle$$

Definition 5.14 (*Inner Product on* $\mathcal{C}_s^+$) Let the binary operator

$$\langle\cdot,\cdot\rangle : \mathcal{C}_s^+ \times \mathcal{C}_s^+ \to \mathbb{R}$$

be defined for any $\lceil K\rceil, \lceil H\rceil \in \mathcal{C}_s^+$ by $\langle\lceil K\rceil, \lceil H\rceil\rangle \triangleq \langle\langle\widehat{K}, \widehat{H}\rangle\rangle$. It is well defined based on Proposition 5.13.

Theorem 5.15 $\left(\mathcal{C}_s^+, +, \cdot, \langle\cdot,\cdot\rangle\right)$ *is an Inner Product Space.*

Proof One needs to check $\langle\cdot,\cdot\rangle$ actually defines an inner product.

- *Symmetry* It directly follows from Definition 5.14.
- *Bilinearity* Let $K_1, K_2, H \in \mathcal{B}_s^+$. If all of K_1, K_2, H have the same structure, then it follows that

$$\begin{aligned}\langle K_1 + K_2, H\rangle &= \sum_{q\in Q}\varpi(q)\sum_{\alpha,\beta\in\Sigma}\log\left(\frac{\tilde{\pi}_{K_1+K_2}(q,\alpha)}{\tilde{\pi}_{K_1+K_2}(q,\beta)}\right)\log\left(\frac{\tilde{\pi}_H(q,\alpha)}{\tilde{\pi}_H(q,\beta)}\right)\\ &= \sum_{q\in Q}\varpi(q)\sum_{\alpha,\beta\in\Sigma}\log\left(\frac{\tilde{\pi}_{K_1}(q,\alpha)\tilde{\pi}_{K_2}(q,\alpha)}{\tilde{\pi}_{K_1}(q,\beta)\tilde{\pi}_{K_2}(q,\beta)}\right)\log\left(\frac{\tilde{\pi}_H(q,\alpha)}{\tilde{\pi}_H(q,\beta)}\right)\\ &= \langle K_1, H\rangle + \langle K_2, H\rangle\end{aligned}$$

Let all of K_1, K_2, H not have the same structure. Then, by assuming minimal representations of their respective classes, it follows that

$$\begin{aligned}\langle K_1 + K_2, H\rangle &= \langle (K_1 \curlyvee K_2 + K_2 \curlyvee K_1) \curlyvee H, H \curlyvee (K_1 \curlyvee K_2 + K_2 \curlyvee K_1)\rangle\\ &= \langle K_1 \curlyvee (K_2 \curlyvee H) + K_1 \curlyvee (K_2 \curlyvee H), H \curlyvee (K_1 \curlyvee K_2)\rangle\\ &= \langle K_1 \curlyvee (K_2 \curlyvee H), H \curlyvee (K_1 \curlyvee K_2)\rangle\\ &\quad + \langle K_1 \curlyvee (K_2 \curlyvee H), H \curlyvee (K_1 \curlyvee K_2)\rangle\\ &\quad \text{(because they all have the same structure} \bar{K}_1 \vee \bar{K}_2 \vee \bar{H})\\ &= \langle K_1, H\rangle + \langle K_2, H\rangle\end{aligned}$$

Now let $K, H \in \mathcal{B}_s^+$ and $\lambda \in \mathbb{R}$. If H and K have the same structure, then it follows that

$$\begin{aligned}\langle\lambda\cdot K, H\rangle &= \sum_{q\in Q}\varpi(q)\sum_{\alpha,\beta\in\Sigma}\log\left(\frac{\tilde{\pi}_K^{\lambda}(q,\alpha)}{\tilde{\pi}_K^{\lambda}(q,\beta)}\right)\times\log\left(\frac{\tilde{\pi}_H(q,\alpha)}{\tilde{\pi}_H(q,\beta)}\right)\\ &= \lambda\langle K, H\rangle\end{aligned}$$

If H and K do not have the same structure, then it follows that

$$\begin{aligned}\langle \lambda \cdot K, H \rangle &= \langle (\lambda \cdot K) \curlyvee H, H \curlyvee (\lambda \cdot K) \rangle \\ &= \langle \lambda \cdot (K \curlyvee H), H \curlyvee K \rangle \\ &= \lambda \langle K \curlyvee H, H \curlyvee K \rangle \\ &= \lambda \langle K, H \rangle\end{aligned}$$

- *Positivity* For all $K \in \mathcal{B}_s^+$, it naturally comes:

$$\langle K, K \rangle = \sum_{q \in Q} \varpi(q) \sum_{\alpha, \beta \in \Sigma} \log^2 \left(\frac{\tilde{\pi}(q, \alpha)}{\tilde{\pi}(q, \beta)} \right) \geq 0$$

because $\varpi(q) > 0$.
- *Definiteness* If $\langle K, K \rangle = 0$, then a sum of positive terms is equal to zero, which means each term is individually equal to zero:

$$\log^2 \left(\frac{\tilde{\pi}(q, \alpha)}{\tilde{\pi}(q, \beta)} \right) = 0 \quad \Rightarrow \quad \tilde{\pi}(q, \alpha) = \tilde{\pi}(q, \beta)$$

Using the normalization condition, it comes that $\tilde{\pi}(q, \sigma) = \frac{1}{|\Sigma|}$, that is $K \equiv e$.

□

In the sequel, a compatible family of weights is proposed by making use of the stationary distribution of PFSA states $\wp$, for which existence and uniqueness are guaranteed by the *Perron–Frobenius theorem* [28,29] for irreducible PFSA. Indeed a stochastic matrix $\pi : Q \times Q \to (0, 1)$ defined by:

$$\pi(q_1, q_2) \triangleq \sum_{\sigma \in \Sigma \ / \ \delta(q_1, \sigma) \ = \ q_2} \tilde{\pi}(q_1, \sigma) \tag{5.8}$$

represents the probability of transition from one state q_1 to another q_2 of an irreducible stochastic matrix π. It guarantees existence and uniqueness of the left eigenvector $\wp_K$ associated to the left eigenvalue 1. The normalized $\wp_K$ vector satisfies the following condition.

$$\sum_{q' \in Q_K} \wp_K(q') \pi_K(q', q) = \wp_K(q) \tag{5.9}$$

Lemma 5.16 *Let $K \preceq H$ (with a merging map μ) be two PFSAs of $\mathcal{B}^+$. Then, for all $q \in Q_K$:*

$$\wp_K(q) = \sum_{\tilde{q} \in \mu^{-1}(q)} \wp_H(\tilde{q})$$

Proof First, for all $\tilde{q}_1 \in Q_H, q_2 \in Q_K$:

$$\begin{aligned}
\pi_K\left[\mu(\tilde{q}_1), q_2\right] &= \sum_{\sigma \in \Sigma \ / \ \delta_K\left[\mu(\tilde{q}_1),\sigma\right]=q_2} \tilde{\pi}_K\left[\mu(\tilde{q}_1), \sigma\right] \\
&= \sum_{\sigma \in \Sigma \ / \ \delta_H(\tilde{q}_1,\sigma)\in\mu^{-1}(q_2)} \tilde{\pi}_H\left[\tilde{q}_1, \sigma\right] \\
&= \sum_{\tilde{q}_2 \in \mu^{-1}(q_2)} \ \sum_{\sigma\in\Sigma \ / \ \delta_H(\tilde{q}_1,\sigma)=\tilde{q}_2} \tilde{\pi}_H\left[\tilde{q}_1, \sigma\right] \\
&= \sum_{\tilde{q}_2 \in \mu^{-1}(q_2)} \pi_H\left(\tilde{q}_1, \tilde{q}_2\right)
\end{aligned}$$

Then define $\tilde{\wp}(q) \triangleq \sum_{\tilde{q}\in\mu^{-1}(q)} \wp_H(\tilde{q})$. What needs to be shown is that $\tilde{\wp} = \wp_K$. Indeed:

$$\begin{aligned}
\sum_{q_1\in Q_K} \tilde{\wp}(q_1)\pi_K(q_1, q_2) &= \sum_{q_1\in Q_K} \sum_{\tilde{q}_1\in\mu^{-1}(q_1)} \wp_H(\tilde{q}_1)\pi_K(\mu(\tilde{q}_1), q_2) \\
&= \sum_{q_1\in Q_K} \sum_{\tilde{q}_1\in\mu^{-1}(q_1)} \wp_H(\tilde{q}_1) \times \left(\sum_{\tilde{q}_2\in\mu^{-1}(q_2)} \pi_H(\tilde{q}_1, \tilde{q}_2)\right) \\
&= \sum_{\tilde{q}_2\in\mu^{-1}(q_2)} \left[\sum_{\tilde{q}_1\in Q_H} \wp_H(\tilde{q}_1)\pi_H(\tilde{q}_1, \tilde{q}_2)\right] \\
&= \sum_{\tilde{q}_2\in\mu^{-1}(q_2)} \wp_H(\tilde{q}_2) \\
&= \tilde{\wp}(q_2)
\end{aligned}$$

Moreover, it is clear that $\tilde{\wp}$ is normalized, and by uniqueness of the left normalized eigenvector, one has $\tilde{\wp} = \wp_K$. □

It follows from Lemma 5.16 that it is possible to obtain a compatible family of weights by putting the (equiprobable) probabilities $\frac{1}{|\Sigma|}$ on the transitions of an FSA G. This is realized by computing the lifting of e relatively to G (see Sect. 4.2) and computing its stationary distribution $\wp$.

Proposition 5.17 *Defining* $\varpi_G = \wp_{\lfloor G \rfloor_e}$*, it follows that* $\{\varpi_G, G \in \mathcal{A}_s\}$ *is a compatible family of weights.*

Proof If $F \lhd G$, then by lifting $\lfloor F \rfloor_e \preceq \lfloor G \rfloor_e$, and the result follows from the Lemma 5.16. □

Remark 5.6 The inner product, introduced in this section, can be interpreted to be analogous to the information contained in a PFSA. Let us compare the norm (induced by the inner product) of a PFSA, defined as:

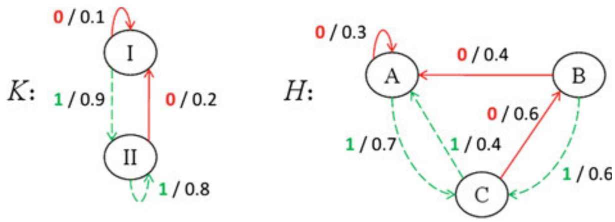


Fig. 3 PFSA K and H to be summed up

$$||K||^2 = \sum_{q \in Q} \varpi(q) \sum_{\alpha, \beta \in \Sigma} \left[\log \tilde{\pi}(q, \alpha) - \log \tilde{\pi}(q, \beta)\right]^2$$

with the entropy rate [30], in the field of Information Theory, defined as:

$$h_K(Q|\Sigma) = -\sum_{q \in Q} \wp(q) \sum_{\sigma \in \Sigma} \tilde{\pi}(q, \sigma) \log \tilde{\pi}(q, \sigma)$$

It follows that the norm corresponds to a mean-square version of the information content. The above two formulae are closely related in the sense that the maximum of entropy rate (achieved for the symbolic white noise) corresponds to the minimum of the norm achieved by $\|e\| = 0$.

6 Validation of theoretical results

Two examples are presented in this section to validate the theoretical work, developed in this paper. The first example is numerical in nature and illustrates the algebraic operations of join and vector addition. The second example analyzes experimental data of sensor time series for pattern recognition in the context of robot motion behavior.

6.1 Example 1: join and vector addition

The objective of this example is to demonstrate the vector addition of two irreducible and synchronizable PFSA. Figure 3 shows two PFSA $K, H \in \mathcal{B}_s^+$ that are constructed over the alphabet $\Sigma = \{0, 1\}$. The synchronizable word for K is w_K =“0” which always leads to state **I**; and the synchronizable word for H is w_H =“01” which always leads to state **C**.

Since the PFSA K and H have different algebraic structures, their join compositions need to be computed as shown in Fig. 4. The behavior of the merging maps μ_K and μ_H (see Eq. 3.2), respectively related to $K \preceq K \curlyvee H$ and $H \preceq H \curlyvee K$, is shown in dash-line circles in Fig. 4. As explained in Sect. 3.2, the concatenated word $w_H w_K$ =“010” is a synchronizable word for $\bar{K} \vee \bar{H}$. This word, indeed, always leads to state **3**. Using Eq. 5.1, the addition is now computed as shown in Fig. 5.

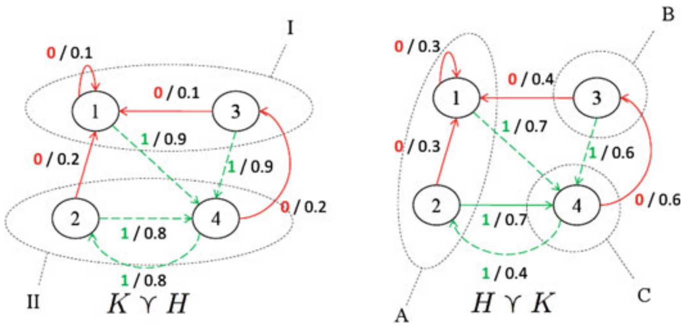


Fig. 4 Join compositions $K \curlyvee H$ and $H \curlyvee K$

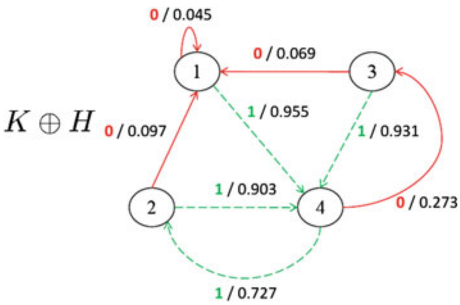
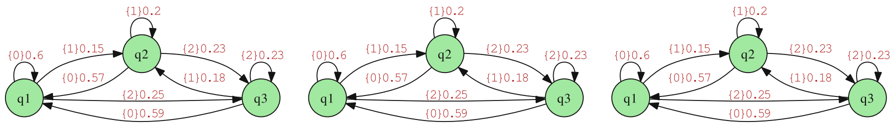


Fig. 5 The vector addition $K \oplus H$

6.2 Example 2: pattern recognition of Robot motion behavior

This subsection presents modeling of robot motion profile from sensor time series and its identification by pattern matching. The results are derived from experimental data in a laboratory environment, consisting of a wireless networked system incorporating mobile robots, robot simulators, and distributed sensors [31]. In this application, D-Markov algorithms [3,6] have been used to construct PFSA as representations of different types of robot motions. To this end, sensor time series data are collected for two different types of robot motion. Accordingly, two PFSA G_1 and G_2 are constructed from the respective training data to serve as reference patterns for these two types of motion. A third PFSA P is constructed from a different time series of test data, where the unknown type of motion profile is to be identified. Figure 6 depicts the three PFSA, G_1, G_2 and P, where each PFSA is constructed with the alphabet cardinality $|\Sigma| = 3$ and depth $D = 1$ of the D-Markov machine [3]. In this example, since all three PFSA have the same structure, it suffices to obtain their respective maps by Eq. 5.1. Now, the problem of robot behavior identification is stated as follows.

> Given a PFSA P constructed from an (experimentally generated) time series of robot motion, the problem is to classify P into one of the two patterns: G_1 and G_2 with the underlying norm induced from the inner product in Definition 5.14.



(a) G_1 **(b)** G_2 **(c)** P

Fig. 6 PFSA representing different motions of robots

$$\widetilde{\Pi}^{P} = \left\{ \begin{matrix} 0.68 & 0.14 & 0.18 \\ 0.61 & 0.14 & 0.25 \\ 0.68 & 0.14 & 0.18 \end{matrix} \right\}$$

$$\xrightarrow[\text{termwise}]{\text{Take inverse}} \left\{ \begin{matrix} 1/0.68 & 1/0.14 & 1/0.18 \\ 1/0.61 & 1/0.14 & 1/0.25 \\ 1/0.68 & 1/0.14 & 1/0.18 \end{matrix} \right\}$$

$$\xrightarrow[\text{Rows}]{\text{Normalize}} \left\{ \begin{matrix} 0.10 & 0.51 & 0.39 \\ 0.13 & 0.55 & 0.32 \\ 0.10 & 0.51 & 0.39 \end{matrix} \right\} = \widetilde{\Pi}^{-P}$$

To verify that the probability map of the PFSA $G_1 - P = G_1 + (-P)$, the matrix $\widetilde{\Pi}^{G_1}$ is first multiplied elementwise (denoted by $.\times$) with $\widetilde{\Pi}^{-P}$ and then each row is normalized.

$$\widetilde{\Pi}^{G_1} . \times \widetilde{\Pi}^{-P} = \left\{ \begin{matrix} 0.60 \times 0.10 & 0.15 \times 0.51 & 0.25 \times 0.39 \\ 0.57 \times 0.13 & 0.20 \times 0.55 & 0.23 \times 0.32 \\ 0.59 \times 0.10 & 0.18 \times 0.51 & 0.23 \times 0.39 \end{matrix} \right\}$$

$$\xrightarrow[\text{Rows}]{\text{Normalize}} \left\{ \begin{matrix} 0.27 & 0.32 & 0.41 \\ 0.29 & 0.43 & 0.28 \\ 0.25 & 0.38 & 0.37 \end{matrix} \right\} = \widetilde{\Pi}^{(G_1 - P)}$$

By Eq. 5.6, where the natural logarithm is used, the distance between the observed pattern P and the reference patterns is obtained as: $\|G_1 - P\| = 0.5743$ and $\|G_2 - P\| = 0.9098$, respectively. Since $\|G_1 - P\| < \|G_2 - P\|$, it is concluded that P is closer to G_1 and hence the robot motion is classified to be G_1.

7 Summary, conclusions, and future research

This paper constructs a vector space on a class of PFSA in an algebraic setting without any direct dependence on the associated concept of probability measures [16,17]. In contrast to the general setting of a HMM built upon a nondeterministic FSA structure [7], this paper assumes a deterministic structure of the underlying FSA (see Definition 2.2) in the PFSA construction. Since a nondeterministic FSA can be exactly represented by a deterministic FSA [20], the proposed algebraic approach might be extendable to a nondeterministic FSA structure, which would, however, increase the complexity of the PFSA's vector space representation. In such cases, a probabilistic approach described in[16,17] could be more suitable to develop a similar mathematical framework for PFSA whose FSA structure is nondeterministic. This issue has not been addressed in this paper.

Instead of considering classes of PFSA generating the same stationary probability-measure, this paper enlightens the notion of *algebraic equivalence*, which is solely described by manipulations of the structure of the automata. This paper is developed in the context of *stationarity*. Not any PFSA is relevant in this context, and only those satisfying the specific properties, namely *irreducibility* (see Definition 2.4) and *synchronizability* (see Definition 3.3), are addressed.

In general, normed spaces can be constructed on this vector space of PFSA. Specifically, the structure of an inner product space is developed on the quotient space of irreducible and synchronizable PFSA. In this respect, two examples are presented to demonstrate applicability of the proposed theoretical work, which includes pattern recognition based on the behavior of robot motion in a laboratory environment.

This mathematical framework is motivated by various applications in dynamical systems modeling, analysis, and control in a stochastic setting. One of the applications of this mathematical framework is pattern recognition in identification of robot motion behavior [31]. While there are many other topics for both theoretical and application-oriented research, a few examples are presented below.

- Extension of the vector space structure to quasi-synchronizable PFSA using a weaker version of the synchronizability condition [24] (See Remark 5.1).
- Extension of the vector space structure to accommodate PFSA whose FSA could be nondeterministic.
- Study of the connection between these algebraic developments and the probability-measures setting from a stationary perspective, and an extension of the vector space structure to the probability-measure space.
- Development of an analytical procedure for selection of an inner product on the vector space of PFSA, which will allow an appropriate choice of the metric for the problem at hand. This is a natural continuation of the present work.
- Formulation of one or more methods of decision and control synthesis in discrete-event dynamical systems based on a vector-space model of PFSA, analogous to classical control synthesis methods in finite-dimensional vector spaces.

Appendix A

Theorem A.1 (Characterization of an irreducible FSA) *Let $G \in \bar{\mathcal{A}}$ with state cardinality $|Q| > 1$. Then G is irreducible if and only if for any $A, B \subset Q$ such that $\{A, B\}$ forms a partition of Q, there exists one transition from A to B.*

Note If $|Q_G| = 1$, there exists no partition of Q_G, on the other hand G is always irreducible.

Proof $\boxed{\Rightarrow}$ Let $\{A, B\}$ be a partition of Q_G, with $q_A \in A$ and $q_B \in B$. Then since G is irreducible, there exists a word $w = \sigma_1 \ldots \sigma_N$ which goes from q_A to q_B. Let $q_1 \triangleq q_A$ and for all $i \in \{2, \ldots, N+1\}$, let $q_i \triangleq \delta(q_{i-1}, \sigma_{i-1})$. Of course one has $q_{N+1} = q_B$.

Let $\Gamma \triangleq \{i \in \{1, \ldots, N+1\} \mid q_i \in B\}$. Then $N+1 \in \Gamma$ and $1 \notin \Gamma$. Since Γ is a non-empty subset of N, it has a minimal element: $i_0 \triangleq \min \Gamma$ with $i_0 > 1$. Then $q_{i_0-1} \in A$ and $q_{i_0} \in B$, and therefore, $\delta(q_{i_0-1}, \sigma_{i_0-1})$ is a transition from A to B.

$\boxed{\Leftarrow}$ Let $q, q' \in Q_G$. First, assume that $q \neq q'$.

A word of transitions from q to q' needs to be constructed.

Let $q_1 \triangleq q$, and then, let be the following recursive construction for $i < |G|$:

$$A_i \triangleq \{q_j, j \leq i\}$$
$$B_i \triangleq Q_G - A_i$$

Since $\{A_i, B_i\}$ forms a partition of Q_G, it is known by assumption that there exist $q_{i+1} \in B_i, \tilde{q}_i \in A_i$ and $\sigma_i \in \Sigma$ such that $q_{i+1} = \delta(\tilde{q}_i, \sigma_i)$.

Inductively, $\forall i \in \{2, \ldots, |G|\}$, let be the assertion P_i: "there exists a word w_i which goes from $q_1 = q$ to q_i".

$\underline{P_2}$: obvious considering $q_2 = \delta(q_1, \sigma_1)$. The word is: $w_2 = \sigma_1$.

$\underline{P_2 \wedge P_3 \wedge \ldots \wedge P_i \Rightarrow P_{i+1}}$: here, there is some $\tilde{q}_i \in A_i$ and $\sigma_i \in \Sigma$ such that $q_{i+1} = \delta(\tilde{q}_i, \sigma_i)$. If $\tilde{q}_i = q$, then simply set $w_{i+1} = \sigma_i$. If not, then there exists j_0 with $2 < j_0 \leq i$ such that $\tilde{q}_i = q_{j_0}$, and then $w_{i+1} = w_{j_0}\sigma_i$ is a word which goes from q to q_i.

Since P_i is true, there is, specifically, a word from q to q'.

Now, if $q = q'$, consider $q'' \neq q$ (which exists because $|G| > 1$), then there exists a word w from q to q'', and a word w' from q'' to q. And finally, ww' goes from q to itself. □

Appendix B

Theorem B.1 *(Join of Two Irreducible FSA) Any two classes of irreducible FSA* $[G_1], [G_2]$ *in* $\mathcal{A}/\mathfrak{S}$ *admit a join denoted as* $G_1 \vee G_2$ *which is irreducible.*

Proof The proof of Theorem B.1 requires two lemmas that are presented below.

Lemma B.2 *Let* $N \in \mathbb{N}^*$ *and* $\{G_i\}_{i=1\ldots N}$ *be FSA. Let* H *be an FSA such that* $G_i \lhd H$ *(with the merging map* $\mu_i : Q_H \to Q_{G_i}$*) for all* i*. If there exist two different states* $q_1, q_2 \in Q_H$ *such that* $\mu_i(q_1) = \mu_i(q_2) = q_i \in Q_i$ *for all* i*, then one of these states can be 'removed'; that is, there exists an FSA* $\tilde{H}$ *such that* $G_i \lhd \tilde{H}$ *for all* i *with* $Q_{\tilde{H}} = Q_H \backslash \{q_2\}$.

Proof Let $Q_{\tilde{H}} = Q_H \backslash \{q_2\}$ and the map $\nu : Q_H \to Q_{\tilde{H}}$ be such that

$$\nu(q) = \begin{cases} q & \text{if } q \neq q_2 \\ q_1 & \text{otherwise} \end{cases}$$

Let now $\delta_{\tilde{H}}$ be defined by $\delta_{\tilde{H}}(q, \sigma) \triangleq \nu[\delta_H(q, \sigma)]$ for all $q \in Q_{\tilde{H}}$.

Now one can build $\tilde{\mu}_i : Q_{\tilde{H}} \to Q_{G_i}$ defined as $\tilde{\mu}_i(q) \triangleq \mu_i(q)$. Clearly, $\tilde{\mu}_i \circ \nu = \mu_i$ and these maps are surjective, because $\tilde{\mu}_i^{-1}(\{q\}) = \nu(\mu_i^{-1}(\{q\}))$ for all $q \in Q_G$, and are merging maps because:

$$
\begin{aligned}
\tilde{\mu}_i\left[\delta_{\tilde{H}}(q,\sigma)\right] &= \tilde{\mu}_i\left[\nu\left(\delta_H(q,\sigma)\right)\right] \\
&= \tilde{\mu}_i \circ \nu\left[\delta_H(q,\sigma)\right] \\
&= \mu_i\left[\delta_H(q,\sigma)\right] \\
&= \delta_G\left[\mu_i(q),\sigma\right] \\
&= \delta_G\left[\tilde{\mu}_i(q),\sigma\right]
\end{aligned}
$$

□

The next lemma makes use of the concept of a stable subset as presented below.

Let $G \in \mathcal{A}$. Then, $Q_{G,S} \subseteq Q_G$ is said to be a *stable subset of states* of Q_G if $Q_{G,S} \neq \emptyset$; and $\delta(q_S,\sigma) \in Q_{G,S}$ for all $q_S \in Q_{G,S}$, $\sigma \in \Sigma$.

Following the characterization of an irreducible FSA in Theorem A.1, one can prove that G is irreducible *if and only if* the only possible stable subset of states of G is Q_G itself.

Lemma B.3 *Let $N \in \mathbb{N}^*$ and let $\{G_i\}_{i=1\ldots N}$ be a family of irreducible FSA and H be an FSA such that $G_i \lhd H$ for all i. If $Q_{H,S}$ is a stable subset of states relatively to H, then H induces an FSA $\tilde{H}$ defined on $Q_{H,S}$ such that $G_i \lhd \tilde{H}$, for all i.*

Proof Let μ_i be the merging maps relatively to $G_i \lhd H$. Let $\tilde{H}$ be defined on $Q_{H,S}$ as follows: $\delta_{\tilde{H}}(q_S,\sigma) \triangleq \delta_H(q_S,\sigma)$ for $q_S \in Q_{H,S}$, which is well defined because of the stability property. Then, the map, $\mu_{\tilde{H},i}(q_S) \triangleq \mu_i(q_S)$ for all i, is surjective.

Indeed, by contradiction, considering $Q_{0,i} \subseteq Q_i$ the set of elements of Q_i which do not admit a pre-image via $\mu_{\tilde{H},i}$, since G_i is irreducible, there must exist a transition from an element $q_1 \in Q_i - Q_{0,i}$ to an element of $q_2 \in Q_{0,i}$. But q_2 has a pre-image set $\mu_i^{-1}(\{q_2\})$ in Q_H via μ_i (since μ_i is surjective). This pre-image set has an empty intersection with $Q_{H,S}$ (because q_2 has no pre-image by $\mu_{\tilde{H},i}$). On the other hand, q_1 admits a pre-image $q_1' \in Q_{H,S}$. Then since there is a transition from q_1 to q_2 in G_i, there must exist a transition from q_1' to some $q_2' \in \mu_i^{-1}(q_2)$ in H which is in contradiction with the fact $Q_{H,S}$ is a stable subset relatively to H.

Finally $\mu_{\tilde{H},i}$ is a merging map because one has:

$$
\begin{aligned}
\mu_{\tilde{H},i}\left[\delta_H(q_S,\sigma)\right] &= \mu_i\left[\delta_H(q_S,\sigma)\right] \\
&= \delta_i\left[\mu_i(q_S),\sigma\right] \\
&= \delta_i\left[\mu_{\tilde{H},i}(q_S),\sigma\right]
\end{aligned}
$$

□

Now the proof of Theorem B.1 is constructed as follows.

Proof By excluding the trivial case, where $|Q_{G_1}| = 1$ or $|Q_{G_2}| = 1$, the join $G_1 \vee G_2$ is constructed as the upper bound of G_1 and G_2 having the minimal number of states. It is subsequently proven that this FSA satisfies with the axioms of the join.

Let Γ be the set of all (not necessarily irreducible) upper bounds of FSA G_1 and G_2. It is not empty because the synchronous composition $G_1 \otimes G_2$ is in Γ. Let $\Gamma' \subset \Gamma$ be the set of upper-bounds having the minimal number of states, say m. Using this minimality condition, the rest of the proof shows that any $H \in \Gamma'$ is a join (and therefore is essentially unique, up to state relabeling).

Claim *For all $H \in \Gamma'$, H is irreducible.*

Indeed, if H is not irreducible, it admits a proper stable subset of states $Q_{H,T}$. Therefore, in virtue of Lemma B.3, since G_1 and G_2 are irreducible, there exists an FSA $\tilde{H} \in \Gamma$ defined on $Q_{H,S}$, with $|\tilde{H}| < m$ which is a contradiction.

To complete the proof, one now needs to show that for all $H \in \Gamma'$, and for all irreducible $F \in \Gamma$, one has $H \lhd F$. For $i \in \{1, 2\}$, there exists two merging maps $\mu_{H,i} : Q_H \to Q_i$ and $\mu_{F,i} : Q_F \to Q_i$ since both H and F are upper bounds of G_1 and G_2. Then, a binary relation $\sim$ is defined over $Q_H \times Q_F$ as follows:

$$q_H \sim q_F \text{ if } \forall i \in \{1, 2\}\,,\ \mu_{H,i}(q_H) = \mu_{F,i}(q_F)$$

Consequently, q_H and q_F are said to be linked together.

Claim *If $q_H \sim q_F$ and $q'_H \sim q_F$, then $q_H = q'_H$.*

Indeed, if not, in virtue of Lemma B.2, q_H and q'_H could be merged together which would contradict the minimality condition of H.

Let $Q_{H,L} \subseteq Q_H$ be the set of the *linked* elements of Q_H, and similarly $Q_{F,L} \subset Q_F$ be the set of *linked* elements of Q_F.

Claim *$Q_{H,L}$ is a stable subset of Q_H; and $Q_{F,L}$ is a stable subset of Q_F.*

Indeed, for $Q_{H,L}$, by contradiction, let $q_{H,L} \sim q_{F,L}$ be two linked elements, $q' \notin Q_{H,L}$ be a non-linked element of Q_H, and $\sigma \in \Sigma$ such that $q' = \delta_H(q_{H,L}, \sigma)$. Then, for all $i \in \{1, 2\}$:

$$\begin{aligned}
\mu_{H,i}(q') &= \mu_{H,i}\left[\delta_H(q_{H,L}, \sigma)\right] \\
&= \delta_i\left[\mu_{H,i}(q_{H,L}), \sigma\right] \\
&= \delta_i\left[\mu_{F,i}(q_{F,L}), \sigma\right] \\
&= \mu_{F,i}\left[\delta_F(q_{F,L}, \sigma)\right]
\end{aligned}$$

And thus, $q' \sim \delta_F(q_{F,L})$, which is a contradiction.

The proof of the claim is identical for $Q_{F,L}$.

Since H and F are irreducible, it directly follows that $Q_{H,L} = Q_H$ and $Q_{F,L} = Q_F$.

It is then possible to define a map $\tilde{\mu} : Q_F \to Q_H$ which to any element $q_F \in Q_F$ associates its only linked element in Q_H.

The map $\tilde{\mu}$ is surjective, because any element of Q_H is *linked*. Finally this map satisfies, for all $q \in Q_F, \sigma \in \Sigma, i \in \{1, 2\}$, with:

$$\begin{aligned}
\mu_{H,i}\left[\delta_H\left(\tilde{\mu}(q), \sigma\right)\right] &= \delta_i\left[\mu_{H,i} \circ \tilde{\mu}(q), \sigma\right] \\
&= \delta_i\left[\mu_{F,i}(q), \sigma\right] \\
&\quad \text{(because } \tilde{\mu}(q) \sim q\text{)} \\
&= \mu_{F,i}\left[\delta_F(q, \sigma)\right]
\end{aligned}$$

That is $\delta_H\left[\tilde{\mu}(q), \sigma\right] \sim \delta_F(q, \sigma)$; and therefore $\tilde{\mu}\left[\delta_F(q, \sigma)\right] = \delta_H\left[\tilde{\mu}(q), \sigma\right]$.
And then, $H \lhd F$.

This last statement implies that for all $H, H' \in \Gamma'$, both $H \lhd H'$ and $H' \lhd H$, that is $H \mathfrak{S} H'$. Hence, all the elements of Γ' are equal up to state relabeling. Let this class of elements be $G_1 \vee G_2$. This completes the proof of Lemma B.3. □

This proves Theorem B.1. □

References

1. Egerstedt MB, Frazzoli E, Pappas G (2006) Special section on symbolic methods for complex control systems. IEEE Trans Autom Control 51(6):921–923
2. Gupta S, Ray A (2009) Statistical mechanics of complex systems for pattern identification. J Stat Phys 134(2):337–364
3. Ray A (2004) Symbolic dynamic analysis of complex systems for anomaly detection. Signal Processing 84(7):1115–1130
4. Gupta S, Ray A, Keller E (2007) Symbolic time series analysis of ultrasonic data for early detection of fatigue damage. Mech Syst Signal Process 21(2):866–884
5. Phoha S, LaPorta T, Griffin C (2006) Sensor network operations. IEEE Press, Piscataway
6. Rajagopalan V, Ray A (2006) Symbolic time series analysis via wavelet-based partitioning. Signal Process 86(11):3309–3320
7. Vidal E, Thollard F, de la Higuera C, Casacuberta F, Carrasco RC (2005) Probabilistic finite-state machines—Part I. IEEE Trans Pattern Anal Mach Intell 27(7):1013–1025
8. Vidyasagar M (2011) The complete realization problem for hidden Markov models: a survey and some new results. Math Control Signals Syst 23:1–65
9. Rabiner L (1989) A tutorial on hidden Markov models and selected applications in speech proccessing. Proc IEEE 77(2):257–286
10. Carrasco R, Oncina J (1994) Learning stochastic regular grammars by means of a state merging method. In: Proceedings of second international colloquium on grammatical inference and applications, Alicante, Spain, pp 139–152
11. Saul L, Pereira F (1997) Aggregate and mixed-order Markov models for statistical language processing. In: Proceedings of the second conference on empirical methods in natural language processing, Brown University, Providence, pp 81–89
12. Ron D, Singer Y, Tishby N (1998) On the learnability and usage of acyclic probabilistic finite automata. J Comput Syst Sci 56(2):133–152
13. Ray A (2005) Signed real measure of regular languages for discrete event supervisory control. Int J Control 78(12):949–967
14. Barfoot T, D'Eleuterio G (2002) 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, UK, pp 14–18
15. Chattopadhyay I, Ray A (2008) Structural transformations of probabilistic finite state machines. Int J Control 81(5):820–835
16. Wen Y, Ray A, Chattopadhyay I, Phoha S (2011) Modeling of symbolic systems: Part I-vector space representation of probabilistic finite state automata. In: Proceedings of American control conference, San Francisco, CA, USA, pp 5133–5138
17. Wen Y, Ray A, Chattopadhyay I, Phoha S (2011) Modeling of SYMBOLIC SYSTems: Part II-Hilbert space construction for model identification and order reduction. In: Proceedings of American control conference, San Francisco, CA, USA, pp 5139–5144
18. Rozenberg G, Salomaa A (eds) (1987) Handbook of formal languages, vol 2, Springer, Berlin
19. Lind D, Marcus M (1995) An introduction to symbolic dynamics and coding. Cambridge University Press, Cambridge
20. Hopcroft JE, Motwani R, Ullman JD (2001) Introduction to automata theory, languages, and computation. 2nd edn. Addison-Wesley, Boston
21. Gill A (1976) Applied algebra for the computer sciences. Prentice-Hall, Englewood Cliffs

22. Shalizi C, Shalizi K (2004) 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, Corvallis, pp 504–511
23. Chattopadhyay I, Wen Y, Ray A, Phoha S (2011) Unsupervised inductive learning in symbolic sequences via recursive identification of self-similar semantics. In: Proceedings of American control conference, San Francisco, CA, USA, pp 125–130 (preprint)
24. Youra H, Inoue T, Masuzawa T, Fujiwara H (1998) On the synthesis of synchronizable finite state machines with partial scan. Syst Comput Japan 29:53–62
25. Eilenberg S (1974) Automata, languages and machines. Academic Press, New York
26. Hartmanis J, Stearns RE (1966) Algebraic structure theory of sequential machines. Prentice Hall, Englewood Cliffs
27. Rudin W (1987) Real and complex analysis, 3rd edn. McGraw-Hill, New York
28. Bapat RB, Raghavan TES (1997) Nonnegative matrices and applications, Chap. 1. Cambridge University Press, Cambridge
29. Berman A, Plemmons RJ (1994) Nonnegative matrices in the mathematical sciences, chap 1. SIAM, Philadelphia
30. Cover TM, Thomas JA (2006) Elements of information theory. 2nd edn. Wiley-Interscience, Hoboken
31. Mallapragada G, Ray A, Jin X (2012) Symbolic dynamic filtering and language measure for behavior identification of mobile robots. IEEE Trans Syst Man Cybern Part B Cybern. doi:10.1109/TSMCB.2011.2172419