Santa Fe Institute Working Paper 10-11-031
arxiv.org:1008.4182 [nlin.CD]

Exact Synchronization for Finite-State Sources

Nicholas F. Travers[1,2,*] and James P. Crutchfield[1,2,3,4,†]

[1] *Complexity Sciences Center*
[2] *Mathematics Department*
[3] *Physics Department*
University of California at Davis,
One Shields Avenue, Davis, CA 95616
[4] *Santa Fe Institute*
1399 Hyde Park Road, Santa Fe, NM 87501

(Dated: December 29, 2010)

We analyze how an observer synchronizes to the internal state of a finite-state information source, using the ϵ-machine causal representation. Here, we treat the case of exact synchronization, when it is possible for the observer to synchronize completely after a finite number of observations. The more difficult case of strictly asymptotic synchronization is treated in a sequel. In both cases, we find that an observer, on average, will synchronize to the source state exponentially fast and that, as a result, the average accuracy in an observer's predictions of the source output approaches its optimal level exponentially fast as well. Additionally, we show here how to analytically calculate the synchronization rate for exact ϵ-machines and provide an efficient polynomial-time algorithm to test ϵ-machines for exactness.

PACS numbers: 02.50.-r 89.70.+c 05.45.Tp 02.50.Ey

I. INTRODUCTION

Synchronization and state estimation for finite-state sources is a central interest in several disciplines, including information theory, theoretical computer science, and dynamical systems [1–5]. Here, we study the synchronization problem for a class of finite-state hidden Markov models known as ϵ-machines [6]. These machines have the important property of *unifilarity*, meaning that the next state is completely determined by the current state and the next output symbol generated. Thus, if an observer is ever able to synchronize to the machine's internal state, it remains synchronized forever using continued observations of the output. As we will see, our synchronization results also have important consequences for prediction. The future output of an ϵ-machine is a function of the current state, so better knowledge of its state enables an observer to make better predictions of the output.

II. BACKGROUND

This section provides the necessary background for our results, including information-theoretic measures of prediction for stationary information sources and formal definitions of ϵ-machines and synchronization. In particular, we identify two qualitatively distinct types of synchronization: exact (synchronization via finite observation sequences) and asymptotic (requiring infinite sequences). The exact case is the subject here; the nonexact case is treated in a sequel [7].

*Electronic address: ntravers@math.ucdavis.edu
†Electronic address: chaos@ucdavis.edu

A. Stationary Information Sources

Let $\mathcal{A}$ be a finite alphabet, and let $X_0, X_1, \ldots$ be the random variables (RVs) for a sequence of observed symbols $x_t \in \mathcal{A}$ generated by an information source. We denote the RVs for the sequence of future symbols beginning at time $t = 0$ as $\overrightarrow{X} = X_0 X_1 X_2 ...$, the *block* of L symbols beginning at time $t = 0$ as $\overrightarrow{X}^L = X_0 X_1 ... X_{L-1}$, and the block of L symbols beginning at a given time t as $\overrightarrow{X}_t^L = X_t X_{t+1} ... X_{t+L-1}$. A *stationary source* is one for which $\Pr(\overrightarrow{X}_t^L) = \Pr(\overrightarrow{X}_0^L)$ for all t and all $L > 0$.

We monitor an observer's predictions of a stationary source using information-theoretic measures [8], as reviewed below.

Definition 1. *The* block entropy $H(L)$ *for a stationary source is:*

$$H(L) \equiv H[\overrightarrow{X}^L] = - \sum_{\{\overrightarrow{x}^L\}} \Pr(\overrightarrow{x}^L) \log_2 \Pr(\overrightarrow{x}^L) \ .$$

The block entropy gives the average uncertainty in observing blocks $\overrightarrow{X}^L$.

Definition 2. *The* entropy rate h_μ *is the asymptotic average entropy per symbol:*

$$\begin{aligned} h_\mu &\equiv \lim_{L \to \infty} \frac{H(L)}{L} \\ &= \lim_{L \to \infty} H[X_L | \overrightarrow{X}^L] \ . \end{aligned}$$

Definition 3. *The entropy rate's* length-L approximation *is:*

$$\begin{aligned} h_\mu(L) &\equiv H(L) - H(L-1) \\ &= H[X_{L-1} | \overrightarrow{X}^{L-1}] \ . \end{aligned}$$

That is, $h_\mu(L)$ is the observer's average uncertainty in the next symbol to be generated after observing the first $L-1$ symbols.

For any stationary process, $h_\mu(L)$ monotonically decreases to the limit h_μ [8]. However, the form of convergence depends on the process. The lower the value of h_μ a source has, the better an observer's predictions of the source output will be asymptotically. The faster $h_\mu(L)$ converges to h_μ, the faster the observer's predictions reach this optimal asymptotic level. If we are interested in making predictions after a finite observation sequence, then the source's true entropy rate h_μ, as well as the rate of convergence of $h_\mu(L)$ to h_μ, are both important properties of an information source.

B. Hidden Markov Models

In what follows we restrict our attention to an important class of stationary information sources known as hidden Markov models. For simplicity, we assume the number of states is finite.

Definition 4. *A finite-state edge-label* hidden Markov machine *(HMM) consists of*

1. *a finite set of states* $\boldsymbol{\mathcal{S}} = \{\sigma_1, ..., \sigma_N\}$*,*
2. *a finite alphabet of symbols* $\mathcal{A}$*, and*
3. *a set of* N *by* N *symbol-labeled transition matrices* $T^{(x)}$*,* $x \in \mathcal{A}$*, where* $T^{(x)}_{ij}$ *is the probability of transitioning from state* σ_i *to state* σ_j *on symbol* x*. The corresponding overall state-to-state transition matrix is denoted* $T = \sum_{x\in\mathcal{A}} T^{(x)}$*.*

A hidden Markov machine can be depicted as a directed graph with labeled edges. The nodes are the states $\{\sigma_1, ..., \sigma_N\}$ *and for all* x, i, j *with* $T^{(x)}_{ij} > 0$ *there is an edge from state* σ_i *to state* σ_j *labeled* $p|x$ *for the symbol* x *and transition probability* $p = T^{(x)}_{ij}$*. We require that the transition matrices* $T^{(x)}$ *be such that this graph is strongly connected.*

A hidden Markov machine M generates a stationary process $\mathcal{P} = (X_L)_{L\geq 0}$ as follows. Initially, M starts in some state σ_{i^*} chosen according to the stationary distribution π over machine states—the distribution satisfying $\pi T = \pi$. It then picks an outgoing edge according to their relative transition probabilities $T^{(x)}_{i^* j}$, generates the symbol x^* labeling this edge, and follows the edge to a new state σ_{j^*}. The next output symbol and state are consequently chosen in a similar fashion, and this procedure is repeated indefinitely.

We denote $\mathcal{S}_0, \mathcal{S}_1, \mathcal{S}_2, \ldots$ as the RVs for the sequence of machine states visited and $X_0, X_1, X_2, \ldots$ as the RVs for the associated sequence of output symbols generated. The sequence of states $(\mathcal{S}_L)_{L\geq 0}$ is a Markov chain with transition kernel T. However, the stochastic process we consider is not the sequence of states, but rather the associated sequence of outputs $(X_L)_{L\geq 0}$, which generally is not Markovian. We assume the observer directly observes this sequence of outputs, but does not have direct access to the machine's "hidden" internal states.

C. Examples

In what follows, it will be helpful to refer to several example hidden Markov machines that illustrate key properties and definitions. We introduce four examples, all with a binary alphabet $\mathcal{A} = \{0, 1\}$.

1. Even Process

Figure 1 gives a HMM for the Even Process. Its transitions matrices are:

$$T^{(0)} = \begin{pmatrix} p & 0 \\ 0 & 0 \end{pmatrix} ,$$
$$T^{(1)} = \begin{pmatrix} 0 & 1-p \\ 1 & 0 \end{pmatrix} . \qquad (1)$$

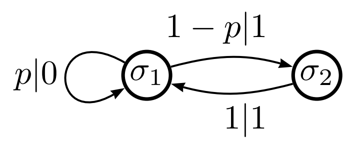

FIG. 1: A hidden Markov machine (the ϵ-machine) for the Even Process. The transitions denote the probability p of generating symbol x as $p|x$.

The support for the Even Process consists of all binary sequences in which blocks of uninterrupted 1s are even in length, bounded by 0s. After each even length is reached, there is a probability p of breaking the block of 1s by inserting a 0. The hidden Markov machine has two internal states, $\boldsymbol{\mathcal{S}} = \{\sigma_1, \sigma_2\}$, and a single parameter $p \in (0, 1)$ that controls the transition probabilities.

2. Alternating Biased Coins Process

Figure 2 shows a HMM for the Alternating Biased Coins (ABC) Process. The transitions matrices are:

$$T^{(0)} = \begin{pmatrix} 0 & 1-p \\ 1-q & 0 \end{pmatrix} ,$$
$$T^{(1)} = \begin{pmatrix} 0 & p \\ q & 0 \end{pmatrix} . \qquad (2)$$

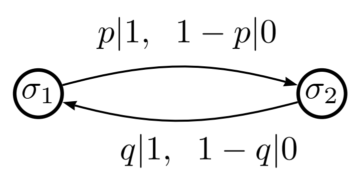

FIG. 2: A hidden Markov machine (the ϵ-machine) for the Alternating Biased Coins Process.

The process generated by this machine can be thought of as alternatively flipping two coins of different biases $p \neq q$.

3. SNS Process

Figure 3 depicts a two-state HMM for the SNS Process which generates long sequences of 1s broken by isolated 0s. Its matrices are:

$$T^{(0)} = \begin{pmatrix} 0 & 0 \\ 1-q & 0 \end{pmatrix} ,$$
$$T^{(1)} = \begin{pmatrix} p & 1-p \\ 0 & q \end{pmatrix} . \quad (3)$$

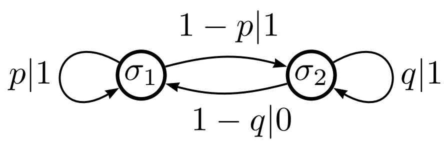

FIG. 3: An HMM for the SNS Process.

Note that the two transitions leaving state σ_1 both emit $x = 1$.

4. Noisy Period-2 Process

Finally, Fig. 4 depicts a nonminimal HMM for the Noisy Period-2 (NP2) Process. The transition matrices are:

$$T^{(0)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1-p & 0 \\ 0 & 0 & 0 & 0 \\ 1-p & 0 & 0 & 0 \end{pmatrix} ,$$
$$T^{(1)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & 1 \\ p & 0 & 0 & 0 \end{pmatrix} . \quad (4)$$

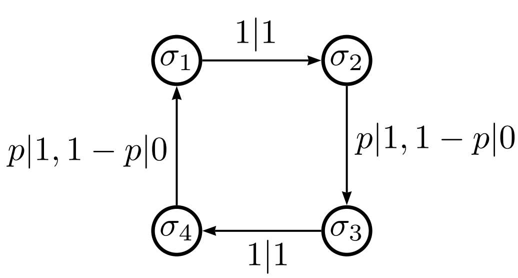

FIG. 4: An HMM for the Noisy Period-2 Process.

It is clear by inspection that the same process can be captured by a hidden Markov machine with fewer states. Specifically, the distribution over future sequences from states σ_1 and σ_3 are the same, so those two states are redundant and can be merged. The same is also true for states σ_2 and σ_4.

D. ϵ-Machines

We now introduce a class of hidden Markov machines that has a number of desirable properties for analyzing synchronization.

Definition 5. *A* finite-state ϵ-machine *is a finite-state edge-label hidden Markov machine with the following properties:*

1. Unifilarity*: For each state $\sigma_k \in \boldsymbol{\mathcal{S}}$ and each symbol $x \in \mathcal{A}$ there is at most one outgoing edge from state σ_k labeled with symbol x.*
2. Probabilistically distinct states*: For each pair of distinct states $\sigma_k, \sigma_j \in \boldsymbol{\mathcal{S}}$ there exists some finite word $w = x_0 x_1 \ldots x_{L-1}$ such that:*
$$\Pr(\overrightarrow{X}^L = w | \mathcal{S}_0 = \sigma_k) \neq \Pr(\overrightarrow{X}^L = w | \mathcal{S}_0 = \sigma_j).$$

The hidden Markov machines given above for the Even Process and ABC Process are both ϵ-machines. The SNS machine of example 3 is not an ϵ-machine, however, since state σ_1 is not unifilar. The NP2 machine of example 4 is also not an ϵ-machine, since it does not have probabilistically distinct states, as noted before.

ϵ-Machines were originally defined in Ref. [6] as hidden Markov machines whose states, known as *causal states*, were the equivalence classes of infinite pasts $\overleftarrow{x}$ with the same probability distribution over futures $\overrightarrow{x}$. This "history" ϵ-machine definition is, in fact, equivalent to the "generating" ϵ-machine definition presented above in the finite-state case. Although, this is not immediately apparent. Formally, it follows from the synchronization results established here and in Ref. [7].

It can also be shown that an ϵ-machine M for a given process $\mathcal{P}$ is unique up to isomorphism [6]. That is, there cannot be two different finite-state edge-label hidden Markov machines with unifilar transitions and probabilistically distinct states that both generate the same process $\mathcal{P}$. Furthermore, ϵ-machines are minimal unifilar generators in the sense that any other unifilar machine M' generating the same process $\mathcal{P}$ as an ϵ-machine M will have more states than M. Note that uniqueness does not hold if we remove either condition 1 or 2 in Def. 5.

E. Synchronization

Assume now that an observer has a correct model M (ϵ-machine) for a process $\mathcal{P}$, but is not able to directly observe M's hidden internal state. Rather, the observer must infer the internal state by observing the output data that M generates.

For a word w of length L generated by M let $\phi(w) = \Pr(\boldsymbol{\mathcal{S}}|w)$ be the observer's *belief distribution* as to the current state of the machine after observing w. That is,

$$\begin{aligned} \phi(w)_k &= \Pr(\mathcal{S}_L = \sigma_k | \overrightarrow{X}^L = w) \\ &\equiv \Pr(\mathcal{S}_L = \sigma_k | \overrightarrow{X}^L = w, \mathcal{S}_0 \sim \pi) \ . \end{aligned}$$

And, define:

$$\begin{aligned} u(w) &= H[\phi(w)] \\ &= H[\mathcal{S}_L | \overrightarrow{X}^L = w], \end{aligned}$$

as the observer's uncertainty in the machine state after observing w.

Denote $\mathcal{L}(M)$ as the set of all finite words that M can generate, $\mathcal{L}_L(M)$ as the set of all length-L words it can generate, and $\mathcal{L}_\infty(M)$ as the set of all infinite sequences $\overrightarrow{x} = x_0 x_1 ...$ that it can generate.

Definition 6. *A word $w \in \mathcal{L}(M)$ is a* synchronizing word *(or* sync word*) for M if $u(w) = 0$; that is, if the observer knows the current state of the machine with certainty after observing w.*

We denote the set of M's infinite synchronizing sequences as $\mathrm{SYN}(M)$ and the set of M's infinite weakly synchronizing sequences as $\mathrm{WSYN}(M)$:

$$\begin{aligned} &\mathrm{SYN}(M) = \{\overrightarrow{x} \in \mathcal{L}_\infty(M) : u(\overrightarrow{x}^L) = 0 \text{ for some } L\}, \ \textit{and} \\ &\mathrm{WSYN}(M) = \{\overrightarrow{x} \in \mathcal{L}_\infty(M) : u(\overrightarrow{x}^L) \to 0 \text{ as } L \to \infty\} \ . \end{aligned}$$

Definition 7. *An ϵ-machine M is* exactly synchronizable *(or simply* exact*) if $\Pr(\mathrm{SYN}(M)) = 1$; that is, if the observer synchronizes to almost every (a.e.) sequence generated by the machine in finite time.*

Definition 8. *An ϵ-machine M is* asymptotically synchronizable *if $\Pr(\mathrm{WSYN}(M)) = 1$; that is, if the observer's uncertainty in the machine state vanishes asymptotically for a.e. sequence generated by the machine.*

The Even Process ϵ-machine, Fig. 1, is an exact machine. Any word containing a 0 is a sync word for this machine, and almost every $\overrightarrow{x}$ it generates contains at least one 0. The ABC Process ϵ-machine, Fig. 2, is not exactly synchronizable, but it is asymptotically synchronizable.

Remark. *If $w \in \mathcal{L}(M)$ is a sync word, then by unifilarity so is wv, for all v with $wv \in \mathcal{L}(M)$. Once an observer synchronizes exactly, it remains synchronized exactly for all future times. It follows that any exactly synchronizable machine is also asymptotically synchronizable.*

Remark. *If $w \in \mathcal{L}(M)$ is a sync word then so is vw, for all v with $vw \in \mathcal{L}(M)$. Since any finite word $w \in \mathcal{L}(M)$ will be contained in almost every infinite sequence $\overrightarrow{x}$ the machine generates, it follows that a machine is exactly synchronizable if (and only if) it has some sync word w of finite length.*

Remark. *It turns out all finite-state ϵ-machines are asymptotically synchronizable; see Ref. [7]. Hence, there are two disjoint classes to consider:* exactly synchronizable machines *and* asymptotically synchronizable machines that are nonexact*. The exact case is the subject of the remainder.*

Finally, one last important quantity for synchronization is the observer's average uncertainty in the machine state after seeing a length-L block of output [8].

Definition 9. *The observer's* average state uncertainty *at time L is:*

$$\begin{aligned} \mathcal{U}(L) &\equiv H[\mathcal{S}_L | \overrightarrow{X}^L] \\ &= \sum_{\{\overrightarrow{x}^L\}} \Pr(\overrightarrow{x}^L) \cdot H[\mathcal{S}_L | \overrightarrow{X}^L = \overrightarrow{x}^L] \ . \end{aligned} \quad (5)$$

That is, $\mathcal{U}(L)$ is the expected value of an observer's uncertainty in the machine state after observing L symbols.

Now, for an ϵ-machine, an observer's prediction of the next output symbol is a direct function of the probability distribution over machine states induced by the previously observed symbols. Specifically,

$$\begin{aligned} &\Pr(X_L = x | \overrightarrow{X}^L = \overrightarrow{x}^L) \\ &\quad = \sum_{\{\sigma_k\}} \Pr(x|\sigma_k) \cdot \Pr(\mathcal{S}_L = \sigma_k | \overrightarrow{X}^L = \overrightarrow{x}^L) \ . \end{aligned} \quad (6)$$

Hence, the better an observer knows the machine state at the current time, the better it can predict the next symbol generated. And, on average, the closer $\mathcal{U}(L)$ is to 0, the closer $h_\mu(L)$ is to h_μ. Therefore, the rate of convergence of $h_\mu(L)$ to h_μ for an ϵ-machine is closely related to the average rate of synchronization.

III. EXACT SYNCHRONIZATION RESULTS

This section provides our main results on synchronization rates for exact machines and draws out consequences for the convergence rates of $\mathcal{U}(L)$ and $h_\mu(L)$.

The following notation will be used throughout:

- $\mathrm{SYN}_L = \{w \in \mathcal{L}_L(M) : w$ is a sync word for $M\}$.
- $\mathrm{NSYN}_L = \{w \in \mathcal{L}_L(M) : w$ is *not* a sync word for $M\}$.
- $\mathrm{SYN}_{L,\sigma_k} = \{w \in \mathcal{L}_L(M) : w$ synchronizes the observer to state $\sigma_k\}$.
- $\mathcal{L}(M, \sigma_k) = \{w : M$ can generate w starting in state $\sigma_k\}$.
- For words $w, w' \in \mathcal{L}(M)$, we say $w \subset w'$ if there exist words u, v (of length ≥ 0) such that $w' = uwv$.
- For a word $w \in \mathcal{L}(M, \sigma_k)$, $\delta(\sigma_k, w)$ is defined to be the (unique) state in $\boldsymbol{\mathcal{S}}$ such that $\sigma_k \xrightarrow{w} \delta(\sigma_k, w)$.
- For a set of states $S \subset \boldsymbol{\mathcal{S}}$, we define:

$$\delta(S, w) = \{\sigma_j \in \boldsymbol{\mathcal{S}} : \sigma_k \xrightarrow{w} \sigma_j \text{ for some } \sigma_k \in S\} \ .$$

A. Exact Machine Synchronization Theorem

Our first theorem states that an observer synchronizes (exactly) to the internal state of any exact ϵ-machine exponentially fast.

Theorem 1. *For any exact ϵ-machine M, there are constants $K > 0$ and $0 < \alpha < 1$ such that:*

$$\Pr(\text{NSYN}_L) \leq K\alpha^L \ , \tag{7}$$

for all $L \in \mathbb{N}$.

Proof. Let M be an exact machine with sync word $w \in \mathcal{L}(M, \sigma_k)$. Since the graph of M is strongly connected, we know that for each state σ_j there is a word v_j such that $\delta(\sigma_j, v_j) = \sigma_k$. Let $w_j = v_j w$, $n = \max_j |w_j|$, and $p = \min_j \Pr(w_j|\sigma_j)$. Then, for all $L \geq 0$, we have:

$$\begin{aligned}
\Pr(w \subset & \overrightarrow{X}^{n+L} | w \not\subset \overrightarrow{X}^L) \\
&\geq \Pr(w \subset \overrightarrow{X}_L^n | w \not\subset \overrightarrow{X}^L) \\
&\geq \min_j \Pr(w \subset \overrightarrow{X}_L^n | \mathcal{S}_L = \sigma_j) \\
&\geq p \ .
\end{aligned} \tag{8}$$

Hence,

$$\Pr(w \not\subset \overrightarrow{X}^{n+L} | w \not\subset \overrightarrow{X}^L) \leq 1 - p \ , \tag{9}$$

for all $L \geq 0$. And, therefore, for all $m \in \mathbb{N}$:

$$\begin{aligned}
\Pr(\text{NSYN}_{mn}) &\leq Pr(w \not\subset \overrightarrow{X}^{mn}) \\
&= \Pr(w \not\subset \overrightarrow{X}^n) \cdot \Pr(W \not\subset \overrightarrow{X}^{2n} | w \not\subset \overrightarrow{X}^n) \\
&\quad \cdots \Pr(w \not\subset \overrightarrow{X}^{mn} | w \not\subset \overrightarrow{X}^{(m-1)n}) \\
&\leq (1-p) \cdot (1-p) \cdots (1-p) \\
&= (1-p)^m \\
&= \beta^m \ ,
\end{aligned} \tag{10}$$

where $\beta \equiv 1 - p$. Or equivalently, for any length $L = mn$ ($m \in \mathbb{N}$):

$$\Pr(NSYN_L) \leq \alpha^L \ , \tag{11}$$

where $\alpha \equiv \beta^{1/n}$. Since $\Pr(\text{NSYN}_L)$ is monotonically decreasing, it follows that:

$$\Pr(\text{NSYN}_L) \leq \frac{1}{\alpha^n} \cdot \alpha^L = K\alpha^L \ , \tag{12}$$

for all $L \in \mathbb{N}$, where $K \equiv 1/\alpha^n$. □

Remark. *In the above proof we implicitly assumed $\beta \neq 0$. If $\beta = 0$, then the conclusion follows trivially.*

B. Synchronization Rate

Theorem 1 states that an observer synchronizes (exactly) to any exact ϵ-machine exponentially fast. However, the sync rate constant:

$$\alpha^* = \lim_{L \to \infty} \Pr(\text{NSYN}_L)^{1/L} \tag{13}$$

depends on the machine, and it may often be of practical interest to know the value of this constant. We now provide a method for computing α^* analytically. It is based on the construction of an auxiliary machine $\widetilde{M}$.

Definition 10. *Let M be an ϵ-machine with states $\boldsymbol{\mathcal{S}} = \{\sigma_1, ..., \sigma_N\}$, alphabet $\mathcal{A}$, and transition matrices $T^{(x)}, x \in \mathcal{A}$. The* possibility machine *$\widetilde{M}$ is defined as follows:*

1. *The alphabet of $\widetilde{M}$ is $\mathcal{A}$.*
2. *The states of $\widetilde{M}$ are pairs of the form (σ, S) where $\sigma \in \boldsymbol{\mathcal{S}}$ and S is a subset of $\boldsymbol{\mathcal{S}}$ that contains σ.*
3. *The transition probabilities are:*

$$\begin{aligned}
\Pr((\sigma, S) &\xrightarrow{x} (\sigma', S')) \\
&= \Pr(x|\sigma) I(x, (\sigma, S), (\sigma', S')) \ ,
\end{aligned}$$

where $I(x, (\sigma, S), (\sigma', S'))$ is the indicator function:

$$\begin{aligned}
I(x, &(\sigma, S), (\sigma', S')) \\
&= \begin{cases} 1 & \text{if } \delta(\sigma, x) = \sigma' \text{ and } \delta(S, x) = S' \\ 0 & \text{otherwise.} \end{cases}
\end{aligned}$$

A state of $\widetilde{M}$ is said to be initial *if it is of the form $(\sigma, \boldsymbol{\mathcal{S}})$ for some $\sigma \in \boldsymbol{\mathcal{S}}$. For simplicity we restrict the $\widetilde{M}$ machine to consist of only those states that are accessible from initial states. The other states are irrelevant for the analysis below.*

The idea is that $\widetilde{M}$'s states represent states of the joint (ϵ-machine, observer) system. State σ is the true ϵ-machine state at the current time, and S is the set of states that the observer believes are currently possible for the ϵ-machine to be in, after observing all previous symbols. Initially, all states are possible (to the observer), so initial states are those in which the set of possible states is the complete set $\boldsymbol{\mathcal{S}}$.

If the current true ϵ-machine state is σ, and then the symbol x is generated, the new true ϵ-machine state must be $\delta(\sigma, x)$. Similarly, if the observer believes any of the states in S are currently possible, and then the symbol x is generated, the new set of possible states to the observer is $\delta(S, x)$. This accounts for the transitions in $\widetilde{M}$ topologically. The probability of generating a given symbol x from (σ, S) is, of course, governed only by the true state σ of the ϵ-machine $\Pr(x|(\sigma, S)) = \Pr(x|\sigma)$.

An example of this construction for a 3-state exact ϵ-machine is given in Appendix A. Note that the graph of the $\widetilde{M}$ machine there has a single recurrent strongly connected component, which is isomorphic to the original machine M. This is not an accident. It will always be the case, as long as the original machine M is exact.

Remark. *If M is an exact machine with more than 1 state the graph of $\widetilde{M}$ itself is never strongly connected. So, $\widetilde{M}$ is not an ϵ-machine or even an HMM in the sense of Def. 4. However, we still refer to $\widetilde{M}$ as a "machine".*

In what follows, we assume M is exact. We denote the states of $\widetilde{M}$ as $\widetilde{\boldsymbol{\mathcal{S}}} = \{q_1, \ldots, q_{\widetilde{N}}\}$, its symbol-labeled transition matrices $\widetilde{T}^{(x)}$, and its overall state-to-state transition matrix $\widetilde{T} = \sum_{x \in \mathcal{A}} \widetilde{T}^{(x)}$. We assume the states are ordered in such a way that the initial states $(\sigma_1, \boldsymbol{\mathcal{S}}), \ldots, (\sigma_N, \boldsymbol{\mathcal{S}})$ are, respectively, $q_1, ..., q_N$. Similarly, the *recurrent states* $(\sigma_1, \{\sigma_1\}), (\sigma_2, \{\sigma_2\}), \ldots, (\sigma_N, \{\sigma_N\})$ are, respectively, $q_{n+1}, q_{n+2}, ..., q_{\widetilde{N}}$, where $n = \widetilde{N} - N$. The ordering of the other states is irrelevant. In this case, the matrix $\widetilde{T}$ has the following block upper-triangular form:

$$\widetilde{T} = \begin{pmatrix} B & B' \\ O & T \end{pmatrix} , \tag{14}$$

where B is a $n \times n$ matrix with nonnegative entries, B' is a $n \times N$ matrix with nonnegative entries, O is a $N \times n$ matrix of all zeros, and T is the $N \times N$ state-to-state transition matrix of the original ϵ-machine M.

Let $\widetilde{\pi} = (\pi_1, ..., \pi_N, 0, ..., 0)$ denote the length-$\widetilde{N}$ row vector whose distribution over the initial states is the same as the stationary distribution π for the ϵ-machine M. Then, the initial probability distribution $\widetilde{\phi}_0$ over states of the joint (ϵ-machine, observer) system is simply:

$$\widetilde{\phi}_0 = \widetilde{\pi} , \tag{15}$$

and, thus, the distribution over states of the joint system after the first L symbols is:

$$\widetilde{\phi}_L = \widetilde{\pi} \widetilde{T}^L . \tag{16}$$

If the joint system is in a recurrent state of the form $(\sigma_k, \{\sigma_k\})$, then to the observer the only possible state of the ϵ-machine is the true state, so the observer is synchronized. For all other states of $\widetilde{M}$, the observer is not yet synchronized. Hence, the probability the observer is not synchronized after L symbols is simply the combined probability of all nonreccurent states q_i in the distribution $\widetilde{\phi}_L$. Specifically, we have:

$$\begin{aligned} \Pr(\mathrm{NSYN}_L) &= \sum_{i=1}^{n} (\widetilde{\phi}_L)_i \\ &= \sum_{i=1}^{n} (\widetilde{\pi} \widetilde{T}^L)_i \\ &= \sum_{i=1}^{n} (\pi^B B^L)_i \\ &= \|\pi^B B^L\|_1 , \end{aligned} \tag{17}$$

where $\pi^B = (\pi_1, ..., \pi_N, 0, ..., 0)$ is the length-n row vector corresponding to the distribution over initial states π. The third equality follows from the block upper-triangular form of $\widetilde{T}$.

Appendix B shows that:

$$\lim_{L \to \infty} \left(\|\pi^B B^L\|_1 \right)^{1/L} = r , \tag{18}$$

where $r = r(B)$ is the (left) spectral radius of B:

$$r(B) = \max\{|\lambda| : \lambda \text{ is a (left) eigenvalue of } B\}. \tag{19}$$

Thus, we have established the following result.

Theorem 2. *For any exact ϵ-machine M, $\alpha^* = r$.*

C. Consequences

We now apply Thm. 1 to show that an observer's average uncertainty $\mathcal{U}(L)$ in the machine state and average uncertainty $h_\mu(L)$ in predictions of future symbols both decay exponentially fast to their respective limits: 0 and h_μ. The decay constant α in both cases is essentially bounded by the sync rate constant α^* from Thm. 2.

Proposition 1. *For any exact ϵ-machine M, there are constants $K > 0$ and $0 < \alpha < 1$ such that:*

$$\mathcal{U}(L) \leq K\alpha^L, \text{ for all } L \in \mathbb{N}. \tag{20}$$

Proof. Let M be any exact machine. By Thm. 1 there are constants $C > 0$ and $0 < \alpha < 1$ such that $\Pr(\mathrm{NSYN}_L) \leq C\alpha^L$, for all $L \in \mathbb{N}$. Thus, we have:

$$\begin{aligned} \mathcal{U}(L) &= \sum_{w \in \mathcal{L}_L(M)} \Pr(w) u(w) \\ &= \sum_{w \in \mathrm{SYN}_L} \Pr(w) u(w) + \sum_{w \in \mathrm{NSYN}_L} \Pr(w) u(w) \\ &\leq 0 + \sum_{w \in \mathrm{NSYN}_L} \Pr(w) \log(N) \\ &\leq \log(N) \cdot C\alpha^L \\ &= K\alpha^L , \end{aligned} \tag{21}$$

where N is the number of machine states and $K \equiv C\log(N)$.

□

Let $h_k \equiv H[X_0|\mathcal{S}_0 = \sigma_k]$ and $h_w \equiv H[X_0|\mathcal{S}_0 \sim \phi(w)]$, be the conditional entropies in the next symbol given the state σ_k and word w.

Proposition 2. *For any exact ϵ-machine M:*

$$h_\mu = H[X_0|\mathcal{S}_0] \equiv \sum_k \pi_k h_k \tag{22}$$

and there are constants $K > 0$ and $0 < \alpha < 1$ such that:

$$h_\mu(L) - h_\mu \leq K\alpha^L \ , \text{ for all } L \in \mathbb{N}. \tag{23}$$

Remark. *The h_μ formula Eq. (22) has been known for some time, although in slightly different contexts. Shannon, for example, derived this formula in his original publication [9] for a type of hidden Markov machine that is similar (apparently unifilar) to an ϵ-machine.*

Proof. Let M be any exact machine. Since we know $h_\mu(L) \searrow h_\mu$ it suffices to show there are constants $K > 0$ and $0 < \alpha < 1$ such that:

$$\left| h_\mu(L) - \sum_k \pi_k h_k \right| \leq K\alpha^L \ , \tag{24}$$

for all $L \in \mathbb{N}$. This will establish both the value of h_μ and the necessary convergence.

Now, by Thm. 1, there are constants $C > 0$ and $0 < \alpha < 1$ such that $\Pr(\text{NSYN}_L) \leq C\alpha^L$, for all $L \in \mathbb{N}$. Also, note that for all L and k we have:

$$\begin{aligned} \pi_k &= \sum_{w \in \mathcal{L}_L(M)} \Pr(w) \cdot \phi(w)_k \\ &\geq \sum_{w \in \text{SYN}_{L,\sigma_k}} \Pr(w) \cdot \phi(w)_k \\ &= \Pr(\text{SYN}_{L,\sigma_k}) \ . \end{aligned} \tag{25}$$

Thus,

$$\sum_k \left(\pi_k - \Pr(\text{SYN}_{L,\sigma_k})\right) \cdot h_k \geq 0 \tag{26}$$

and

$$\begin{aligned} &\sum_k (\pi_k - \Pr(\text{SYN}_{L,\sigma_k})) \cdot h_k \\ &\quad \leq \sum_k (\pi_k - \Pr(\text{SYN}_{L,\sigma_k})) \cdot \log|\mathcal{A}| \\ &\quad = \log|\mathcal{A}| \cdot \left(\sum_k \pi_k - \sum_k \Pr(\text{SYN}_{L,\sigma_k}) \right) \\ &\quad = \log|\mathcal{A}| \cdot (1 - \Pr(\text{SYN}_L)) \\ &\quad = \log|\mathcal{A}| \cdot \Pr(\text{NSYN}_L) \\ &\quad \leq \log|\mathcal{A}| \cdot C\alpha^L. \end{aligned} \tag{27}$$

Also, clearly,

$$\sum_{w \in \text{NSYN}_L} \Pr(w) \cdot h_w \geq 0 \tag{28}$$

and

$$\begin{aligned} \sum_{w \in \text{NSYN}_L} \Pr(w) \cdot h_w &\leq \log|\mathcal{A}| \cdot \Pr(\text{NSYN}_L) \\ &\leq \log|\mathcal{A}| \cdot C\alpha^L \ . \end{aligned} \tag{29}$$

Therefore, we have for all $L \in \mathbb{N}$:

$$\begin{aligned} &\left| h_\mu(L+1) - \sum_k \pi_k h_k \right| \\ &= \left| \sum_{w \in \mathcal{L}_L(M)} \Pr(w) h_w - \sum_k \pi_k h_k \right| \\ &= \left| \sum_{w \in \text{NSYN}_L} \Pr(w) h_w + \sum_{w \in SYN_L} \Pr(w) h_w - \sum_k \pi_k h_k \right| \\ &= \left| \sum_{w \in \text{NSYN}_L} \Pr(w) h_w + \sum_k \Pr(\text{SYN}_{L,\sigma_k}) h_k - \sum_k \pi_k h_k \right| \\ &= \left| \sum_{w \in \text{NSYN}_L} \Pr(w) h_w - \sum_k (\pi_k - \Pr(\text{SYN}_{L,\sigma_k})) h_k \right| \\ &\leq C \log|\mathcal{A}| \alpha^L \ . \end{aligned} \tag{30}$$

The last inequality follows from Eqs. (26)-(29), since $|x - y| \leq z$ for all nonnegative real numbers x, y, and z with $x \leq z$ and $y \leq z$.

Finally, since:

$$\left| h_\mu(L+1) - \sum_k \pi_k h_k \right| \leq C \log|\mathcal{A}| \alpha^L \ , \tag{31}$$

for all $L \in \mathbb{N}$, we know that:

$$\left| h_\mu(L) - \sum_k \pi_k h_k \right| \leq K\alpha^L \ , \tag{32}$$

for all $L \in \mathbb{N}$, where $K \equiv (\log|\mathcal{A}|/\alpha) \cdot \max\{C, 1\}$. □

Remark. *For any $\alpha > \alpha^*$ there exists some $K > 0$ for which Eq. (7) holds. Hence, by the constructive proofs above, we see that the constant α in Props. 1 and 2 can be chosen arbitrarily close to α^*: $\alpha = \alpha^* + \epsilon$.*

IV. CHARACTERIZATION OF EXACT ϵ-MACHINES

In this section we provide a set of necessary and sufficient conditions for exactness and an algorithmic test for exactness based upon these conditions.

A. Exact Machine Characterization Theorem

Definition 11. *States σ_k and σ_j are said to be* topologically distinct *if $\mathcal{L}(M, \sigma_k) \neq \mathcal{L}(M, \sigma_j)$.*

Definition 12. *States σ_k and σ_j are said to be* path convergent *if there exists $w \in \mathcal{L}(M,\sigma_k) \cap \mathcal{L}(M,\sigma_j)$ such that $\delta(\sigma_k, w) = \delta(\sigma_j, w)$.*

If states σ_k and σ_j are topologically distinct (or path convergent) we will also say the pair (σ_k, σ_j) is topologically distinct (or path convergent).

Theorem 3. *An ϵ-machine M is exact if and only if every pair of distinct states (σ_k, σ_j) satisfies at least one of the following two conditions:*

(i) The pair (σ_k, σ_j) is topologically distinct.

(ii) The pair (σ_k, σ_j) is path convergent.

Proof. It was noted above that an ϵ-machine M is exact if and only if it has some sync word w of finite length. Therefore, it is enough to show that every pair of distinct states (σ_k, σ_j) satisfies either (i) or (ii) if and only if M has some sync word w of finite length.

We establish the "if" first: *If M has a sync word w, then every pair of distinct states (σ_k, σ_j) satisfies either (i) or (ii).*

Let w be a sync word for M. Then $w \in \mathcal{L}(M, \sigma_k)$ for some k. Take words v_j, $j = 1, 2, ...N$, such that $\delta(\sigma_j, v_j) = \sigma_k$. Then, the word $v_j w \equiv w_j \in \mathcal{L}(M, \sigma_j)$ is also a sync word for M for each j. Therefore, for each $i \neq j$ either $w_j \notin \mathcal{L}(M, \sigma_i)$ or $\delta(\sigma_i, w_j) = \delta(\sigma_j, w_j)$. This establishes that the pair (σ_i, σ_j) is either topologically distinct or path convergent. Since this holds for all $j = 1, 2, ..., N$ and for all $i \neq j$, we know every pair of distinct states is either topologically distinct or path convergent.

Now, for the "only if" case: *If every pair of distinct states (σ_k, σ_j) satisfies either (i) or (ii), then M has a sync word w.*

If each pair of distinct states (σ_k, σ_j) satisfies either (i) or (ii), then for all k and j ($k \neq j$) there is some word w_{σ_k,σ_j} such that one of the following three conditions is satisfied:

1. $w_{\sigma_k,\sigma_j} \in \mathcal{L}(M, \sigma_k)$, but $w_{\sigma_k,\sigma_j} \notin \mathcal{L}(M, \sigma_j)$.
2. $w_{\sigma_k,\sigma_j} \in \mathcal{L}(M, \sigma_j)$, but $w_{\sigma_k,\sigma_j} \notin \mathcal{L}(M, \sigma_k)$.
3. $w_{\sigma_k,\sigma_j} \in \mathcal{L}(M, \sigma_k) \cap \mathcal{L}(M, \sigma_j)$ and $\delta(\sigma_k, w_{\sigma_k,\sigma_j}) = \delta(\sigma_j, w_{\sigma_k,\sigma_j})$.

We construct a sync word $w = w_1 w_2 ... w_m$ for M, where each $w_i = w_{\sigma_{k_i},\sigma_{j_i}}$ for some k_i and j_i, as follows.

- Let $\boldsymbol{\mathcal{S}}^0 = \{\sigma_1^0, \ldots, \sigma_{N_0}^0\} \equiv \boldsymbol{\mathcal{S}} = \{\sigma_1, ..., \sigma_N\}$. Take $w_1 = w_{\sigma_1^0,\sigma_2^0}$.
- Let $\boldsymbol{\mathcal{S}}^1 = \{\sigma_1^1, \ldots, \sigma_{N_1}^1\} \equiv \delta(\boldsymbol{\mathcal{S}}^0, w_1)$. Since $w_1 = w_{\sigma_1^0,\sigma_2^0}$ satisfies either condition (1), (2), or (3), we know $N_1 < N_0$. Take $w_2 = w_{\sigma_1^1,\sigma_2^1}$.
- Let $\boldsymbol{\mathcal{S}}^2 = \{\sigma_1^2, \ldots, \sigma_{N_2}^2\} \equiv \delta(\boldsymbol{\mathcal{S}}^1, w_2)$. Since $w_2 = w_{\sigma_1^1,\sigma_2^1}$ satisfies either condition (1), (2), or (3) we know $N_2 < N_1$. Take $w_3 = w_{\sigma_1^2,\sigma_2^2}$.

.
.
.

Repeat until $|\boldsymbol{\mathcal{S}}^m| = 1$ for some m. Note that this must happen after a finite number of steps since $N = N_0$ is finite and $N_0 > N_1 > N_2 > \cdots$.

By this construction $w = w_1 w_2 ... w_m \in \mathcal{L}(M)$ is a sync word for M. After observing w, an observer knows the machine must be in state σ_1^m. $\square$

B. A Test for Exactness

We can now provide an algorithmic test for exactness using the characterization theorem of exact machines. We begin with subalgorithms to test for topological distinctness and path convergence of state pairs. Both are essentially the same algorithm and only a slight modification of the deterministic finite-automata (DFA) table-filling algorithm to test for pairs of equivalent states [10].

Algorithm 1. *Test States for Topological Distinctness.*

1. Initialization*: Create a table containing boxes for all pairs of distinct states (σ_k, σ_j). Initially, all boxes are blank. Then,*

   Loop over distinct pairs (σ_k, σ_j)
   Loop over $x \in \mathcal{A}$
   If $\{x \in \mathcal{L}(M, \sigma_k)$ but $x \notin \mathcal{L}(M, \sigma_j)\}$
   or $\{x \in \mathcal{L}(M, \sigma_j)$ but $x \notin \mathcal{L}(M, \sigma_k)\}$,
   then mark box for pair (σ_k, σ_j).
   end
   end

2. Induction*: If $\delta(\sigma_k, x) = \sigma_{k'}$, $\delta(\sigma_j, x) = \sigma_{j'}$, and the box for pair $(\sigma_{k'}, \sigma_{j'})$ is already marked, then mark the box for pair (σ_k, σ_j). Repeat until no more inductions are possible.*

Algorithm 2. *Test States for Path Convergence.*

This algorithm is identical to Algorithm 1 except that the if-statement in the initialization step is replaced with the following:

If $x \in \mathcal{L}(M, \sigma_k) \cap \mathcal{L}(M, \sigma_j)$ and $\delta(\sigma_k, x) = \delta(\sigma_j, x)$, then mark box for pair (σ_k, σ_j).

With Algorithm 1 all pairs of topologically distinct states end up with marked boxes. With Algorithm 2 all pairs of path convergent states end up with marked boxes. These facts can be proved, respectively, by using induction on the length of the minimal distinguishing or path converging word w for a given pair of states. The proofs are virtually identical to the proof of the standard DFA table filling algorithm, so the details have been omitted.

Note also that both of these are polynomial-time algorithms. Step (1) has run time $O(|\mathcal{A}|N^2)$. The inductions in Step (2), if done in a reasonably efficient fashion, can also be completed in run time $O(|\mathcal{A}|N^2)$. (See, e.g., the analysis of DFA table filling algorithm in Ref. [10].) Therefore, the total run time of these algorithm is $O(|\mathcal{A}|N^2)$.

Algorithm 3. *Test for Exactness.*

1. *Use Algorithm 1 to find all pairs of topologically distinct states.*
2. *Use Algorithm 2 to find all pairs of path convergent states.*
3. *Loop over all pairs of distinct states (σ_k, σ_j) to check if they are either (i) topologically distinct or (ii) path convergent. By Thm. 3, if all distinct pairs of states satisfy (i) or (ii) or both, the machine is exact, and otherwise it is not.*

This, too, is a polynomial-time algorithm. Steps (1) and (2) have run time $O(|\mathcal{A}|N^2)$. Step (3) has run time $O(N^2)$. Hence, the total run time for this algorithm is $O(|\mathcal{A}|N^2)$.

V. CONCLUSION

We have analyzed the process of exact synchronization to finite-state ϵ-machines. In particular, we showed that for exact machines an observer synchronizes exponentially fast. As a result, the average uncertainty $h_\mu(L)$ in an observer's predictions converges exponentially fast to the machine's entropy rate h_μ—a phenomenon first reported for subshifts estimated from maps of the interval [11]. Additionally, we found an efficient (polynomial-time) algorithm to test ϵ-machines for exactness.

In Ref. [7] we similarly analyze asymptotic synchronization to nonexact ϵ-machines. It turns out that qualitatively similar results hold. That is, $\mathcal{U}(L)$ and $h_\mu(L)$ both converge to their respective limits exponentially fast. However, the proof methods in the nonexact case are substantially different.

In the future we plan to extend these results to more generalized model classes, such as to ϵ-machines with a countable number of states and to nonunifilar hidden Markov machines.

Acknowledgments

NT was partially supported on a VIGRE fellowship. The work was partially supported by the Defense Advanced Research Projects Agency (DARPA) Physical Intelligence project via subcontract No. 9060-000709. The views, opinions, and findings here are those of the authors and should not be interpreted as representing the official views or policies, either expressed or implied, of the DARPA or the Department of Defense.

Appendix A

We construct the possibility machine $\widetilde{M}$ for the three-state ϵ-machine shown in Fig. 5. The result is shown in Fig. 6.

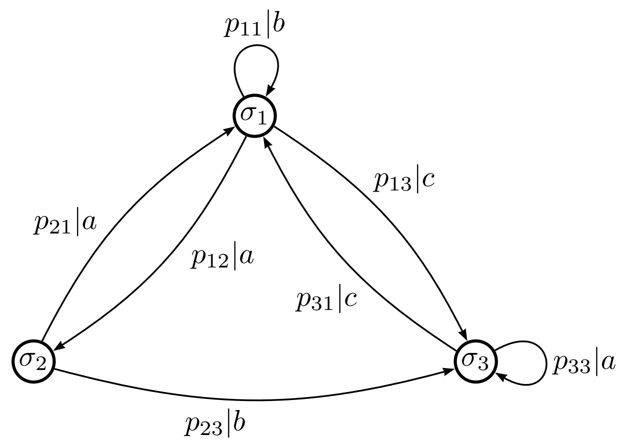

FIG. 5: A three-state ϵ-machine M with alphabet $\mathcal{A} = \{a, b, c\}$.

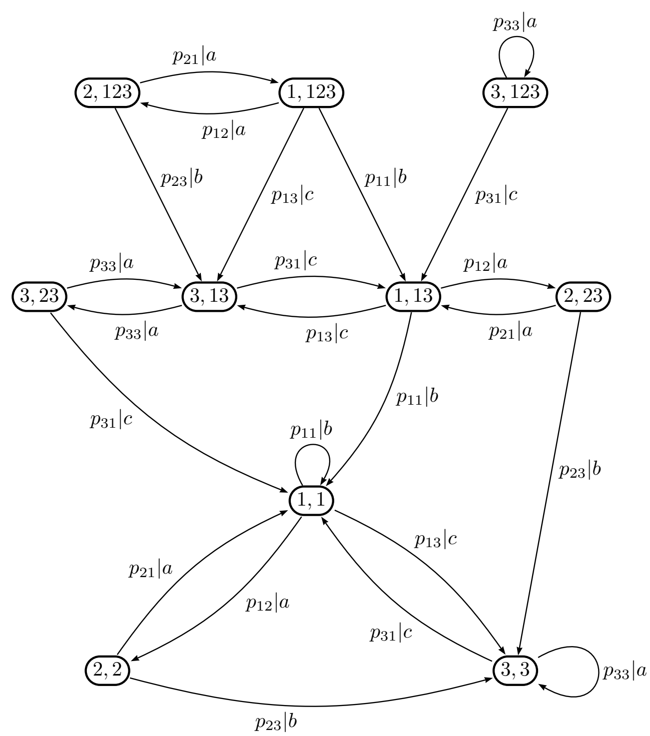

FIG. 6: The possibility machine $\widetilde{M}$ for the three-state ϵ-machine M of Fig. 5. The state names have been abbreviated for display purposes: e.g., $(\sigma_1, \{\sigma_1, \sigma_2, \sigma_3\}) \to (1, 123)$.

Appendix B

We prove Eq. (18) in Sec. III B. (Restated here as Lemma 1.)

Lemma 1. *For any exact ϵ-machine M,*

$$\lim_{L\to\infty} \|\pi^B B^L\|_1^{1/L} = r(B) \ . \tag{B1}$$

In what follows A denotes an arbitrary $m \times m$ matrix and $\overrightarrow{v}$ and $\overrightarrow{w}$ denote row m-vectors. Unless otherwise specified, the entries of matrices and vectors are assumed to be complex.

Definition 13. *The* (left) matrix p-norms ($1 \leq p \leq \infty$) *are defined as:*

$$\|A\|_p = \max\{\|\overrightarrow{v}A\|_p : \|\overrightarrow{v}\|_p = 1\} \ . \tag{B2}$$

The following facts will be used in our proof.

Fact 1. *If A is a matrix with real nonnegative entries and $\overrightarrow{v} = (v_1, \ldots, v_m)$ is a vector with real nonnegative entries, then:*

$$\|\overrightarrow{v}A\|_1 = \sum_{k=1}^{m} \|(v_k \overrightarrow{e}_k)A\|_1 \ , \tag{B3}$$

where $\overrightarrow{e}_k = (0, \ldots, 1, \ldots, 0)$ is the k_{th} standard basis vector.

Fact 2. *Let A be a matrix with real nonnegative entries, let $\overrightarrow{v} = (v_1, \ldots, v_m)$ be a vector with complex entries, and let $\overrightarrow{w} = (w_1, \ldots, w_m) = (|v_1|, \ldots, |v_m|)$. Then:*

$$\|\overrightarrow{v}A\|_1 \leq \|\overrightarrow{w}A\|_1 \ . \tag{B4}$$

Fact 3. *For any matrix $A = \{a_{ij}\}$, the matrix 1-norm is the largest absolute row sum:*

$$\|A\|_1 = \max_i \sum_{j=1}^{m} |a_{ij}| \ . \tag{B5}$$

Fact 4. *For any matrix A, $L \in \mathbb{N}$, and $1 \leq p \leq \infty$:*

$$\|A^L\|_p \leq \|A\|_p^L \ . \tag{B6}$$

Fact 5. *For any matrix A and $1 \leq p \leq \infty$:*

$$\lim_{L\to\infty} \|A^L\|_p^{1/L} = r(A) \ , \tag{B7}$$

where $r(A)$ is the (left) spectral radius of A:

$$r(A) = \max\{|\lambda| : \lambda \textit{ is a (left) eigenvalue of } A\}. \tag{B8}$$

(This is, of course, the same as the right spectral radius, but we emphasize the left eigenvalues for the proof of Lemma 1 below.)

Fact 1 can be proved by direct computation, and Fact 2 follows from the triangle inequality. Fact 3 is a standard result from linear algebra. Facts 4 and 5 are finite-dimensional versions of more general results established in Ref. [12] for bounded linear operators on Banach spaces.

Using these facts we now prove Lemma 1.

Proof. By Fact 5 we know:

$$\limsup_{L\to\infty} \|\pi^B B^L\|_1^{1/L} \leq r(B) \ . \tag{B9}$$

Thus, it suffices to show that:

$$\liminf_{L\to\infty} \|\pi^B B^L\|_1^{1/L} \geq r(B) \ . \tag{B10}$$

Let us define the *B-machine* to be the restriction of the $\widetilde{M}$ machine to its nonreccurent states. The state-to-state transition matrix for this machine is B. We call the states of this machine *B-states* and refer to paths in the associated graph as *B-paths*. Note that the rows of $B = \{b_{ij}\}$ are substochastic:

$$\sum_j b_{ij} \leq 1 \ , \tag{B11}$$

for all i, with strict inequality for at least one value of i as long as M has more than 1 state.

By the construction of the B-machine we know that for each of its states q_j there exists some initial state $q_i = q_{i(j)}$ such that q_j is accessible from $q_{i(j)}$. Define l_j to be the length of the shortest B-path from $q_{i(j)}$ to q_j, and $l_{max} = \max_j l_j$. Let $c_j > 0$ be the probability, according to the initial distribution π^B, of both starting in state $q_{i(j)}$ at time 0 and ending in state q_j at time l_j:

$$c_j = (\pi_{i(j)} \overrightarrow{e}_{i(j)} B^{l_j})_j \ .$$

Finally, let $C_1 = \min_j c_j$.

Then, for any $L > l_{max}$ and any state q_j we have:

$$\begin{aligned} \|\pi^B B^L\|_1 &\geq \|\pi_{i(j)} \overrightarrow{e}_{i(j)} B^L\|_1 && \text{(B12)} \\ &= \|(\pi_{i(j)} \overrightarrow{e}_{i(j)} B^{l_j}) B^{L-l_j}\|_1 && \text{(B13)} \\ &\geq \|c_j \overrightarrow{e}_j B^{L-l_j}\|_1 && \text{(B14)} \\ &\geq C_1 \|\overrightarrow{e}_j B^{L-l_j}\|_1 && \text{(B15)} \\ &\geq C_1 \|\overrightarrow{e}_j B^L\|_1 \ . && \text{(B16)} \end{aligned}$$

Equation (B12) follows from Fact 1. The decomposition in Eq. (B13) is possible since $L > l_{max} \geq l_j$. Equation (B14) follows from Fact 1 and the definition of c_j. Equation (B15) follows from the definition of C_1. Finally, Eq. (B16) follows from Fact 3, Fact 4, and Eq. (B11).

Now, take a normalized (left) eigenvector $\overrightarrow{y} = (y_1, ..., y_n)$ of B whose associated eigenvalue is maximal. That is, $\|\overrightarrow{y}\|_1 = 1$, $\overrightarrow{y}B = \lambda\overrightarrow{y}$, and $|\lambda| = r(B)$. Define

$\overrightarrow{z} = (z_1, ..., z_n) = (|y_1|, ..., |y_n|)$. Then, for any $L \in \mathbb{N}$:

$$\begin{aligned} \sum_{k=1}^{n} z_k \|\overrightarrow{e}_k B^L\|_1 &= \|\overrightarrow{z} B^L\|_1 && \text{(B17)} \\ &\geq \|\overrightarrow{y} B^L\|_1 && \text{(B18)} \\ &= \|\lambda^L \overrightarrow{y}\|_1 && \text{(B19)} \\ &= |\lambda|^L \cdot \|\overrightarrow{y}\|_1 && \text{(B20)} \\ &= r(B)^L \ , && \text{(B21)} \end{aligned}$$

where Eq. (B17) follows from Fact 1 and Eq. (B18) from Fact 2. Therefore, for each L we know there exists some $j = j(L)$ in $\{1, ..., n\}$ such that:

$$z_{j(L)} \|\overrightarrow{e}_{j(L)} B^L\|_1 \geq \frac{r(B)^L}{n} \ . \qquad \text{(B22)}$$

Now, $r(B)$ may be 0, but we can still choose the $j(L)'s$ such that $z_{j(L)}$ is never zero. And, in this case, we may divide through by $z_{j(L)}$ on both sides of Eq. (B22) to obtain, for each L:

$$\begin{aligned} \|\overrightarrow{e}_{j(L)} B^L\|_1 &\geq \frac{r(B)^L}{n \cdot z_{j(L)}} \\ &\geq C_2 \cdot r(B)^L, && \text{(B23)} \end{aligned}$$

where $C_2 > 0$ is defined by:

$$C_2 = \min_{z_j \neq 0} \frac{1}{n \cdot z_j} \ .$$

Therefore, for any $L > l_{max}$ we know:

$$\begin{aligned} \|\pi^B B^L\|_1 &\geq C_1 \cdot \|\overrightarrow{e}_{j(L)} B^L\|_1 && \text{(B24)} \\ &\geq C_1 \cdot \left(C_2 \cdot r(B)^L\right) && \text{(B25)} \\ &= C_3 \cdot r(B)^L \ , && \text{(B26)} \end{aligned}$$

where $C_3 \equiv C_1 C_2$. Equation (B24) follows from Eq. (B16) and Eq. (B25) follows from Eq. (B23). Finally, since this holds for all $L > l_{max}$, we have:

$$\begin{aligned} \liminf_{L \to \infty} \|\pi^B B^L\|_1^{1/L} &\geq \liminf_{L \to \infty} \left(C_3 \cdot r(B)^L\right)^{1/L} \\ &= r(B) \ . && \text{(B27)} \end{aligned}$$

$\square$

[1] G. D. Forney Jr. The Viterbi algorithm: A personal history. *CoRR*, abs/cs/0504020, 2005.

[2] A. J. Viterbi. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Trans. Info. Th.*, 13(2):260–269, 1967.

[3] N. Jonoska. Sofic shifts with synchronizing presentations. *Theor. Comput. Sci.*, 158(1-2):81–115, 1996.

[4] S. Sandberg. Homing and synchronizing sequences. In M. Broy et al, editor, *Lect. Notes Comp. Sci.*, volume 3472, pages 5–33. Springer, Berlin, 2005.

[5] A. Paz. *Introduction to Probabilistic Automata.* Academic Press, New York, 1971.

[6] J. P. Crutchfield and K. Young. Inferring statistical complexity. *Phys. Rev. Let.*, 63:105–108, 1989. J. P. Crutchfield, C. J. Ellison, J R. Mahoney, and R. G. James, Synchronization and Control in Intrinsic and Designed Computation: An Information-Theoretic Analysis of Competing Models of Stochastic Computation, CHAOS 20:3 (2010) 037105.

[7] N. Travers and J. P. Crutchfield. Asymptotic synchronization for finite-state sources. 2010. SFI Working Paper 10-11-029; arxiv.org:1011.1581 [nlin.CD].

[8] T. M. Cover and J. A. Thomas. *Elements of Information Theory.* Wiley-Interscience, New York, second edition, 2003. Extensions and notation used here are from J. P. Crutchfield and D. P. Feldman, Regularities Unseen, Randomness Observed: Levels of Entropy Convergence, CHAOS 13:1 (2003) 25–54.

[9] C. E. Shannon. A mathematical theory of communication. *Bell Sys. Tech. J.*, 27:379–423, 623–656, 1948.

[10] J. E. Hopcroft, R. Motwani, and J. D. Ullman. *Automata Theory, Languages, and Computation.* Addison-Wesley, Reading, 2007.

[11] J. P. Crutchfield and N. H. Packard. Symbolic dynamics of noisy chaos. *Physica*, 7D:201 – 223, 1983.

[12] M. Reed and B. Simon. *Functional Analysis.* Academic Press, 1980.