STOCHASTIC DYNAMICS: MARKOV CHAINS
AND RANDOM TRANSFORMATIONS

FELIX X.-F. YE, YUE WANG AND HONG QIAN

Department of Applied Mathematics
University of Washington
Seattle, WA 98195-3925, USA

Abstract. This article outlines an attempt to lay the groundwork for understanding stochastic dynamical descriptions of biological processes in terms of a discrete-state space, discrete-time random dynamical system (RDS), or random transformation approach. Such mathematics is not new for continuous systems, but the discrete state space formulation significantly reduces the technical requirements for its introduction to a much broader audience. In particular, we establish some elementary contradistinctions between Markov chain (MC) and RDS descriptions of a stochastic dynamics. It is shown that a given MC is compatible with many possible RDS, and we study in particular the corresponding RDS with maximum metric entropy. Specifically, we show an emergent behavior of an MC with a unique absorbing and aperiodic communicating class, after all the trajectories of the RDS synchronizes. In biological modeling, it is now widely acknowledged that stochastic dynamics is a more complete description of biological reality than deterministic equations; here we further suggest that the RDS description could be a more refined description of stochastic dynamics than a Markov process. Possible applications of discrete-state RDS are systems with fluctuating law of motion, or environment, rather than inherent stochastic movements of individuals.

1. Introduction. Mathematical theories of biological dynamics are usually divided into “deterministic” and “stochastic” models. Hodgkin-Huxley’s theory of membrane action potential belongs to the former, the Wright-Fisher model for random mating is one of the latter [33, 1, 11]. Characterizing the solution of a dynamic model with given initial conditions, either analytically or computationally, is one of the main tasks in mathematical biology.

The modern theory of nonlinear dynamical systems [45, 23], however, articulates a qualitative, yet global understanding of dynamics by “looking at the evolution of the whole phase space of initial conditions” [22]. Dynamical systems with discrete or continuous times are usually referred to as iterative maps or flows, respectively [15]. Phase portrait and local linear stability analysis of ordinary differential equations (ODEs) are now routines of analyzing biological dynamics. Important concepts emerge from this type of studies are the notions of attractors, phase portrait, local and global vector field bifurcation, topological equivalence and canonical forms, Lyapunov exponents, among others [33, 11]. One is interested in the simultaneous trajectories with different initial conditions.

2010 Mathematics Subject Classification. Primary: 60G10, 92B05; Secondary: 34F05. 
Key words and phrases. Stochastic process, Markov chain, random dynamical system, entropy. The last author is partially supported by NIH grant R01GM109964.
Most stochastic models in biology are based on the theory of stochastic processes. A large class of biological models is based on discrete state, discrete time Markov chain (MC), with either finite or countable states space [1]. The stochastic counterpart of dynamical systems theory is known as random transformation [21], or discrete-time random dynamical system (RDS) [2]. It has a mathematical setup that is rather different from the theory of Markov processes, and the literature is somewhat inaccessible to a nonspecialist. RDS approach has found interesting applications in studies of synchronization, a concept inherited from the theory of non-autonomous ODEs, of neural firing in recent years [26]; it will have wide applications in studying cellular automata [51] within a fluctuating environment as well as many other complex dynamics, such as in finance [46]. With a fixed environment, a cellular automaton is a discrete-state discrete-time deterministic dynamical system which has an unambiguously defined response follows an unambiguous stimulus [49].

The goal of the present paper is to initiate an applied mathematical study of RDS with discrete state space. In particular, we like to establish the relation, as intuitive as possible, between the theory of MC and the theory of random transformation. Many of the materials presented in the present paper were contained in highly abstract mathematical theorems. We believe, however, that reducing the disciplinary barriers and developing a unified applied mathematics of stochastic dynamics in a discrete setting is critical for the wider applications of this new frontier of dynamic modeling in biology.

The paper is organized as follows. In Sec. 2, we establish the connection between the discrete i.i.d RDS and the finite-state Markov chain. For a given finite i.i.d RDS, the transition probability on state space can be defined and it further induces a finite-state Markov chain. In the mean time, for a given MC, it is shown that there exists a representation by means of i.i.d random transformations with some weak conditions [7]. In the finite state space, it is equivalent to decompose the induced MC transition matrix into the convex combination of deterministic transition matrices, with each coefficient being the probability for the corresponding deterministic transformation in the RDS. This RDS representation of an MC, however, is not unique. Different RDS representations of the same MC yield different behaviors, such as synchronization among trajectories of different initial states. From the viewpoint of random dynamical system, we can also study some properties of a given MC, such as invariant measure and reversibility.

In Sec. 3, we introduce the metric entropy of an RDS, and use it to characterize the different RDS representations of an MC. The upper and lower bounds of the metric entropy associated with an MC are analyzed. Actually the metric entropy of an RDS is no less than the Shannon-Khinchin entropy [20] of the induced MC [21]. We discover that when the equality holds, the transformations with positive probability have no “common dynamics”. The upper bound of the metric entropy among all the RDS representations, on the other hand, can be obtained explicitly, called Maximum-Metric-Entropy representation. This RDS has statistically independent movements given current positions being different.

In Sec. 4, we introduce the notion of synchronization in a RDS. Also it is shown that the sufficient and necessary condition for this maximum metric entropy RDS representation synchronizes is the Markov chain having a unique absorbing and aperiodic communicating class. This concept turns out to be intimately related to the Doeblin’s coupling in the theory of Markov process. We also present some
numerical results that characterize the synchronization phenomenon in different RDS representations of an MC. In Sec. 5, we explore the relation between RDS and another body of literature: the coupled stochastic processes. We show that the latter can be viewed as a generalization of the RDS to *random Markov system*. The mathematical models for biological motors proteins in terms of the coupled (or switching) diffusion process, called *Brownian ratchet*, is in fact one such example [38]. This subject is too broad to be discussed comprehensively. We shall illustrate one issue: the reversibility and Kolmogorov cycle condition. The paper concludes with Sec. 6 which contains a summary as well as some suggestions for future work.

Some of the mathematical materials can be found in the Appendices.

2. Discrete i.i.d. Random dynamical systems. The theory of random dynamical system (RDS) studies the action of random maps, drawn from a collection with prescribed probability, on a state space. Heuristically, the difference between an MC and an RDS is that the randomness in the former arises in a particular dynamics while in the latter is embedded in the “law of motion”. In this section, we present a brief overview of some essential concepts of RDS that are particularly relevant; we refer the reader to [21] for more mathematical discussion. The presentation is pedagogically self-contained.

2.1. RDS with discrete state space. We shall adopt the notations from [5], and focus on the case that the state space is finite.

An i.i.d random dynamical system is described by a triplet \((\mathcal{S}, \Gamma, Q)\), where \(\mathcal{S}\) is the state space, \(\Gamma\) is a family of maps from \(\mathcal{S}\) into itself and \(Q\) is the probability measure on the \(\sigma\)-field of \(\Gamma\), denoted as \(\mathcal{F}\). So \((\Gamma, \mathcal{F}, Q)\) forms a probability space. The set \(\Gamma\) is interpreted as the set of all admissible laws of dynamics.

In what follows, while all the mathematical definitions are general enough for a continuous state space \(\mathcal{S}\), we will give explicit examples in terms of a finite state space \(\mathcal{S} = \{s_1, s_2, \ldots, s_n\}\). Any \(\alpha \in \Gamma\) is called a deterministic transformation on \(\mathcal{S}\). Each such transformation has a \(n \times n\) matrix representation, called deterministic transition matrix. A deterministic transition matrix is a binary matrix that has all entries from \(\{0, 1\}\) but it has exactly one entry 1 in each row and 0s otherwise. Therefore, the cardinality \(|\Gamma| \leq n^n\). For convenience, later in this paper, \(n^n\) is denoted as \(N\). Among all deterministic transition matrices, there are \(n!\) permutation matrices and they correspond all invertible mappings of the finite state space \(\mathcal{S}\) to itself; all other \(N-n!\) matrices correspond to non-invertible maps, which necessarily have at least one column of 0s.

2.2. Eigenvalues and singular values of deterministic transition matrix.

If a deterministic transition matrix \(A\) is invertible, and we denote its transpose as \(A^T\), then \(AA^T\) is the identity matrix. In fact, \(A^T = A^{-1}\) is the inverse of the corresponding one-to-one transformation. Therefore, invertible deterministic transition matrix has all eigenvalues on the unit circle and singular value being 1. If \(A\) is a non-invertible deterministic transition matrix, then its eigenvalues are either on the unit circle or zero. It necessarily has at least one column of 0’s. Note, however, that both

\[
A_1 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{pmatrix}
\] (1)
have one column of 0’s. However, eliminating the transient state 2 yields

\[
\begin{pmatrix}
0 & 1 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 1 \\
1 & 0 & 0
\end{pmatrix}.
\]

This indicates that state 3 in \(A_2\) is also a transient state; in fact state 4 is also transient. The deterministic dynamics represented by \(A_2\) is \(2 \to 3 \to 4 \to 1\). It has a global “attractor” with a single state 1; thus its zero eigenvalue has a multiplicity of 3. In contrast, the deterministic transformation represented by \(A_1\) is \(2 \to 3 \to 4 \to 1 \to 3\). Its global “attractor” is a 3-state cycle. This is the dynamical reason why the zero eigenvalue of \(A_1\) has only multiplicity of 1. We can rearrange rows and columns of \(A_1\) to make it block upper triangular, then the above statement is clear. The singular values of a non-invertible deterministic transition matrix \(A_i, i = 1, 2\), provide different kind of information on the dynamics: The square of the singular values are the number of pre-image of a state that are also exactly the eigenvalues of the diagonal matrix, \(A_iA_i^T\). For example, both \(A_1\) and \(A_2\) have a same set of singular values \((\sqrt{2}, 1, 1, 0)\).

2.3. RDS and MC. The i.i.d RDS can be intuitively described as follows [5]: initially, the system is in some state \(x_0\) in \(S\), a map \(\alpha_1\) in \(\Gamma\) is chosen according to the probability measure \(Q\) and the system moves to the state \(x_1 = \alpha_1(x_0)\) in step 1. Again, independently of previous maps, another map \(\alpha_2\) is chosen according to the probability measure \(Q\) and the system moves to the state \(x_2 = \alpha_2(x_1)\). The procedure repeats. The initial state \(x_0\) can be a fixed state or an \(S\)-valued random variable \(X_0\) independent of all maps \(\alpha_n\). The random variable \(X_n\) is constructed by means of composition of independent random maps, \(X_n = \alpha_n \circ \alpha_{n-1} \circ \cdots \circ \alpha_1(X_0)\). Clearly, \(X_n\) is a Markov chain (MC) and its transition probability \(P(x, G)\) is expressed

\[
P(x, G) = Q\{\alpha : \alpha(x) \in G\},
\]

for any \(x \in S\) and any measurable set \(G \in \mathcal{B}(S)\). In terms of the \(n \times n\) deterministic transition matrix \(P_\alpha\) corresponding to \(\alpha \in \Gamma\), the transition probability matrix for the MC can be simply expressed as \(M = \mathbb{E}^Q[P_\alpha]\), where the expectation is taken with respect to probability measure \(Q\).

From a stochastic dynamics perspective, therefore, there is no difference between an i.i.d RDS and its induced MC if only a single sample trajectory is simulated. In other words, the one-point motion in this RDS is the Markov chain with transition probability \(P(x, G)\). The difference between them comes out when one is interested in two simultaneous sample trajectories with different initial states (but same \(\{\alpha_n\}\)) since two sequences \(\{X_n\}\) in the RDS are not independent. This is sometime called “identical noise realizations” [8] or “two-point motion” [4]. The difference in the two theories has been described as “two cultures” in [3].

For a given RDS, Eq. 2 uniquely defines an induced Markov chain. In the present work, we are interested in the reverse question: Can, and how, a Markov chain be represented by the compositions of i.i.d. random transformations? In the world of stochastic modeling, this provides a more refined stochastic description of dynamics that is consistent with a Markov model.

In more precise mathematical writings: Given the transition probability \(P(x, \cdot)\), \(x \in S\), does there exist a probability measure on \(\Gamma\), such that \(Q\{\alpha : \alpha(x) \in G\} = P(x, G)\), for all \(x \in S\) and any measurable set \(G \in \mathcal{B}(S)\)? There is a proof of the “can” for a general continuous transformations, given in [7] and [21]. In the finite
state situation, it follows the Theorem 2.1, which is an analog of the Birkhoff-von Neumann theorem in the theory of doubly stochastic matrices [6] and [31].

**Theorem 2.1.** The set of \( n \times n \) Markov transition matrices, \( \Omega_n \), forms a convex polyhedron with deterministic transition matrices as its vertices.

**Proof.** See Appendix A.

In more plain words, if matrix \( M \in \Omega_n \), then

\[
M = \sum_{\gamma \in \Gamma} a_\gamma P_\gamma,
\]

(3)

where \( P_\gamma \) are deterministic transition matrices and \( a_\gamma \) are nonnegative numbers, called weights, satisfying \( \sum_{\gamma=1}^{N} a_\gamma = 1 \). This implies that there always exists at least one RDS representation for any finite Markov chain with the probability mass function \( \{a_\gamma\}^N \) on the set \( \Gamma \).

The proof is based on a min-max algorithm. Here is the scratch of the algorithm: at each step the weight \( \alpha \) is the minimum of maximum entry in each row of the matrix \( M \) and the corresponding deterministic matrix \( P \) is the index of maximum entry in each row; then redefine \( M - \alpha P \) as \( M \) and keep this iteration until \( M \) becomes a zero matrix.

Such representation in general is not unique. This gives rise to the question of “how”: Which representation is reasonable under some prior information or requirements.

Here is an example that illustrates the existence and proves non-uniqueness.

**Example 1.** If the state space \( S = \{1, 2\} \) and the set of the transformations, expressed in terms of deterministic transition matrices, is

\[
\Gamma = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \right\}. \tag{4}
\]

Then a Markov transition probability matrix

\[
M = \begin{pmatrix} 0.4 & 0.6 \\ 0.7 & 0.3 \end{pmatrix}
\]

can be represented by the following decomposition by the min-max algorithm.

\[
M = 0.6 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + 0.3 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + 0.1 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{5}
\]

In the mean time, another decomposition could be

\[
M = 0.18 \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} + 0.28 \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + 0.42 \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + 0.12 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{6}
\]

This gives rise to the question of how many ways can a given Markov transition probability matrix be expressed in the form of (3).

In fact, one can ask the following combinatorial question: In the representations of a Markov transition matrix in the form (3), what is the least possible number of deterministic transition matrices, \( \kappa_s(M) \)? Some results of upper bounds for \( \kappa_s(M) \) can be obtained. Unlike a stochastic differential equation which has a well-defined deterministic counterpart, a Markov chain has never had an unambiguous deterministic reference. The “least”, therefore, could be interpreted as the “closest to a deterministic dynamics”. The following theorem gives a loose estimation.
Theorem 2.2. If $M \in \Omega_n$, then
\[ \kappa_*(M) \leq n^2 - n + 1. \] (7)

Proof. There are $n$ linear conditions on the row sums of a $n \times n$ Markov transition matrix. Therefore, $\dim \Omega_n = n^2 - n$. Carathéodory’s theorem shows that every matrix $M \in \Omega_n$ is in a convex hull of $n^2 - n + 1$ deterministic transition matrices. Then $\kappa_*(M) \leq n^2 - n + 1$. \qed

Notice that in Example 1, this upper bound has been reached.

For irreducible periodic Markov transition matrices, the upper bound can be further improved by using the period of the matrix [29].

The sequences of maps in the RDS can be generated other than i.i.d process. For example, another common type is to generate the random transformation via a Markov process [35, 12]: The probability which map being chosen at each step depends on the previous map. This RDS also induces a stochastic process on the state space $S$, but its Markovian property is lost in general. Here is an example that no matter how many steps of memory one keeps, it will not be a Markov chain.

Example 2. Consider a set of four deterministic transition matrices
\[
A = \begin{pmatrix}
1 & 0 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0
\end{pmatrix}, \quad
B = \begin{pmatrix}
0 & 0 & 1 \\
1 & 0 & 0 \\
1 & 0 & 0
\end{pmatrix}, \quad
C = \begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
1 & 0 & 0
\end{pmatrix}, \quad
D = \begin{pmatrix}
0 & 1 & 0 \\
1 & 0 & 0 \\
1 & 0 & 0
\end{pmatrix}.
\]

Define the transition probability between different maps as following:
\[
P_{A\to A} = 0.5, \quad P_{A\to B} = 0.5, \quad P_{B\to A} = 1, \quad P_{C\to C} = 0.5, \quad P_{C\to D} = 0.5, \quad P_{D\to C} = 1.
\]

It is illustrated by the Figure. 1

Choosing the initiate map with equal probability (0.5) to be $A$ or $C$. Then possible map sequences can be $\text{AAABAABA} \cdots$ or $\text{CDCCDCCC} \cdots$. Consider the RDS trajectory starting from state $X_0 = 1$, then the corresponding state sequences are $11131131 \cdots$ and $112112111 \cdots$. Notice that 2 and 3 cannot appear.
in the same sequence. Assume it is a Markov chain with respect to the memory of previous \( n \) digits. Then we have

\[
0.5 = \Pr \{X_{m+1} = 2 | X_{m-n} = 2, X_{m-n+1} = \cdots = X_m = 1 \} = \Pr \{X_{m+1} = 2 | X_{m-n+1} = \cdots = X_m = 1 \} = \Pr \{X_{m+1} = 2 | X_{m-n} = 3, X_{m-n+1} = \cdots = X_m = 1 \} = 0,
\]

which is a contradiction. Thus the stochastic process \( \{X_t\} \) is not a Markov chain with any finite length of memory.

2.4. Subshift of finite type. In i.i.d RDS, each map is chosen independently from previous chosen maps. The concept of independence in probability has to be defined in a product space. In our case, this product space will be infinite product of the copies of \( \Gamma \), \( \Omega_1 = \Gamma^\mathbb{Z}_+ \) and it is the set of all possible one-sided infinitely long sequence of deterministic transformations [21].

\[
\Omega_1 = \{ \omega : (\alpha_0\alpha_1\alpha_2\cdots\alpha_k\cdots) | \alpha_k \in \Gamma \}. \tag{8}
\]

A shift map \( \theta \) acting on each sequence shifts all the symbols to the left, i.e, \( \theta(\alpha_0\alpha_1\alpha_2\cdots\alpha_k\cdots) = (\alpha_1\alpha_2\cdots\alpha_k\cdots) \). At each step, the first transformation in this one-sided sequence will be the one applies on the state space. The sequence space together with the shift map, \( (\Omega_1, \theta) \), is called a subshift of finite type (SFT). Very insightfully, in terms of SFT one can show the shift map of an RDS is a deterministic dynamics with chaos, while in the state space the dynamics is nevertheless stochastic [27]. Probability measure on this product space \( \Omega_1 \) are defined on the cylinder set,

\[
\Pr_1(\alpha_0\alpha_1\alpha_2\cdots\alpha_k) = Q(\alpha_0)Q(\alpha_1)\cdots Q(\alpha_k). \tag{9}
\]

Clearly maps at different steps are chosen independently with the same distribution \( Q \).

The concept of SFT can also be defined for an MC. It would be interesting to discuss the relationship between SFT in an i.i.d RDS and SFT in its induced MC. Recall that the induced Markov chain is the one-point motion of the i.i.d RDS, with a finite state space of \( n \) alphabets, \( S \), a transition probability matrix \( M = \{p_{xy} | x, y \in S \} = Q(\alpha : \alpha(x) = y) \), and an initial distribution \( \{p_z, z \in S \} \). Then the sequence space \( \Omega_2 \) for the MC is the set of all possible one-sided infinite sequences of states,

\[
\Omega_2 = S^{\mathbb{Z}_+} = \{ \omega : (x_0x_1x_2\cdots x_k\cdots) | x_k \in S \}. \tag{10}
\]

Again the shift map \( \nu \) is \( \nu(x_0x_1x_2\cdots) = (x_1x_2x_3\cdots) \) and the probability measure on the cylinder set

\[
\Pr_2(x_0x_1x_2\cdots x_k) = p_{x_0}p_{x_0x_1}p_{x_1x_2}\cdots p_{x_{k-1}x_k}. \tag{11}
\]

In the studies of dynamical systems, especially in ergodic theory, a measure-preserving dynamical system is an important concept. For the shift map, it is defined as follows: \( \nu \) is measure preserving if \( P_2(\nu^{-1}(A)) = P_2(A) \), for all cylinder sets \( A \). Alternatively we can say that the measure \( P_2 \) is invariant under the shift map \( \nu \) if \( \nu \) is measure preserving [23]. One can show \( P_2 \) is invariant if and only if the initial distribution \( p_z \) is the invariant density of the MC, \( \pi : \pi_y = \sum_x \pi_xp_{xy} \). In i.i.d RDS, \( \theta \) is measure-preserving.
Each element \( \omega_1 = \alpha_0\alpha_1\alpha_2\cdots \) in the product space \( \Omega_1 \) is a sequence of deterministic transformations. If this sequence applies on \( n \) different initial conditions, i.e., \( s_1, s_2, \ldots, s_n \), it produces \( n \) different sequences in the product space \( \Omega_2 \). In the mean time, multiple elements in \( \Omega_1 \) apply on the same initial conditions may have exactly the same sequence of states. This implies that knowing one sample trajectory in the MC may not be enough to fully determine which transformation was picked in the i.i.d RDS view. We use “may” because in some special situations, this can actually be uniquely determined. We shall call such RDS’s having no common dynamics. It will be explored more in Sec. 3.3.

Question concerning two simultaneous trajectories, e.g., the joint probability of two sequences of states \( \omega \in \Omega_2 \) and \( \omega' \in \Omega_2 \), can not be asked; it is not defined in this model. Such probability has to be defined on the new product space \( \Omega_2 \otimes \Omega_2 \). We will discuss this more in the Sec. 4.2 in the context of synchronization.

3. The metric entropy. The concept of metric entropy was first introduced by Kolmogorov and further improved by Sinai [50, 34, 10]. It has been very successfully used in solving the isomorphism, or conjugacy, of dynamical systems. The metric entropy measures the maximal rate of information production a system is capable of generating [19]. It is a well-developed subject but can be technical in nature.
In the present work, we only give some very intuitive notions, and then focus on several relevant results for the metric entropy. Some mathematical derivations, in a heuristic fashion, can be found in Appendix B.

3.1. Metric and topological entropies of Markov chain. The notions of metric and topological entropies are very much motivated by the Gibbs’ and Shannon’s entropy, which is based on the large deviation of probability [48], and Boltzmann’s entropy, which is based on simple counting [13].

In this section, we only focus on the irreducible MC such that it will have the unique invariant density \( \pi \). The irreducibility of MC is equivalent with the dynamic being ergodic. Aperiodicity is not required for the entropy theory of MC.

The metric entropy \( h \) of a Markov chain with transition probability \( p_{ij}, i, j \in S \), is the asymptotic exponent of the vanishing probability for a single stochastic trajectory \( i_0 i_1 i_2 \cdots i_\ell \) with increasing \( \ell \),

\[
e^{-h \ell}
\]

\( (12) \)

\( p_{i_0 i_1} p_{i_1 i_2} \cdots p_{i_k i_{k+1}} \cdots p_{i_{\ell-1} i_\ell} = \exp \left( \sum_{k=1}^{\ell} \ln p_{i_{k-1} i_k} \right) = e^{-h \ell}, \)

and

\[
h = \lim_{\ell \to \infty} -\frac{1}{\ell} \sum_{k=1}^{\ell} \ln p_{i_{k-1} i_k} = - \sum_{i,j \in S} \pi_i p_{ij} \ln p_{ij}.
\]

(13)

Here we stipulate that \( 0 \ln 0 = 0 \).

The last step to replace time average by expectation from invariant distribution, is based on the ergodicity of the MC: As \( \ell \to \infty \), the frequency of the state \( i \) in the sequence goes to \( \pi_i \) and the frequency of the pair \( ij \) becomes \( \pi_i p_{ij} \).

Note that for a deterministic transformation with probability 1 for transition \( i_k \to i_{k+1} \), the probability in (12) is a constant. Therefore, its metric entropy \( h = 0 \).

The topological entropy of a Markov chain \( \eta \) characterizes what is possible and what is not; it is independent of actual values of transition probabilities. Two Markov chains with transition probabilities \( p_{ij} \) and \( q_{ij} \), \( i, j \in S \), has a same topological entropy when \( p_{ij} = 0 \) if and only if \( q_{ij} = 0 \). They both induce a same SFT, and the topological entropy \( \eta \) simply counts the number of possible trajectories generated by the Markov process. Consider an \( n \times n \) irreducible binary matrix \( A \), \( n = \|S\| \), satisfying \( A_{ij} = 1 \) when \( p_{ij} > 0 \), \( A_{ij} = 0 \) when \( p_{ij} = 0 \). We call it the “topological skeletal matrix” of this MC. The number of possible trajectories increases asymptotically with \( \ell \):

\[
\Omega_B(\ell) \overset{\Delta}{=} (1, 0, \cdots, 0) A^\ell \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \to C_1 e^{(\ln \lambda_A) \ell},
\]

(14)

in which \( \lambda_A \) is the largest positive eigenvalue of the matrix \( A \). The topological entropy \( \eta \), following Boltzmann’s notion of entropy

\[
\eta = \lim_{\ell \to \infty} \frac{\ln \Omega_B(\ell)}{\ell} = \ln \lambda_A.
\]

(15)

It is easy to show that \( \eta \geq h \). For a Markov chain with all \( p_{ij} = \frac{1}{n} \), the equality is attained \( \eta = h = \ln n \).
The definition of topological entropy according to (14) also clearly indicates that \( \eta \) is the same for a stochastic process with “increasing number of distinct trajectories” and an endomorphism with “increasing number of distinct pre-image” [41].

For an MC, its topological entropy \( \eta \) is intimately related to how many connected neighbours a state has, or “dimensions” in a lattice system [27]. For an MC with exact \( \nu \) non-zero elements in each and every row of its transition matrix, \( \nu < n \), then the largest eigenvalue of its “topological skeletal matrix” \( \mathbf{A} \) is \( \lambda_{\mathbf{A}} = \nu \). The corresponding right eigenvector is clearly \( (1, 1, \ldots, 1) \). This is the “topology” of the MC, which is also the metric entropy of the Markov chain with \( p_{ij} = A_{ij}/\nu \). Its corresponding invariant distribution, e.g. the left-eigenvector is \( \left( \frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n} \right) \):

\[
\sum_{i=1}^{n} \frac{1}{n} A_{ij} = \frac{1}{n}.
\] (16)

Then we have the following result which could be very useful for sparse Markov networks:

**Proposition 1.** The metric entropy of a Markov chain

\[
h \leq \sum_{i \in S} \pi_i \ln \nu_i,
\]

where \( \nu_i \) is cardinality of \( \{ j : p_{ij} > 0 \} \), and \( \{ \pi_i \} \) is the stationary distribution. In particular, if the MC has the maximal number of neighboring states being \( \nu \), then \( h \leq \ln \nu \).

**Proof.** Let us now consider all the possible Markov chains with state state \( S \) which have the same “topological skeletal matrix” \( \mathbf{A} : p_{ij} = p_{ij}/A_{ij} \). Denote \( \nu_i = \sum_j A_{ij} \). Then for each \( i \), the transition probability \( p_{ij} \) has lower entropy than the uniform \( A_{ij}/\nu_i \) whose entropy is simply \( \ln \nu_i \). The metric entropy for the Markov chain then is:

\[
h = - \sum_{i,j \in S} \pi_i p_{ij} \ln p_{ij} \\
\leq \sum_{i \in S} \pi_i \ln \nu_i.
\]

3.2. Metric entropy of i.i.d. RDS. In the finite i.i.d RDS, there is a i.i.d process to choose the deterministic transformation at each step. This i.i.d process is isomorphic to the 1-sided \( (a_1, a_2, \ldots, a_N) \) Bernoulli shift, so the entropy for this i.i.d process is naturally defined as \( h = -\sum_{i=1}^{N} a_i \ln(a_i) \) [27]. Moreover, the deterministic transformation with finite state has zero metric entropy once it is chosen. Therefore, the randomness of this system is solely generated by the i.i.d process.

The metric entropy for this RDS is given as follows.

**Proposition 2.** Suppose the state space \( S \) is finite with \( ||S|| = n \), and \( \{ a_i \}^N \) is the probability measure on the set of all deterministic transformations, \( \Gamma \). Then the metric entropy for the RDS is \( h_{RDS} = -\sum_{i=1}^{N} a_i \ln(a_i) \).

It is worth mentioning that the metric entropy of the RDS in many cases can be infinite. This is mainly because when there are countably infinite transformations in \( \Gamma \), the infinite sum in the entropy may not converge. A different notion of entropy of i.i.d. RDS that remedies the difficulty is defined as the weighted mean of the
metric entropy for all deterministic transformations with probability mass as their weight [21]. We do not need to be concerned with this since our $S$ is always finite.

We are now in the position to address the question that for a given MC, which RDS representation is reasonable according to certain requirement. In fact, the metric entropy of its corresponding RDS becomes a good characterization for different representations. It is natural to ask the lower bound and upper bound of them for a given MC.

3.3. Lower bound of metric entropy. Kifer notified this question and gave the result that $h_{RDS} \geq h_{MC}$ [21]. In the finite i.i.d RDS, different sequences of deterministic transformations applies on a same initial condition might induce the same MC trajectory, as we discussed in Sec. 2.4. That means more information is required to determine which deterministic transformation is chosen at each step. In other words, the RDS generates more information than the MC at each step. With this intuition, Kifer’s result becomes clear; a general proof can be found in his book [21]. We will provide an elementary proof here for the finite i.i.d RDS and illustrate the condition under which the equality is attained.

Denote a deterministic transition matrix $P \in \Gamma$ as $P_{i_1,i_2,\ldots,i_n}$ if $P_{i,i} = 1$. Denote its corresponding probability as $a_{i_1,i_2,\ldots,i_n}$. For example, $P_{1,2,1}$ represents the matrix
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
1 & 0 & 0
\end{pmatrix}
\]
The transition probability matrix of the induced MC then is
\[
M = \sum_{i_1,i_2,\ldots,i_n} a_{i_1,i_2,\ldots,i_n} P_{i_1,i_2,\ldots,i_n}.
\]

Definition 3.1. The deterministic transformation $\varphi_1$ and $\varphi_2$ have common dynamics if there exists a state $s$ in the space $S$ such that $\varphi_1(s) = \varphi_2(s)$.

By the representation of deterministic transition matrices $P_{i_1,i_2,\ldots,i_n}$ and $Q_{j_1,j_2,\ldots,j_n}$, the definition is equivalent with $P - Q$ has at least one row being all zeroes. In our notation, that is, $\exists k$ such that $i_k = j_k$. If two deterministic transformations have no common dynamics, then $P - Q$ has no rows that being all zeroes. This definition can be extended to multiple deterministic transformations. If no pair of two deterministic transformations has common dynamics, we call all these deterministic transformations have no common dynamics.

Theorem 3.2. In the finite state space $S$, $h_{RDS} \geq h_{MC}$. The equality is reached if and only if deterministic transformations with positive probability have no common dynamics.

Proof. The metric entropy of the Markov chain is
\[
h_{MC} = -\sum_{j,k=1}^{n} \pi_j M_{jk} \ln M_{jk}
\]
\[
= -\sum_{j,k=1}^{n} \pi_j \left( \sum_{i_j=k} a_{i_1,i_2,\ldots,i_n} \right) \ln \left( \sum_{i_j=k} a_{i_1,i_2,\ldots,i_n} \right)
\]
\leq - \sum_{j,k=1}^{n} \pi_j \sum_{i_j=k}^{n} \left( a_{i_1,i_2,...,i_n} \ln a_{i_1,i_2,...,i_n} \right)
= - \sum_{i_1,i_2,...,i_n} \left( a_{i_1,i_2,...,i_n} \ln a_{i_1,i_2,...,i_n} \right) \sum_{j}^{n} \pi_j
= - \sum_{i_1,i_2,...,i_n} a_{i_1,i_2,...,i_n} \ln a_{i_1,i_2,...,i_n} = h_{\text{RDS}}.

The equality holds if $M_{jk}$ is positive, among all deterministic transition matrices satisfying $i_j = k$, only one has positive probability mass and others are 0. This positive probability of course is $M_{jk}$. Therefore, all deterministic transformations with positive probability have no common dynamics. \hfill \square

In the case that the equality is attained, and when simulating a trajectory in the state space $S$ with such RDS, the deterministic transformation is uniquely identified in each step. The following corollaries illustrate more about the RDS and its induced Markov chain as the equality attained.

**Corollary 1.** If $h_{RDS} = h_{MC}$, the cardinality of the deterministic matrices with positive probability is no greater than $n$, i.e., $\| \Gamma \| \leq n$

**Corollary 2.** If $h_{RDS} = h_{MC}$, each row of the induced Markov transition matrix is a permutation of the first row.

In fact, $h_{RDS} = h_{MC}$ if and only if we can recover the RDS map sequence from the MC trajectory.

For a given Markov transition probability matrix, it is not necessarily true that there exists an RDS representation whose metric entropy is the same as that of the MC. In fact, for most Markov transition matrices, it is not reachable. If exists, such RDS might not be unique as well. Nevertheless, it provides a possible lower bound of the metric entropy. In order to give a reachable lower bound, it usually need to solve a non-convex problem and no desirable solution is available.

### 3.4. Upper bound of metric entropy

It is natural to ask what is the upper bound of the metric entropy of all RDS representations given a finite MC. It turns out we can find an analytic expression for the decomposition.

**Lemma 3.3.** $\sum_{t_1,t_2,...,t_n=1}^{n} M_{1t_1}M_{2t_2} \cdots M_{nt_n} = 1$.

So there exists a decomposition, that is, $a_{i_1,i_2,...,i_n} = M_{1t_1}M_{2t_2} \cdots M_{nt_n}$.

**Theorem 3.4.** The metric entropy $h_{RDS} \leq - \sum_{j,k} M_{kj} \ln M_{kj}$ and the equality holds if and only if $a_{i_1,i_2,...,i_n} = \prod_k M_{ki_k}$.

**Proof.** We can show

\[- \sum_{k,j} M_{kj} \ln M_{kj} - h_{RDS} = - \sum_{k,j} \left( \sum_{i_k=j} a_{i_1,i_2,...,i_n} \right) \ln M_{kj} - h_{RDS} \]
\[- = - \sum_{i_1,i_2,...,i_n} \left( a_{i_1,i_2,...,i_n} \ln (\prod_k M_{ki_k}) - h_{RDS} \right) \]
\[- = \sum_{i_1,i_2,...,i_n} a_{i_1,i_2,...,i_n} \ln \frac{a_{i_1,i_2,...,i_n}}{\prod_k M_{ki_k}} \geq 0 \]

So there exists a decomposition, that is, $a_{i_1,i_2,...,i_n} = M_{1t_1}M_{2t_2} \cdots M_{nt_n}$.
The Gibbs inequality is applicable since \( a_{i_1,i_2,...,i_n} \) and \( \Pi_k M_{ki_k} \) are probability mass functions. The equal sign holds if and only if these two functions are identical, i.e., \( a_{i_1,i_2,...,i_n} = \Pi_k M_{ki_k} \).

The example in (6) is the maximum-metric-entropy representation for the given Markov transition matrix \( M \). It is easy to see if all entries of the transition matrix \( M \) are positive, then all \( n^n \) deterministic matrices will have positive probability. Furthermore, the most probable deterministic transition matrix corresponds to the deterministic transformation that maps to the state with the largest probability given the current state, i.e., its deterministic matrix has entry one in the position that is maximum in each row of the transition matrix \( M \). It is a very insightful result since this is the most reasonable “deterministic counterpart” for a given MC with transition probability matrix \( M \).

4. Synchronization. Each realization of an RDS is a non-autonomous dynamical systems [2]. Synchronization is a well-developed concept in the theory of non-autonomous ODEs [26]. Through studying synchronization in an RDS, one appreciates the fact that RDS formulation is a more refined model of stochastic dynamics than an MC. The same phenomenon has also been terms as random sink [4, 26] in neuron dynamics applications. It has a deep relation to the concept of coupling in the theory of MC [25, 47].

4.1. Defining synchronization. Recall the SFT \((\Omega_1, \theta)\) derived from the finite i.i.d. RDS defined in Sec 2.4, for each element \( \omega_1 \) in the product space, we apply this sequence of deterministic transformations on multiple initial conditions simultaneously, and it induces multiple sequences of states. These sequences are not independent; actually once they collide at some instance, they will be together forever. This phenomenon is called synchronization. In fact, it is easy to see that if any pair of states synchronize in the finite step almost surely, any multiple sequences will synchronize to one single sequence almost surely. So we can reduce this to the study of two-point motion [4].

We shall first give a mathematical definition for synchronization: Denoting \( f(\omega) = \omega(1) \), that is the first transformation in this one-sided infinite sequence \( \omega \) and \( f^{(n)}(\omega) = f(\theta^{n-1} \omega) \circ \cdots \circ f(\theta \omega) \circ f(\omega) = \omega(n) \circ \cdots \circ \omega(2) \circ \omega(1) \), this is the composition of first \( n \) transformations in \( \omega \).

Definition 4.1. The finite i.i.d RDS \((\Omega, F, \Pr, \theta)\) synchronizes if for any two different initial states \( s_1 \) and \( s_2 \) in \( S \), i.e, \( s_1 \neq s_2 \)

\[
\lim_{n \to +\infty} \Pr \left\{ \omega : f^{(n)}(\omega)s_1 = f^{(n)}(\omega)s_2, n \in \mathbb{Z}^+ \right\} = 1.
\]

There are always some sequences that they will never synchronize but we claim the probability of these sequences is 0. So they are insignificant.

4.2. Maximum metric entropy representation and Doeblin’s coupling. We note, however, that not every RDS will possess such property. Given a Markov chain with the transition probability matrix \( M \), we would like to have a survey on the synchronization of its RDS representations. We are particular interested in the maximum metric entropy representation, i.e, \( a_{i_1, i_2, ..., i_n} = M_{1i_1} M_{2i_2} \cdots M_{ni_n} \).

In studying synchronization, one needs simultaneous construction of two infinite sequences by applying the sequence \( \omega_1 \) on the two initial states as a pair \((x_0, y_0)\),
so the product space will be $\Omega_2 \otimes \Omega_2$ and the shift map can be induced from $\theta$ in the RDS. It describes a two-point motion.

In terms of this RDS, it becomes a new Markov chain $(X, Y)$ whose state space is $S \times S$. Its transition probability matrix $W$ is defined as

$$W_{(i,j)\rightarrow(k,\ell)} = \begin{cases} \sum_{t_i=t_k,t_j=\ell} a_{t_1,t_2,...,t_n} & \text{when } i \neq j, \\ M_{ik} & \text{when } i = j \text{ and } k = \ell, \\ 0 & \text{otherwise.} \end{cases}$$  \(17\)

In Eq. 17, $a_{t_1,t_2,...,t_n}$ is the probability mass of the deterministic matrix $P_{t_1,t_2,...,t_n}$. $W$ is an $n^2 \times n^2$ stochastic matrix. Each component of this new Markov chain is the one-point motion in the RDS, namely the original Markov chain, since summing out the probability of the other component will degenerate the transition probability $W$ into the original transition probability matrix $M$.

The synchronization of the RDS means that only the states $\{(1,1), (2,2), \ldots, (n,n)\}$ are possibly recurrent, while all other states are transient. Moreover, the transition probability among these recurrent states are exactly the same as the original transition probability because the two-point motion collapses down to a copy of the one-point motion. In the maximum-metric-entropy decomposition, the previous expression can be further simplified as

$$W_{(i,j)\rightarrow(k,\ell)} = \begin{cases} M_{ik}M_{j\ell} & \text{when } i \neq j, \\ M_{ik} & \text{when } i = j \text{ and } k = \ell, \\ 0 & \text{otherwise.} \end{cases}$$  \(18\)

In probability theory, Eq. 18 is the transition probability of a Markov chain, as the most basic example of coupling of two Markov chains, first used by Wolfgang Doeblin. This Markov chain $(X, Y)$ behaves as follows: if $X_n \neq Y_n$, then the two components make independent movements according to the transition matrix $M$ at each step; if $X_n = Y_n$, they make the same movements. Thus, $\{(1,1), (2,2), \ldots, (n,n)\}$ is absorbing.

Actually, in terms of Eq. 19 below, it is very clear why Doeblin’s coupling is different from two independent Markov chains: The transition probability matrix $W'$ for the two-point motion of two independent Markov chains is

$$W'_{(i,j)\rightarrow(k,\ell)} = M_{ik}M_{j\ell}. \quad \text{(19)}$$

$W$ is the same as $W'$ except at these $n$ rows of $(1,1), (2,2), \ldots, (n,n)$.

With the introduction of $W$, we can find the precise condition under which an RDS synchronizes. We have the following theorem.

**Theorem 4.2.** For a Markov chain, its corresponding RDS with the maximum-metric-entropy representation synchronizes if and only if this MC has unique absorbing communicating class, and the MC restricted on this communicating class is aperiodic.

**Remark.** From this theorem, it is easy to see that if a Markov chain is irreducible and aperiodic, then the maximum-metric-entropy RDS synchronizes.

**Proof.** $\Rightarrow$: Remember that once a trajectory falls into an absorbing communicating class, it will never get out again. If the Markov chain has at least two absorbing communicating classes, then states from these two classes will never synchronize.
If the MC restricted on an absorbing communicating class is \( k \)-periodic, then this communicating class can be divided into \( k \) subclasses. Each time the MC jumps from one subclass to the next. States from different subclasses will never synchronize.

\[ \Leftarrow: \] Denote the absorbing communicating class by \( \mathcal{C} \). Since the absorbing communicating class is unique, there exists a positive integer \( n_1 \) such that for any state \( i \in \mathcal{S} \setminus \mathcal{C} \), there exists a state \( j \in \mathcal{C} \) to make \( p^m(i,j) > 0 \). Since the Markov chain restricted on \( \mathcal{C} \) is irreducible and aperiodic, there exists a positive integer \( n_2 \) such that for any two states \( i, j \in \mathcal{C} \), \( p^{n_2}(i,j) > 0 \). Set \( n = n_1 + n_2 \), then for any two states \( i \in \mathcal{S}, j \in \mathcal{C}, p^n(i,j) > 0 \).

Now for any two initial states \( i_0, j_0 \in \mathcal{S} \), we can find admissible sequences to reach any state \( k_0 \in \mathcal{C} \) after \( n \)th step:

\[
i_{0,1,2,\ldots,i_{n-1},i_n} = k_0 \quad \text{and} \quad j_{0,1,2,\ldots,j_{n-1},j_n} = k_0.
\]

If this two sequence collide before the step \( n \), i.e., \( m = \min \{ k : i_k = j_k \} < n \), we truncate the sequence up to \( m \) steps. Now there are two admissible sequences which reach the same state at step \( m \) in the first time. So in the intermediate states, \( i_k \neq j_k \), for \( 1 \leq k \leq m - 1 \). The probabilities of both sequences in MC are strictly positive.

Now we want to calculate the probability for both sequences with initial states \( i_0 \) and \( j_0 \), \( p(i_0, j_0) \), in the maximum-metric-entropy representation RDS.

\[
p(i_0, j_0) = \Pr \{ \omega : f^{(k)}(\omega)i_0 = i_k, f^{(k)}(\omega)j_0 = j_k, 1 \leq k \leq m \} = (M_{i_0,i_k}M_{j_0,j_k}M_{i_k,i_{k+1}}M_{j_k,j_{k+1}}\cdots M_{i_{m-1},i_m}M_{j_{m-1},j_m}) \cdot \Pr \{ i_0, i_1, i_2, \ldots, i_{m-1}, i_m \} \cdot \Pr \{ j_0, j_1, j_2, \ldots, j_{m-1}, j_m \}.
\]

This probability is strictly positive. \( p(i_0, j_0) \) gives a lower bound of the probability of sequences starting with \( i_0 \) and \( j_0 \) synchronize within \( n \) steps. For convention, if \( i_0 = j_0 \), \( p(i_0, j_0) = 1 \). Among all different pairs of initial states, define \( p = \min_{i,j} p(i,j) \), which is the lower bound of the probability for any two initial states to synchronize within \( n \) steps. The probability that any two sequences don’t synchronize within \( n \) steps is \( P_{\text{non-sync}} = \Pr \{ \omega : f^{(n)}(\omega)s_1 \neq f^{(n)}(\omega)s_2 \} \leq (1 - p) \).

So the probability for any two sequences do not synchronize within \( kn \) steps is \( P_{\text{non-sync}}^k < (1 - p)^k \). It implies any two sequences do not synchronize in finite steps has probability 0. Thus, the finite i.i.d RDS synchronizes by definition.

4.3. **Examples and numerical results.** We shall show several examples of synchronization through numerical simulations. We consider Markov chain with \( 4 \times 4 \) transition matrix, thus \( ||\Gamma|| \leq 256 \). We will simulate four sample trajectories in the RDS setting and these trajectories start from four different initial conditions and stop once all four collide into one trajectory; count down steps required. We are interested in two questions: first, given a Markov transition matrix, what kind of RDS representation will synchronize and what kind of will not?; second, if it synchronizes, what is the probability distribution of the synchronization steps (or coupling time)? A more rigorous definition for the random variable \( N_s \), the synchronization steps is

\[
N_s(\omega) = \min \left\{ n : f^{(n)}(\omega)s_1 = f^{(n)}(\omega)s_2 = f^{(n)}(\omega)s_3 = f^{(n)}(\omega)s_4, n \in \mathbb{Z}^+ \right\} \tag{20}
\]

where \( s_1, s_2, s_3, s_4 \) are four different initial conditions.
For the first question, we have shown the sufficient and necessary condition for the maximum-metric-entropy representation. It is possible that an irreducible and aperiodic MC in other RDS representations will not synchronize. A trivial example is an irreducible and aperiodic doubly stochastic Markov matrix being decomposed into a convex combination of permutation matrices. However, if $\Gamma$ contains non-permutation matrices, it is still possible that RDS will not synchronize.

**Example 3.** Consider the $4 \times 4$ Markov transition matrix $M$ and it is decomposed into the following way,

\[
\begin{pmatrix}
0.4 & 0.3 & 0.1 & 0.2 \\
0.3 & 0.4 & 0.2 & 0.1 \\
0.1 & 0.2 & 0.4 & 0.3 \\
0.2 & 0.1 & 0.6 & 0.1
\end{pmatrix}
= 0.3
\begin{pmatrix}
1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0
\end{pmatrix}
+ 0.1
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}
+ 0.2
\begin{pmatrix}
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 \\
0 & 1 & 0 & 0 \\
1 & 0 & 0 & 0
\end{pmatrix}
+ 0.3
\begin{pmatrix}
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0
\end{pmatrix}
\]

The first four deterministic matrices are permutation matrices and the last one is not. In such RDS representation, it is easy to see that two trajectories starting from state 1 (or state 2) and state 3 (or state 4) will not synchronize into one, even though the Markov transition matrix is irreducible and aperiodic.

For the second question, it is very hard to calculate the distribution of synchronization steps due to deterministic transition matrices being non-commutative. We simulate various examples of Markov transition matrices in maximum metric entropy representation and min-max representation and histogram the frequency of synchronization steps. In Figure 3, we give two examples of Markov transition matrices $M_1$ and $M_2$. They are

\[
M_1 = \begin{pmatrix}
0 & 0.38 & 0.41 & 0.21 \\
0 & 0 & 0 & 1 \\
0.18 & 0.30 & 0.44 & 0.08 \\
0.39 & 0.11 & 0 & 0.5
\end{pmatrix}, \quad M_2 = \begin{pmatrix}
0 & 0 & 0.86 & 0.14 \\
0.14 & 0.51 & 0 & 0.35 \\
0.09 & 0.38 & 0.53 & 0 \\
0 & 0.47 & 0.46 & 0.07
\end{pmatrix}
\]

Both matrices are irreducible and aperiodic so maximum-metric-entropy representation will synchronize. We also verified min-max representations synchronize through finding the absorbing states after constructing the $16 \times 16$ matrixes $W$ in (17). From the Figure 3, the maximum-metric-entropy representation synchronizes faster than the min-max representation on average for the first Markov transition matrix; however, the min-max one is faster on average for the second. The tail frequency for both representations shows an exponential decaying, as illustrated by the semi-log linear fits of the tail probability $p(x) \propto a \exp(-bx)$. The slope $b$ is related to the second largest eigenvalue of the transition matrix of the two-point motion, $W$.

5. **Random Markov systems.** In the present work, RDS is defined as a sequence of identical, independently distributed deterministic transformations on discrete state space $S$. If we replace the collection of all possible deterministic transformations by a set of stochastic matrices, then we obtain a generalization of RDS, which we shall term random Markov system (RMS). The description of RMS is similar with RDS if one-point motion is considered.
Figure 3. The top left and bottom left figures are the histograms of the synchronization steps for Markov transition matrices $M_1$ and $M_2$ after 150,000 independent simulations in each representation. The top right and bottom right figures are the logarithm of the frequency scatter plots of the histograms with the linear fits.

There already exists a large body of literature on this subject, under various names, ranging from coupled (or switching) diffusion processes [42, 28, 52, 54] to theories of random evolution and random medium [36, 32]. A key motivation of this class of models is to distinguish intrinsic noise that causes stochastic transitions and extrinsic noise, which is the origin of time-dependent parameters of a dynamical system.

In biophysical chemistry, one of the central issues for this class of problems is the nature of detailed balance, or reversibility [9, 16]. See [39] for an extensive discussion on how extrinsic noise can be a source of free energy that sustains a biochemical system out of equilibrium. In mathematical terms, a random i.i.d. sequence of reversible stochastic matrices, in general, defines a Markov process that is not necessarily reversible. This can be easily shown from the following example:

Example 4.

$$\frac{1}{3} \begin{pmatrix} 0.3 & 0.7 & 0.0 \\ 0.35 & 0.15 & 0.5 \\ 0.0 & 0.1 & 0.9 \end{pmatrix} + \frac{2}{3} \begin{pmatrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{pmatrix} = \begin{pmatrix} \frac{1.1}{3} & \frac{1.3}{3} & 0.2 \\ \frac{0.95}{3} & \frac{0.95}{3} & \frac{1.1}{3} \\ \frac{0.2}{3} & \frac{0.7}{3} & \frac{1.7}{3} \end{pmatrix}. \quad (23)$$

The first stochastic matrix has a detail-balanced stationary probability $\frac{1}{13}(1, 2, 10)$, and the second stochastic matrix has a detail-balanced stationary probability $\frac{1}{5}(1, 1, 1)$. However, the stochastic matrix on the right-hand-side has a Kolmogorov cycle criterion [53, 14], also known as entropy production per cycle [40]

$$\ln \left( \begin{pmatrix} 1.3 \times 1.1 \times 0.2 \\ 0.95 \times 0.7 \times 0.2 \end{pmatrix} \right) = \ln \left( \begin{pmatrix} 1.43 \\ 0.665 \end{pmatrix} \right) \neq 0.$$
Therefore, it is not reversible.

**Lemma 5.1.** If a collection of i.i.d. reversible stochastic matrices all have a same invariant probability, then the RMS is detailed balanced.

**Proof.** Let $P^{(k)} = \{p^{(k)}_{ij}\}$ be the transition probability matrix for the $k$th Markov system in the collection, $1 \leq k \leq K$, all with a same invariant probability distribution $\{\pi_i\}$. Let $a_k$ be the probability of the $k$th system being sampled in the i.i.d. RMS. Then the induced Markov chain has transition probability matrix $M$ with elements $M_{ij} = \sum_{k=1}^K a_k p^{(k)}_{ij}$. The first assertion is immediate from the definition of detailed balance: $\pi_i p^{(k)}_{ij} = \pi_j p^{(k)}_{ji} \forall k$, which yields

$$\pi_i M_{ij} = \pi_i \sum_{k=1}^K a_k p^{(k)}_{ij} = \sum_{k=1}^K a_k \pi_j p^{(k)}_{ji} = \pi_j M_{ji}.$$  

\[ \Box \]

Treating $P$ as a matrix-valued random variable with probability measure $\{a_i; 1 \leq i \leq K\}$, the pair $(P(t), X(t))$ is also a Markov chain with transition probability $P_{(M_k, x_i)\rightarrow(M_k, x_j)} = a_k p^{(k)}_{ij}$.

Then its stationary entropy production is [54]:

$$\frac{1}{2} \sum_{\ell, k} \sum_{i, j \in S} \left( a_\ell \pi_i P_{(M_\ell, x_i)\rightarrow(M_k, x_j)} - a_k \pi_j P_{(M_k, x_j)\rightarrow(M_\ell, x_i)} \right) \times \ln \left( \frac{a_\ell \pi_i P_{(M_\ell, x_i)\rightarrow(M_k, x_j)}}{a_k \pi_j P_{(M_k, x_j)\rightarrow(M_\ell, x_i)}} \right)$$

$$= \frac{1}{2} \sum_{\ell, k} a_\ell a_\ell \sum_{i, j \in S} \left( \pi_i p^{(k)}_{ij} - \pi_j p^{(\ell)}_{ji} \right) \ln \left( \frac{\pi_i p^{(k)}_{ij}}{\pi_j p^{(\ell)}_{ji}} \right).$$

If $P^{(k)}$ are not detailed balance, but they all still have a common stationary distribution $\{\mu_i\}$. Then,

**Lemma 5.2.** Assuming a same invariant distribution, the stationary entropy production rate (epr) of a full RMS is never less than the epr of the composite Markov chain:

$$\sum_{i, j \in S} \sum_{k, \ell} a_k a_\ell \pi_i p^{(k)}_{ij} \ln \left( \frac{\pi_i p^{(k)}_{ij}}{\pi_j p^{(\ell)}_{ji}} \right) \geq \sum_{i, j \in S} \pi_i M_{ij} \ln \left( \frac{\pi_i M_{ij}}{\pi_j M_{ji}} \right). \quad (24)$$

**Proof.** We note the left-hand-side and the right-hand-side of (24) are, respectively,

$$\sum_{i, j \in S} \sum_{k, \ell} a_k a_\ell \pi_i p^{(k)}_{ij} \ln \left( \frac{a_k a_\ell \pi_i p^{(k)}_{ij}}{a_k a_\ell \pi_j p^{(\ell)}_{ji}} \right),$$

$$\sum_{i, j \in S} \left( \sum_{k, \ell} a_k a_\ell \pi_i p^{(k)}_{ij} \right) \ln \left( \frac{\sum_{k, \ell} a_k a_\ell \pi_i p^{(k)}_{ij}}{\sum_{k, \ell} a_k a_\ell \pi_j p^{(\ell)}_{ji}} \right). \quad (25)$$

and

$$\sum_{i, j \in S} \left( \sum_{k, \ell} a_k a_\ell \pi_i p^{(k)}_{ij} \right) \ln \left( \frac{\sum_{k, \ell} a_k a_\ell \pi_i p^{(k)}_{ij}}{\sum_{k, \ell} a_k a_\ell \pi_j p^{(\ell)}_{ji}} \right). \quad (26)$$
Based on the log-sum inequality, which states that for two sets of non-negative numbers \((\alpha_1, \alpha_2, \cdots, \alpha_n)\) and \((\beta_1, \beta_2, \cdots, \beta_n)\),

\[
\sum_{i=1}^{n} \alpha_i \ln \left( \frac{\alpha_i}{\beta_i} \right) \geq \left( \sum_{i=1}^{n} \alpha_i \right) \ln \left\{ \frac{\sum_{i=1}^{n} \alpha_i}{\sum_{i=1}^{n} \beta_i} \right\},
\]

we have

\[
\sum_{k,\ell} a_k a_\ell \pi_i p_{ij}^{(k)} \ln \left( \frac{a_k a_\ell \pi_i p_{ij}^{(k)}}{a_k a_\ell \pi_i p_{ji}^{(\ell)}} \right) \geq \left( \sum_{k,\ell} a_k a_\ell \pi_i p_{ij}^{(k)} \right) \ln \left\{ \frac{\sum_{k,\ell} a_k a_\ell \pi_i p_{ij}^{(k)}}{\sum_{k,\ell} a_k a_\ell \pi_i p_{ji}^{(\ell)}} \right\}.
\]

Summing over all \(i, j\), the two sides leads to (25) and (26).

The equality in (27) holds true if and only if \(\alpha_i \beta_i\) are equal for all \(i\). Thus, the equality in the Lemma is achieved if and only if \(p_{ij}^{(k)} = p_{ji}^{(\ell)}\) for all \(i, j, k, \ell\); there is no difference among the Markov processes being sampled.

Therefore, if not all the Markov systems are detail balanced, then the total stationary entropy of the RMS, in terms of the pair \((P(t), X(t))\), is greater than the entropy production of the composite MC, \(X(t)\), with transition probability matrix \(M = E^Q[P]\) defined similarly in Sec. 2.3. As discussed in [44], the details in these three systems decrease as a form of coarse graining, thus the entropy production.

Interestingly, the mean entropy production rate of a set of Markov chains in a RMS can actually be greater than the total stationary entropy production rate of the RMS.

**Example 5.** Consider \(a_1 = a_2 = 0.5\), with two Markov chains

\[
P^{(1)} = \begin{pmatrix}
d & 1-2d & d \\
d & d & 1-2d \\
1-2d & d & d\end{pmatrix},
\]

\[
P^{(2)} = \begin{pmatrix}
d & d & 1-2d \\
1-2d & d & d \\
d & 1-2d & d\end{pmatrix}.
\]

Their common stationary distribution is \((\frac{1}{3}, \frac{1}{3}, \frac{1}{3})\). The entropy production rate for both is \((1-3d) \ln \left( \frac{1-2d}{d} \right)\), thus the mean entropy production rate. The left-hand-side of (24), since \(p_{ij}^{(1)} = p_{ji}^{(2)}\), is only half:

\[
\frac{1}{2} \sum_{i,j=1}^{3} \pi_i p_{ij}^{(k)} \ln \left( \frac{p_{ij}^{(k)}}{p_{ji}^{(k)}} \right) = \frac{1}{2} \ln \left( \frac{1-3d}{d} \right).
\]

**6. Conclusions.** In the era of BIG DATA, a quote from [8], which has partially inspired our work, is still very relevant: “In the study of deterministic dynamical systems, environmental noise tends to be suppressed or, at most, plays a secondary role, whereas in the study of statistics the deterministic dynamic kernel of the random generating mechanism tends to give way to the more macroscopic characterization such as the mean functions, the covariance functions, the spectral functions and so on.”

In the present work, we have seen that environmental noise and the dynamic kernel of random generating mechanism, which we termed extrinsic and intrinsic noises respectively, actually play very different roles in a stochastic dynamics. Random dynamical system (RDS) is a more refined stochastic dynamic description of...
random processes. For single sample paths generated by a Markov chain, there are many compatible i.i.d. RDS. In the present work, we have termed them RDS representations of the Markov chain. Since the representation is not unique, one can characterize an RDS representation in terms of the metric entropy of the i.i.d. system. It can be shown that the Shannon-Khinchin entropy of the Markov chain is the lower bound of the metric entropies of all the RDS representations. On the other hand, the RDS representation with the maximum metric entropy has a very special feature that if currently synchronization has not yet occurred between two trajectories, then their next transitions are independent.

One of the particularly attractive features of the discrete state formulation of RDS and MC is the possibility of various in-depth investigations of complex biological dynamics using a broad spectum of mathematical tools that are accessible to mathematical biologists. For example, applying the theory of random matrices [30, 35]. In essence, by focusing on discrete state space, our approach described here has reverted to the original motivation of Oseledec’s multiplicative ergodic theorem [3], which is at the heart of RDS. Other possible directions include the theory of irreversible Markov process and its entropy production [53, 14], convex analysis [43], and maybe even group theoretical investigations. Also, understanding chaotic dynamics in a discrete setting [27] could have implications to uncertainty quantification of numerical algorithms [17]. Finally, but not the least, recent development in stochastic thermodynamics has suggested that the fundamental entities in a nonequilibrium stochastic dynamics are cycles rather than states [40, 18]. The discrete formulation of RDS will provide ample opportunity to explore this insight.

Acknowledgments. We thank Professors Renato Feres, Gary Froyland, Da-Quan Jiang, Yao Li, Eric Shea-Brown, and Eli Shlizerman for help discussions. H. Q. particularly thanks Drs. Kevin Lin (University of Arizona), Wen-Xian Shen (Auburn University), both of whom he first met at Institute for Mathematics and its Applications Workshop on Stochastic Modeling of Biological Processes (May 13–17, 2013), Jian-Sheng Xie (Fudan University) and Cheng-Hung Chang (National Chiao-Tung University), for their keen interests, helpful discussions, and continuous encouragements in the initial phase of this project.

REFERENCES

[1] L. J. S. Allen, An Introduction to Stochastic Processes with Applications to Biology, 2nd edition, CRC Press, Boca Raton, FL, 2011.
[2] L. Arnold, Random Dynamical Systems, Springer, New York, 1998.
[3] L. Arnold and H. Crauel, Random dynamical systems, in Lyapunov Exponents, (eds. L. Arnold, H. Crauel and J.-P. Eckmann), Springer, Berlin, 1486 (2006), 1–22.
[4] P. H. Baxendale, Statistical equilibrium and two-point motion for a stochastic flow of diffeomorphisms. Spatial stochastic processes, Progr. Probab., 19 (1991), 189–218.
[5] R. Bhattacharya and M. Majumdar, Random Dynamical Systems: Theory and Applications, Cambridge Univ. Press, U.K., 2007.
[6] G. Birkhoff, Three observations on linear algebra, Univ. Nac. Tucumán. Revista A, 5 (1946), 147–151.
[7] R. M Blumenthal and H. K. Corson, On continuous collections of measures, Proc. 6th Berkeley Symp. on Math. Stat. and Prob., 2 (1972), 33–40.
[8] K.-S. Chan and H. Tong, Chaos: A Statistical Perspective, Springer, New York, 2001.
[9] Y.-D. Chen, Asymmetry and external noise-induced free energy transduction, Proc. Natl. Acad. Sci. U.S.A., 84 (1987), 729–733.
[10] T. Downarowicz, Entropy in Dynamical Systems, Cambridge Univ. Press, UK, 2011.
[11] S. P. Ellner and J. Guckenheimer, Dynamic Models in Biology, Princeton Univ. Press, NJ, 2006.
[12] G. Froyland, Extracting dynamical behavior via Markov models, Nonlinear dynamics and statistics (Cambridge, 1998), Birkhäuser Boston, Boston, MA, (2001), 281–321.
[13] G. Gallavotti, Statistical Mechanics: A Short Treatise, Springer, New York, 1999.
[14] H. Ge, M. Qian and H. Qian, Stochastic theory of nonequilibrium steady states (Part II): Applications in chemical biophysics, Phys. Rep., 510 (2012), 87–118.
[15] B. Hasselblatt and A. Katok, Principal structures, in Handbook of Dynamical Systems, vol. 1A (eds. B. Hasselblatt and A. Katok), Elsevier, Amsterdam, 1 (2002), 1–203.
[16] W. Horsthemke and R. Lefever, Noise-Induced Transitions: Theory and Applications in Physics, Chemistry, and Biology, Springer, New York, 1984.
[17] E. Isaacson and H. B. Keller, Analysis of Numerical Methods, Dover, New York, 1966.
[18] S. L. Kalpazidou, Cycle Representations of Markov Processes, 2nd edition, Springer, New York, 2006.
[19] M. Keane, Ergodic theory and subshifts of finite type in Ergodic theory, Symbolic Dynamics and Hyperbolic Spaces (eds. T. Bedford, M. Keane and C. Series), Oxford, (1991), 35–70.
[20] A. I. Khinchin, Mathematical Foundations of Information Theory, Dover, New York, 1957.
[21] Yu. Kifer, Ergodic Theory of Random Transformations, Birkhäuser, Basel, 1986.
[22] Yu. Kifer and P.-D. Liu, Random dynamics, in Handbook of Dynamical Systems, (eds. B. Hasselblatt and A. Katok), Elsevier, Amsterdam, 1 (2006), 379–499.
[23] A. Lasota and M. C. Mackey, Chaos, Fractals, and Noise: Stochastic Aspects of Dynamics, 2nd edition, Springer-Verlag, New York, 1994.
[24] V. Lecomte, C. Appert-Rolland and F. van Wijland, Thermodynamic formalism for systems with Markov dynamics, J. Stat. Phys., 127 (2007), 51–106.
[25] T. Liggett, The coupling technique in interacting particle systems, In Doeblin and Modern Probability, (ed. H. Cohn), AMS, Providence, 149 (1993), 73–83.
[26] K. K. Lin, Stimulus-response reliability of biological networks, in Nonautonomous and Random Dynamical Systems in Life Sciences (eds. P. Kloeden and C. Poetzsche), Springer, New York, 2102 (2012), 135–161.
[27] D. Lind and B. Marcus, An Introduction to Symbolic Dynamics and Coding, Cambridge Univ. Press, U.K., 1995.
[28] X. Mao and C. Yuan, Stochastic Differential Equations with Markovian Switching, Imperial College Press, UK, 2006.
[29] M. Marcus, H. Minc and B. Moyls, Some results on non-negative matrices, J. Res. Nat. Bur. Standards Sec. B, 65 (1961), 205–209.
[30] M. L. Mehta, Random Matrices, 3rd edition, Academic Press, New York, 2004.
[31] H. Minc, Nonnegative Matrices, John Wiley & Sons, New York, 1988.
[32] J. A. Morrison and J. McKenna, Analysis of some stochastic ordinary differential equations, in Stochastic Differential Equations, SIAM-AMS Proc., Vol. 6 (eds. J. B. Keller and H. P. McKean), Amer. Math. Soc., Providence, R.I., (1973), 97–161.
[33] J. D. Murray, Mathematical Biology: I. An Introduction, 3rd edition, Springer, New York, 2002.
[34] D. S. Ornstein, Ergodic theory, randomness, and “chaos”, Science, 243 (1989), 182–187.
[35] Y. Peres, Analytic dependence of Lyapunov exponents on transition probabilities, in Lyapunov Exponents, (eds. L. Arnold, H. Crauel and J.-P. Eckmann), Springer, Berlin, 1486 (1991), 64–80.
[36] M. A. Pinsky, Lectures on Random Evolution, World Scientific, Singapore, 1991.
[37] H. Qian and J. A. Schellman, Helix-coil theories: A comparative studies for finite length polypeptides, J. Phys. Chem., 96 (1992), 3987–3994.
[38] H. Qian, The mathematical theory of molecular motor movement and chemomechanical energy transduction, J. Math. Chem., 27 (2000), 219–234.
[39] H. Qian, Cooperativity in cellular biochemical processes: Noise-enhanced sensitivity, fluctuating enzyme, bistability with nonlinear feedback, and other mechanisms for sigmoidal responses, Annu. Rev. Biophys., 41 (2012), 179–204.
[40] H. Qian, S. Kjelstrup, A. B. Kolomeisky and D. Bedeaux, Entropy production in mesoscopic stochastic thermodynamics — Nonequilibrium steady state cycles driven by chemical potentials, temperatures, and mechanical forces, J. Phys. Cond. Matt. 28 (2016), 153004. arXiv:1601.04018
Appendix A. The proof of Theorem 2.1.

Proof. We will use the induction on the number of positive entries in the transition matrix $M$, denoted as $\pi(M)$. If $\pi(M) = n$, then $M$ is a deterministic transition matrix and the theorem holds with $i = 1$.

Assume $\pi(M) > n$ and the theorem is true for all matrices in $\Omega_n$ with less than $\pi(M)$ positive entries. For every Markov transition matrix, there exists a deterministic transition matrix $P = P_{i_1,i_2,\ldots,i_n}$ with the positive coefficient $a = \min\{M_{j,k} : M_{j,k} > 0, j, k = 1, 2, \ldots, n\}$ such that $M - aP$ is a nonnegative matrix. Clearly, $a < 1$ because $a = 1$ would imply that $M$ has 1 in positions $(k, i_k)$ for each row. Assume the minimum is attained at row $t$, that is, $k_t = t$.

The matrix $B = \frac{1}{1-a}(M - aP)$ is a Markov transition matrix, because $B$ is nonnegative matrix and

$$\sum_{j=1}^{n} B_{ij} = \sum_{j=1}^{n} \frac{1}{1-a} (M_{ij} - aP_{ij}) = (1-a)/(1-a) = 1 \text{ for } i = 1, 2, \ldots, n$$

Now $\pi(B) \leq \pi(M) - 1$, since $B$ has zero entries where $M$ has zeros as well, and in addition, $B_{t,i_t} = 0$. By induction,

$$B = \sum_{j=1}^{N} \beta_j P_j$$

where $\beta_j$ are the coefficients to be determined.
where \( P_j \) are deterministic matrices, \( \beta_j \) are nonnegative and \( \sum_j \beta_j = 1 \). Therefore,
\[
M = (1 - a)B + aP = \left( \sum_{j=1}^{N} (1 - a)\beta_j P_j \right) + aP = \sum_{i=1}^{N} a_i P_i
\]
Where \( P_{N+1} = P \), \( a_{N+1} = a \), \( a_i = (1 - a)\beta_i \) for \( i = 1, 2, \ldots, n \). Obviously, \( a_i \) are nonnegative. Moreover,
\[
\sum_{i=1}^{N} a_i = \sum_{i=1}^{N} (1 - a)\beta_i + a = 1 - a + a = 1.
\]

It is easy to see that a deterministic matrix cannot be expressed as a convex combination of other deterministic matrices. Thus deterministic matrices are the vertices of this polyhedron.

**Appendix B. Metric entropy.** In this section, we review some important concepts in connection with metric entropy. This material can be found in standard textbooks [21, 50, 19, 10]; it is presented for the convenience of the readers.

Let \( T \) be a measure preserving transformation of a probability space \((\mathcal{S}, \mathcal{F}, \mu)\). A good motivation for the notion of metric entropy \( h_T \), associated with \( T \), is in terms of measurements. A **measurement** is a finite partition of the space \( \mathcal{S} \), \( P = \{P_1, \cdots, P_k\} \). All these measurable sets \( P_i \) are disjoint and their union is \( \mathcal{S} \). Now consider a finite portion of an orbit, with length \( n \), generated by \( T \), starting from an initial condition \( s \in \mathcal{S} \),
\[
s, T(s), T^2(s), \cdots, T^{n-1}(s).
\]
Each of the points \( T^i(s) \) belongs to exactly one of the sets of the partition \( P \),
\[
s \in P_{k_0}, T(s) \in P_{k_1}, \cdots, T^{n-1}(s) \in P_{k_{n-1}}.
\]
We say \( k = \{k_0, k_1, \cdots, k_{n-1}\} \) the **address** of \( s \) with respect with the partition \( P \). It is possible that another orbit will have the same address as \( k \) with respect with the partition \( P \) so we can collect all initial points such their orbits have the same address \( k \).
\[
P^n(k) = \{ s \in \mathcal{S} : \text{address of } s = k \} \quad \text{(B1)}
\]
In fact, \( P^n = \{P^n(k) : k \text{ is any address of length } n\} \) is also a partition of \( \mathcal{S} \). Moreover, it can be shown that \( P^n \) is the join of the partitions \( P, T^{-1}P, \ldots, T^{-(n-1)}P \).
One likes to quantify the amount of information in an address of an orbit with length \( n \). This is given by the Gibbs-Shannon entropy
\[
H(P^n) = - \sum_{j=1}^{m} \mu(P^n_j) \ln \mu(P^n_j) \quad \text{(B2)}
\]
where \( m \) is the number of partitions of \( P^n \).

The amount of information per unit is then \( \frac{1}{n} H(P^n) \). Therefore, the information per unit in a measurement \( P \) is defined as \( H(P, T) = \lim_{n \to \infty} \frac{1}{n} H(P^n) \) and the metric entropy is the supremum of this value over all possible finite measurements.
$h_T = \sup_P H(P, T)$. A discovery due to Sinai helps the computation: The supremum is attained for all partitions are generators. A generator is a partition $P$ such that two different points of $S$ have a different address.

In the case of a Bernoulli sequence of only finite states, e.g., “symbols”, with probability $(p_1, \ldots, p_m)$, the generator is simply the natural partition $P = (1, 2, \ldots, m)$ since an address of $s$ is $(s_0, s_1, s_2, \ldots)$ which is $s$ itself. For each address of length $n$, $k = (k_0, k_1, \ldots, k_{n-1})$, we will have $\mu(P^n(k)) = p_{k_0} p_{k_1} \cdots p_{k_{n-1}}$ and hence

$$H(P^n) = - \sum_k \mu(P^n(k)) \log \mu(P^n(k)) = -n \left( \sum_{i=1}^m p_i \ln p_i \right).$$

Therefore the metric entropy is $h(T) = - \sum_{i=1}^m p_i \ln p_i$. Bernoulli trial is a sequence of i.i.d. random variables; one can similarly derive for the metric entropy of a Markov chain. For Markov chain with transition probability $p_{ij}$ and invariant distribution $\pi_i$, $h = - \sum_{i,j=1}^m \pi_i p_{ij} \ln p_{ij}$. See also Eq. C9.

### Appendix C. Sequence space and thermodynamic formalism.

Consider an aperiodic, irreducible, and recurrent Markov chain with transition probability $p_{ij}$ and stationary distribution $\{\pi_i\}, i, j \in S$ with $|S| = n$. If one considers the space of all the possible sequences generated by the Markov process, then there is an isomorphism between the problem of stochastic dynamical systems and the technique used in statistical thermodynamics. This is known as the thermodynamic formalism [24]. Both metric entropy and topological entropy emerge in this formalism, as follows.

One calls $E_{ij} \overset{\text{def.}}{=} - \ln p_{ij}$ “interaction energy” between nearest neighbors in state $i$ and $j$ in a given sequence $\sigma_\ell = i_0 i_1 \cdots i_{\ell-1}$ of length $\ell$. Then the total energy of the sequence $\sigma_\ell$

$$E_{\sigma_\ell} = \sum_{k=1}^\ell E_{i_k i_{k-1}} ,$$

and

$$\Xi(\beta) = \sum_{\sigma_\ell}^\text{all} e^{-\beta E_{\sigma_\ell}}$$

is called a “partition function” and $\Lambda(\beta) \overset{\text{def.}}{=} - \frac{1}{\beta} \ln \Xi(\beta)$ is called the “free energy”. The “all $\sigma_\ell$” in the summation includes forbidden sequences with $p_{ij} = 0$, which corresponds to $E_{ij} = -\infty$. Note that when $\beta \to 0$, all $\beta E_{ij} = 0$ for those with $p_{ij} \neq 0$, but still $\beta E_{ij} = -\infty$ for those with $p_{ij} = 0$. Therefore, the matrix $e^{-\beta E_{ij}}$ becomes the A matrix in (14) when $\beta \to 0$.

To compute the partition function $\Xi$ in (C2), one can use a technique based on matrix multiplication. In molecular biophysics of proteins and DNA, this has been applied to the theory of helix-coil transition [37]. It can be easily shown that

$$\Xi(\beta, \ell) = \left( 0, \ldots, \underbrace{1, \ldots, 1}_{i_0} \right) Q^\ell \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} ,$$

in which matrix $Q$ has elements

$$Q_{ij} = e^{-\beta E_{ij}} = e^{\beta \ln p_{ij}} .$$
Entropy in thermodynamics is defined as
\[
S \overset{\text{def.}}{=} \ln \Xi - \beta \frac{\partial \ln \Xi}{\partial \beta} = \frac{\partial (\beta^{-1} \ln \Xi)}{\partial \beta^{-1}},
\]
(C5)
and
\[
U \overset{\text{def.}}{=} \sum_{\sigma_t} E_{\sigma_t} e^{-\beta E_{\sigma_t}} = -\frac{\partial \ln \Xi}{\partial \beta}.
\]
(C6)
is called “mean internal energy”. \(U - \beta^{-1} S = \Lambda\) then is free energy. We note that when \(\beta = 1\), \(Q_{ij} = p_{ij}\) and \(\Xi(1, \ell) = 1\). Therefore when \(\beta = 1\), \(\Lambda = 0\) and \(U = S\) \(\forall \ell\).

C.1. Thermodynamic limit. Taking \(\ell \to \infty\) is called “thermodynamic limit”, in which one has the free energy
\[
\Lambda(\beta) = -\beta^{-1} \ln \Xi(\beta, \ell) \simeq -\ell \beta^{-1} \ln \lambda_Q(\beta),
\]
(C7)
where \(\lambda_Q(\beta)\) is the largest eigenvalue of \(Q\). Furthermore,
\[
S(\beta) = \ell \ln \lambda_Q(\beta) - \ell \beta \frac{\partial \ln \lambda_Q(\beta)}{\partial \beta}.
\]
(C8)
In fact, \(\frac{S(0)}{\ell}\) is the topological entropy of the Markov chain, and \(\frac{S(\infty)}{\ell} = 0\). At \(\beta = 1\), eigenvalue perturbation calculation yields
\[
\frac{S(1)}{\ell} = -\sum_{i,j \in S} \pi_i p_{ij} \ln p_{ij},
\]
(C9)
which is the metric entropy.

Received June 2015; revised February 2016.

E-mail address: yexf308@u.washington.edu
E-mail address: yuewang@u.washington.edu
E-mail address: hqian@u.washington.edu