THE CUTOFF PHENOMENON AND MIXING BY CHAOTIC MAPS

T. LIANG* AND M. WEST†

Abstract. We extend the definition of a cutoff from finite Markov chains to the evolution of probability distributions by the Perron-Frobenius operator of 1D maps. We prove that if the map has full shift symbolic dynamics then, for appropriately chosen initial distributions, such an evolution exhibits a total variation cutoff. Moreover, the initial distributions can be chosen so that the limiting evolution has the same normal shape as is found in many Markov chain problems.

Key words. Cutoff phenomenon, chaotic mixing, symbolic dynamics.

AMS subject classifications. 37A25, 60J05, 37B10

1. Introduction.

1.1. Cutoff phenomenon. How many riffle shuffles are required to adequately mix 52 cards? This question has been answered by Bayer and Diaconis [4]. It is quite surprising that the cards are highly ordered for the first several shuffles (up to 6 shuffles for 52 cards) and then become randomized almost abruptly. This kind of sharp change in some measure of order has been discovered in many finite Markov Chains, and is termed the *cutoff phenomenon* [1]. A good review of the cutoff phenomenon for random walks on finite groups has been given by Saloff-Coste [17].

In general, proving the existence of a cutoff in a certain sequence of Markov chains is a difficult task. It usually requires some clever insight about the system, such as the famous example of the "rising sequence" formulation of the riffle shuffle problem [4]. Showing cutoff numerically is equally difficult because it usually entails the simulation of very large Markov chains. As a result, our knowledge about cutoff phenomena is limited. Diaconis [6, 8] asks, "How widespread is the cutoff phenomenon for families of finite ergodic Markov chains and how can one recognize it?" In this paper we provide a new class of systems that exhibit cutoff phenomena using chaotic maps.

1.2. Chaotic map mixing. In other work, the mixing process of chaotic maps or flows has been observed to have similar multistage behavior [20, 19, 22] to that of cutoff. Think of cream being stirred into a cup of coffee: initially the cream is stretched and folded, but remains distinct from the coffee, but then they rapidly mix and become indistinguishable. Chaotic mixing have been observed in experiments [16, 23], can be simulated by large-scale computations [13, 22], and have been applied in, for example, microfluidic mixing channel design [14, 25, 15], random search strategies and generating random numbers.

A widely observed phenomenon in the chaotic mixing process with small diffusion is the two- or three-stage transition [20, 10, 3]. The map does not mix a scalar function with a constant rate in general. When the variance of the scalar function is measured during the mixing process, one generally observes a relatively flat decay followed by a super-exponential change, before it finally tends to an exponential decay. We are interested in when these transitions happen, why they happen, and how to predict the slope of the exponential region. A good review and physical interpretation is given by Thiffeault [19].

*Department of Aeronautics and Astronautics, Stanford University, Stanford, CA 94305, USA (hartree50@gmail.com).

†Department of Mechanical Science and Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA (mwest@illinois.edu).

1.3. Cutoff for chaotic maps. In this article, we build an explicit relationship between the cutoff phenomenon for finite Markov chains and the chaotic mixing process by developing a generalized symbolic dynamics for chaotic maps. The result shows cutoff in the mixing process can be produced by chaotic maps with full shift symbolic dynamics and certain initial conditions. In related work [12], we numerically investigate the mixing processes induced by the 2D standard map and provide numerical evidence that it exhibits cutoff behavior.

The outline of this paper is as follows. In Section 2, we briefly review cutoff phenomena and point out how a simple chaotic map can produce cutoff. Section 3 introduces symbolic dynamics for chaotic maps and defines a new object called a stochastic symbol sequence, which serves as our main tool to build the bridge between the cutoff phenomenon and chaotic mixing. The main result is given in Section 4, and finally conclusions are in Section 5.

2. Background.

2.1. Cutoff phenomenon. The cutoff phenomenon was discovered by Aldous, Diaconis, and Shahshahani [2, 1, 9], and formalized by Aldous and Diaconis [6, 2]. Some of the most interesting cases are found in random walks on finite groups with the measure of total variation distance, and most known Markov Chains that present cutoffs can be shown to belong to this category [17]. Here we state the definition of a cutoff given by Diaconis [8], but we do not require the probability spaces to be finite.

Assume that to any set Ω and any pair of probability measures ω, $\bar{\omega}$ on Ω is associated a real number $D(\omega, \bar{\omega})$ such that $D(\omega, \bar{\omega}) \in [0, 1]$,

$$\max_{\Omega, \omega, \bar{\omega}} D(\omega, \bar{\omega}) = 1, \tag{2.1}$$

and $D(\omega, \bar{\omega}) = 0$ if and only if $\bar{\omega} = \omega$. Consider a sequence of probability spaces $(\Omega_n, \bar{\omega}_n)_{n=1}^{\infty}$, each equipped with a sequence $(\omega_n^k)_{k=0}^{\infty}$ of probability measures, such that

$$\lim_{k \to \infty} D(\omega_n^k, \bar{\omega}_n) = 0 \tag{2.2}$$

for each n. The definition of a cutoff follows.

DEFINITION 2.1. *A family* $(\Omega_n, \bar{\omega}_n, (\omega_n^k)_{k=0}^{\infty})_{n=1}^{\infty}$ *presents a* D-cutoff *if there exists a sequence* $(t_n)_{n=1}^{\infty}$ *of positive reals such that, for any* $\epsilon \in (0, 1)$,

1. $\lim_{n \to \infty} D(\omega_n^{k_n}, \bar{\omega}_n) = 0$ *if* $k_n > (1 + \epsilon) t_n$;
2. $\lim_{n \to \infty} D(\omega_n^{k_n}, \bar{\omega}_n) = 1$ *if* $k_n < (1 - \epsilon) t_n$.

The numbers t_n *are called* cutoff times.

We are interested in sequences of probability distributions generated by Markov chains. That is, for each n we will take a Markov chain P_n on Ω_n and an initial distribution ω_n^0, so we can then define the sequence by $\omega_n^{k+1} = P_n \omega_n^k$. The distribution $\bar{\omega}_n$ will be an invariant distribution of P_n that the sequence limits to as $k \to \infty$.

To see if these sequences exhibit a cutoff, we consider $D(\omega_n^k, \bar{\omega}_n)$ as a trajectory in k for each n. For cutoff, the trajectories should remain near 1 initially, then transition to 0, where they then remain. The cutoff times t_n characterize the iteration when the transition is occurring. If the transition time divided by the cutoff time tends to zero as $n \to \infty$ then there is a cutoff. Equivalently, the trajectories can be rescaled by the transition times t_n, and we have a cutoff if the rescaled trajectories limit to a step function as $n \to \infty$.

The distance function D that we use is the total variation distance, as follows.

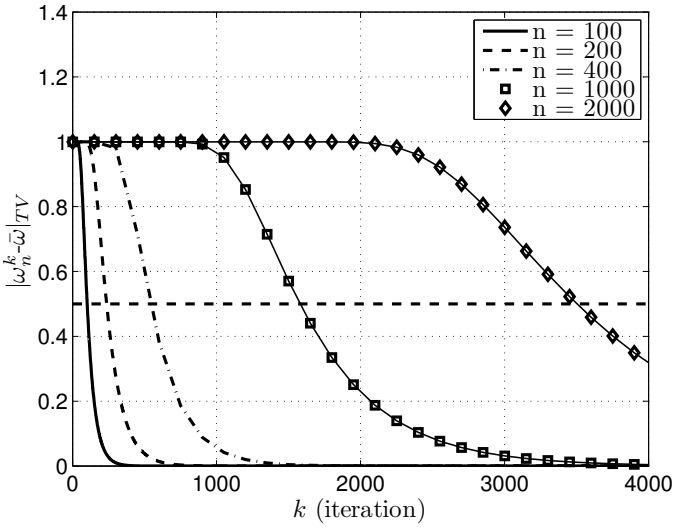

FIG. 2.1. *Sample trajectories of $|\omega_n^k - \bar{\omega}|_{TV}$ versus iteration number k for the random walk on a n-dimensional hypercube described in Example 1. As n increases, the trajectory stays close to* 1 *for more iterations before it drops to zero.*

DEFINITION 2.2. *For two probability distributions μ and ν on Ω, the* total variation distance *(TV) between μ and ν is*

$$|\mu - \nu|_{TV} = \frac{1}{2} \int_\Omega |\mu - \nu|.$$

The shape of a cutoff can be characterized in several ways [5]. We are specifically interested in the "normal" shape, which is addressed by the following well-known cutoff example.

EXAMPLE 1 (Random walk on an n-dimensional hypercube [7]). *Let Ω_n be the vertices of an n-dimensional hypercube. A particle starts at $\mathbf{0}$ and moves to one of its nearest neighbors (or stays fixed) with equal probability at each step. We take ω_n^k to be the probability distribution of the particle's position after k steps, and ω_n is the uniform distribution. Let $k = \frac{1}{4} n \log n + cn$ for a fixed $c \in \mathbb{R}$. Then, as $n \to \infty$,*

$$|\omega_n^k - \bar{\omega}|_{TV} \sim \operatorname{erf}\left(\frac{e^{-2c}}{\sqrt{8}}\right). \tag{2.3}$$

Some example evolutions for this system are shown in Figure 2.1. ◊

In the above example, the cutoff shape is called "normal" because of the relationship between the normal distribution and the error function. In fact, most interesting cutoff examples have normal shapes. When $n \to \infty$ the evolution of the total variation distance to the uniform distribution can be calculated by evaluating the total variation between two normal distributions. This fact is crucial in our later discussion.

Proving that a given system exhibits cutoff is typically challenging, and only a small number of such proofs are known. In the case of the random walk on a hypercube, an elegant method of proof uses symmetry reduction to the so-called Ehrenfest urn problem, which has only $n + 1$ states [7]. In this case, the invariant distributions $\bar{\omega}_n$ are binomial with certain means, so the distributions ω_n^k have a shape

that limits to a binomial, and a mean that limits to the mean of $\bar{\omega}_n$. Cutoff phenomena involve the interaction of these two processes [21]. For large n, binomial distributions are well approximated by normal distributions, giving the behavior (2.3).

2.2. Mixing by chaotic maps. To extend the study of cutoff phenomena to the chaotic map setting, we begin by recalling how the Perron-Frobenius operator evolves probability distributions for a map.

We take a space Ω with a map $S : \Omega \to \Omega$, and we let Γ be the set of $\omega \in L^1\Omega)$ that are probability distributions on Ω. That is, $\omega \in \Gamma$ is non-negative and has a 1-norm of one.

DEFINITION 2.3. *Consider a measure space $(\Omega, \mathcal{A}, \mu)$ with a map $S : \Omega \to \Omega$. Let $\omega \in \Gamma$, and suppose that for every $A \in \mathcal{A}$ the operator $P : \Gamma \to \Gamma$ satisfies*

$$\int_A (P\omega)(x)\mu(dx) = \int_{S^{-1}(A)} \omega(x)\mu(dx). \tag{2.4}$$

Then P is the Perron-Frobenius operator *associated to S.* The Perron-Frobenius operator can be thought of as evolving probability distributions by S.

EXAMPLE 2 (Tent map cutoff.). *Consider the tent map on $\Omega = [0, 1]$ defined by*

$$S_{tent}(x) = 1 - 2\left|x - \frac{1}{2}\right| \tag{2.5}$$

with the initial distributions in Γ given by

$$\omega_n^0 = \begin{cases} \frac{1}{\mu_n} & \text{if } x \le \mu_n, \\ 0 & \text{otherwise,} \end{cases} \tag{2.6}$$

where $\mu_1 = 1$, and $\mu_{n+1} = \mu_n/2$. The Perron-Frobenius operator of the tent map can be explicitly computed to give

$$\omega_n^{k+1}(x) = (P_{tent}\omega_n^k)(x) = \frac{1}{2}\left(\omega_n^k\left(\frac{x}{2}\right) + \omega_n^k\left(1 - \frac{x}{2}\right)\right). \tag{2.7}$$

The invariant distribution $\bar{\omega}$ of the tent map is uniform. Let $\nu_n^k = |\omega_n^k - \bar{\omega}|_{TV}$; we find that

$$\nu_n^k = \begin{cases} 1 - 2^{1+k-n} & \text{if } k \le n-1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.8}$$

The trajectories of ν_n^k with varying n are shown in Figure 2.2. It shows a cutoff. $\Diamond$

Hence we state the following simple theorem (a more general version is proved in Section 4).

THEOREM 2.4 (Tent map cutoff). *The family $(\Omega, \bar{\omega}, (\omega_n^k)_{k=0}^{\infty})_{n=1}^{\infty}$, where $\Omega = [0, 1]$, $\bar{\omega}$ is uniform in $[0, 1]$ and ω_n^k are defined by (2.6) and (2.7), presents a total variation-cutoff.*

Proof. This is a special case of Theorem 4.1 with $m = 1$. □

This example above shows that the tent map can present sharp changes in total variation distance. We will generalize this observation to a set of initial distributions and all 1D chaotic maps that have full shift symbolic dynamics.

3. Symbolic Dynamics and Stochastic Symbol Sequence.

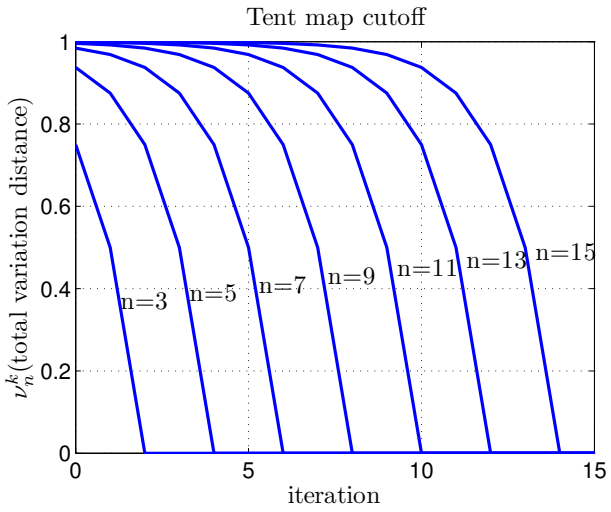

FIG. 2.2. *Trajectories of ν_n^k versus iteration number k for the tent map described in Example 2. This system presents a cutoff.*

3.1. Symbolic Dynamics. We briefly recall the definition of symbolic dynamics for chaotic maps. We focus on 1D chaotic maps, whose symbolic dynamics are semi-infinite sequences.

Let $\{L, R\}$ be a set of emphsymbols and define Σ to be the collection of all semi-infinite sequences of symbols. That is, $s \in \Sigma$ implies

$$s = \{.s_0 s_1 \cdots s_n \cdots\} \tag{3.1}$$

with $s_i \in \{L, R\}$ for all i. We refer to Σ as the space of semi-infinite sequences of two symbols. We consider a map $\sigma : \Sigma \to \Sigma$, called the shift map, defined by $\sigma(s)_i = s_{i+1}$. For $s = \{.s_0 s_1 \cdots s_n \cdots\}$,

$$\sigma(s) = \{.s_1 s_2 \cdots s_n \cdots\}. \tag{3.2}$$

That is, the shift operator σ simply deletes the first element of the sequence. There are rich results about the relationship between symbolic dynamics and chaotic maps [24, 11]. Given a map S with invariant set Λ, we partition $\Lambda = \Lambda_L \cup \Lambda_R$ into a disjoint union. Then we define the function $\phi(x) : \Lambda \to \Sigma$, which maps a point in $x \in \Lambda$ to a semi-infinite sequence, by

$$\phi(x)_i = \begin{cases} L & \text{if } S^i(x) \in \Lambda_L \\ R & \text{if } S^i(x) \in \Lambda_R. \end{cases} \tag{3.3}$$

With this definition we see that

$$\sigma \circ \phi = \phi \circ S. \tag{3.4}$$

Then, roughly speaking, we establish that ϕ is a homeomorphism, so S acting on Λ and σ acting on Σ are topologically conjugate. This relationship can be used to precisely define the sensitivity to initial conditions of chaotic maps.

All the results we derive in the next section for semi-infinite stochastic symbol sequences can be extended to bi-infinite sequences without difficulty.

3.2. Stochastic Symbol Sequences. Since our goal is to study how the probability density is evolved by a map, and symbolic dynamics itself does not provide this information, we need a new object which we call a stochastic symbol sequence.

DEFINITION 3.1. *A* stochastic symbol sequence δ *is a semi-infinite sequence*

$$\delta = \{.\delta_0\delta_1 \cdots \delta_n \cdots\}, \tag{3.5}$$

where $\delta_i \in [0,1]$ *for each* $i \geq 0$. *We denote the set of all such sequences by* Δ.

For the symbol L we define the map $\psi^L : \Gamma \to \Delta$, which maps a probability distribution ω on Λ to a stochastic symbol sequence $\delta^L = \psi^L(\omega)$ defined by

$$\delta_i^L = \int_{\Lambda_L} (P^i\omega)(z)\, dz, \text{ for all } i, \tag{3.6}$$

where P is the Perron-Frobenius operator of map S. That is, if $X \in \Lambda$ is a random variable with probability distribution ω, then

$$\delta_i^L = \text{Prob}(\phi(X)_i = L) = \text{Prob}(S^i(X) \in \Lambda_L). \tag{3.7}$$

We see that the stochastic symbol sequence δ^L gives the probabilities of a random point X having $S^i(X)$ in Λ_L, whereas the regular symbol sequence $s = \phi(x)$ tells us directly whether a fixed point x has $S^i(x)$ in Λ_L.

We can also define ψ^R and δ^R similarly for the symbol R, giving $\delta_i^R = 1 - \delta_i^L$. It is also straightforward to extend these definitions to more than two symbols, but we will not do so here. Later, we will write simply δ for δ^L.

It is clear that the map $\psi^L : \Gamma \to \Delta$, as defined in (3.6), maps a probability distribution $\omega \in \Gamma$ to the stochastic symbol sequence δ^L uniquely. Moreover, we apply the shift σ to the stochastic symbol sequence space Δ, so that

$$\sigma(\delta^L)_i = \delta_{i+1}^L. \tag{3.8}$$

and we see that shifts on stochastic symbol sequences correspond to evolution by the Perron-Frobenius operator as follows.

LEMMA 3.2. *The Perron-Frobenius operator* P *on the space* Γ *of probability distributions is related to the shift map* σ *on the space* Δ *of stochastic symbol sequences by*

$$\psi^L \circ P = \sigma \circ \psi^L, \tag{3.9}$$

where ψ^L *is defined by (3.6).*

Proof. By definition, $\delta_{i+1}^L(\omega) = \int_{\Lambda_L} (P^{i+1}\omega)(z)dz = \int_{\Lambda_L} (P^i(P\omega))(z)dz = \delta_i^L(P\omega)$.

□

However, unlike the map ϕ from points to (non-stochastic) symbol sequences, the function ψ mapping probability distributions to stochastic symbol sequences is not invertible. There are many ωs that map to the same δ^L. To resolve this problem, we define a subset $\bar{\Gamma} \subset \Gamma$:

$$\bar{\Gamma} = \left\{ \omega \in \Gamma \;\middle|\; \omega(x) = \lim_{n\to\infty} \prod_{i=0}^{n} \beta_i^{\phi(x)_i} \text{ for some } \beta^L \in \Delta \text{ and } \beta_i^R = 1 - \beta_i^L \text{ for all } i \right\}. \tag{3.10}$$

Recall that $\phi(x)_i$ is the i-th component of the symbol sequence of x associated with the map S, so having full shift symbolic dynamics is a necessary condition for having $\bar{\Gamma}$.

The following lemma justifies the existence of ψ^{-1} acting on $\bar{\Gamma}$.

LEMMA 3.3. *For $\omega \in \bar{\Gamma}$ corresponding to $\beta^L \in \Delta$ and $\delta^L = \psi^L(\omega)$, we have*

$$\delta^L = \beta^L. \tag{3.11}$$

Proof. By definition, $\delta_k^L = \int_{\Lambda_L} P^k \omega(x) dx = \int_{S^{-k}(\Lambda_L)} \omega(x) dx$, so from (3.10) we have

$$\delta_k^L = \lim_{n\to\infty} \sum_{\substack{s\in\Sigma \\ s^k = L}} \prod_{i=0}^{n} \beta_i^{s_i} = \beta_k^L \lim_{n\to\infty} \sum_{s\in\Sigma} \prod_{\substack{i=0 \\ i\neq k}}^{n} \beta_i^{s_i} = \beta_k^L.$$

□

Hence, the map $\psi^L : \Gamma \to \Delta$ is invertible on the subset $\bar{\Gamma} \subset \Gamma$. From now on, we restrict our attention to the space of probability distributions $\bar{\Gamma}$. It is also easy to check that

$$P\omega \in \bar{\Gamma} \text{ if } \omega \in \bar{\Gamma}, \tag{3.12}$$

and since ψ^L restricted to $\bar{\Gamma}$ is invertible we have

$$P = (\psi^L)^{-1} \circ \sigma \circ \psi^L. \tag{3.13}$$

just like (non-stochastic) symbolic dynamics. This means that the Perron-Frobenius operator P and the shift operator σ are conjugate in the spaces $\{\bar{\Gamma}, \Delta\}$.

To further characterize the space $\bar{\Gamma}$, we give the following lemma.

LEMMA 3.4. *For $X \in \Lambda$ a random variable with probability distribution $\omega \in \bar{\Gamma}$, and any symbol sequence $s \in \Sigma$, then for any i, j with $i \neq j$, the events*

$$\phi(X)_i = s_i \text{ and } \phi(X)_j = s_j \tag{3.14}$$

are independent.

Proof. Let $\delta^L = \psi^L(\omega)$ and $\delta^R = \psi^R(\omega)$, so

$$\text{Prob}(\phi(X) = s) = \lim_{n\to\infty} \prod_{i=0}^{n} \delta_i^{s_i} = \lim_{n\to\infty} \prod_{i=0}^{n} \text{Prob}(\phi(X)_i = s_i),$$

where the second equality is from (3.7). This justifies the claim of independence. □

For any two stochastic symbol sequences, the convex combination defined element-wise gives another stochastic symbol sequence. Similarly, convex combinations of probability distributions are also probability distributions. These are related as follows.

LEMMA 3.5. *For $\omega_1, \omega_2 \in \bar{\Gamma}$ and $\alpha \in [0, 1]$, we have*

$$\psi^L(\alpha\omega_1 + (1-\alpha)\omega_2) = \alpha\psi^L(\omega_1) + (1-\alpha)\psi^L(\omega_2). \tag{3.15}$$

Proof. This follows immediately from the linearity of (3.6). □

To use the stochastic symbol sequences δ^L and δ^R associated to a probability distributions ω for studying cutoff, we need to express the total variation distance directly in terms of the stochastic symbol sequences. Given $\omega, \hat{\omega} \in \hat{\Gamma}$ with stochastic symbol sequences $\{\delta^L, \delta^R\}$ and $\{\hat{\delta}^L, \hat{\delta}^R\}$, respectively, we have

$$|\omega - \hat{\omega}|_{TV} = \frac{1}{2} \lim_{n\to\infty} \sum_{s\in\Sigma} \left| \prod_{i=0}^{n} \delta_i^{s_i} - \prod_{i=0}^{n} \hat{\delta}_i^{s_i} \right|. \tag{3.16}$$

This expression is difficult to evaluate in general, so let us consider a simpler case, when δ^L and $\hat{\delta}^L$ only have p different digits, i.e., $\delta_i^L = \hat{\delta}_i^L$ when $i \notin \theta$, and $|\theta| = p$. Let Σ_p be the space of all symbol sequence of length p with two symbols. Then we have

$$|\omega - \hat{\omega}|_{TV} = \frac{1}{2} \sum_{s\in\Sigma_p} \left| \prod_{i\in\theta} \delta_i^{s_i} - \prod_{i\in\theta} \hat{\delta}_i^{s_i} \right|. \tag{3.17}$$

Importantly, we note that the total variation distance is invariant under any simultaneous reordering of the symbols in the sequences, so is only a function of the elements $s_i, i \in \theta$. Furthermore, (3.17) can serve as a lower bound when the symbols for $i \notin \theta$ are unknown.

3.3. General Results. We now turn from general stochastic symbol sequences to cases where one of our sequences corresponds to an invariant distribution of the map S. In particular, we will consider

$$\bar{\delta} = \left\{ .\frac{1}{2}\frac{1}{2}\frac{1}{2}\cdots \right\}. \tag{3.18}$$

This is invariant under the shift operator action, so it corresponds to an invariant distribution $\bar{\omega} = \psi^{-1}(\bar{\delta})$. In these and future expressions, we will write δ for δ^L and $1 - \delta$ for δ^R. Similarly, ψ is the map ψ^L.

The following convexity lemma is the basis of all later results.

LEMMA 3.6. *For $\delta, \hat{\delta} \in \Delta$ and $\alpha \in [0, 1]$, we have*

$$|\psi^{-1}(\alpha\delta + (1-\alpha)\hat{\delta}) - \bar{\omega}|_{TV} \le \alpha|\psi^{-1}(\delta) - \bar{\omega}|_{TV} + (1-\alpha)|\psi^{-1}(\hat{\delta}) - \bar{\omega}|_{TV}. \tag{3.19}$$

Proof. Since $\omega \mapsto |\omega - \bar{\omega}|_{TV}$ is convex, by using (3.15) we see that $\delta \mapsto |\psi^{-1}(\delta) - \bar{\omega}|_{TV}$ is also convex, giving the desired result. □

The above lemma means that even if we cannot directly calculation the total variation distance between stochastic symbol sequences, we can use the convexity to deduce useful bounds.

LEMMA 3.7. *Suppose δ and $\hat{\delta}$ are two stochastic symbol sequences corresponding to the probability distributions ω and $\hat{\omega}$, respectively. Suppose there exists a bijection $\gamma : i \mapsto j$ such that $\delta_i = \hat{\delta}_j$ for all $i \in \mathbb{Z} \setminus \{k\}$. If $|\delta_k - \frac{1}{2}| \ge |\hat{\delta}_{\gamma(k)} - \frac{1}{2}|$, then*

$$|\omega - \bar{\omega}|_{TV} \ge |\hat{\omega} - \bar{\omega}|_{TV}. \tag{3.20}$$

Proof. Since $|\psi^{-1}(\delta) - \bar{\omega}|_{TV}$ is independent of the order of the sequence, we can assume $\delta = \hat{\delta}$ for all $i \in \mathbb{Z} \setminus \{k\}$. Let $\tilde{\delta} = \delta$ for all $i \in \mathbb{Z} \setminus \{k\}$, and $\tilde{\delta}_k = 1 - \delta_k$. Then $|\psi^{-1}(\delta) - \bar{\omega}|_{TV} = |\psi^{-1}(\tilde{\delta}) - \bar{\omega}|_{TV}$. When $|\delta_k - \frac{1}{2}| \ge |\hat{\delta}_k - \frac{1}{2}|$, it is always possible to

choose $\alpha \in [0,1]$ such that $\hat{\delta}_i = \alpha\delta_i + (1-\alpha)\tilde{\delta}_i$. Applying (3.19) to δ and $\tilde{\delta}$ with the α above, we have

$$\begin{aligned}|\psi^{-1}(\hat{\delta}) - \bar{\omega}|_{TV} &\leq \alpha|\psi^{-1}(\delta) - \bar{\omega}|_{TV} + (1-\alpha)|\psi^{-1}(\tilde{\delta}) - \bar{\omega}|_{TV} \\ &= |\psi^{-1}(\delta) - \bar{\omega}|_{TV}.\end{aligned}$$

□

LEMMA 3.8. *Suppose δ and $\hat{\delta}$ are two stochastic symbol sequences corresponding to the probability distributions ω and $\hat{\omega}$, respectively. Suppose there exists a bijection $\gamma : i \mapsto j$ such that $|\delta_i - \frac{1}{2}| \geq |\hat{\delta}_j - \frac{1}{2}|$ for all $i \in \mathbb{Z}$. Then*

$$|\omega - \bar{\omega}|_{TV} \geq |\hat{\omega} - \bar{\omega}|_{TV}. \tag{3.21}$$

Proof. Apply Lemma 3.7 repeatedly. □

The above lemma allows us to compare $|\omega - \bar{\omega}|_{TV}$ and $|\hat{\omega} - \bar{\omega}|_{TV}$ if some relation of their stochastic symbol sequences is known. Now for a given ω we want to bound $|\omega - \bar{\omega}|_{TV}$ by choosing an $\hat{\omega}$ such that $|\hat{\omega} - \bar{\omega}|_{TV}$ is easy to calculate. The choices we make are the sequences with the following form:

$$\begin{aligned}\delta^m &= \{.\underbrace{m\,m\cdots m}_{p \text{ times}}\frac{1}{2}\frac{1}{2}\cdots\}, \\ \delta^M &= \{.\underbrace{M\,M\cdots M}_{p \text{ times}}\frac{1}{2}\frac{1}{2}\cdots\}.\end{aligned} \tag{3.22}$$

As we will see later, $|\omega - \bar{\omega}|_{TV}$ can be evaluated easily for such sequences.

3.3.1. Lower Bound. We now construct an explicit lower bound for the total variation, using the framework above.

THEOREM 3.9. *Suppose δ and $\bar{\delta}$ are two stochastic symbol sequences corresponding to the probability distribution ω and the invariant distribution $\bar{\omega}$, respectively, with $\bar{\delta}$ being (3.18). Suppose there is a set $\theta \subset \mathbb{N}_0 = 0, 1, 2, \ldots$, $|\theta| = p$, such that for all $i \in \theta$, $|\delta_i - \frac{1}{2}| > |m - \frac{1}{2}|$. Then*

$$|\omega - \bar{\omega}|_{TV} \geq \mathbf{I}_{\frac{1}{2}}(p - q^*, q^* + 1) - \mathbf{I}_{1-m}(p - q^*, q^* + 1), \tag{3.23}$$

where

$$q^* = \left\lfloor p\frac{\log 2 + \log(\frac{1}{2} - \epsilon)}{\log(\frac{1}{2} - \epsilon) - \log(\frac{1}{2} + \epsilon)} \right\rfloor, \tag{3.24}$$

$\epsilon = |m - \frac{1}{2}|$, and $\mathbf{I}$ is the regularized incomplete beta function.

Proof. Since the total variation is independent of the order of the sequence, we can assume that $\theta = \{0, 1, \ldots, p-1\}$. Also, without loss of generality, we assume for $i \in \theta$, $\delta_i \geq \epsilon + \frac{1}{2} = m$. Using equation (3.17) and Lemma 3.8, we have

$$\begin{aligned}|\omega - \bar{\omega}|_{TV} &\geq \frac{1}{2}\sum_{s\in\Sigma_p}\left|\prod_{i\in\theta}\delta_i^{s_i} - \prod_{i\in\theta}\bar{\delta}_i^{s_i}\right| \\ &\geq \frac{1}{2}\sum_{q=0}^{p}\binom{p}{q}\left|m^q(1-m)^{p-q} - \frac{1}{2^p}\right| \\ &= \frac{1}{2}\sum_{q=0}^{p}\left|\binom{p}{q}m^q(1-m)^{p-q} - \binom{p}{q}\frac{1}{2^p}\right|.\end{aligned} \tag{3.25}$$

So the total variation distance can be lower bounded by the difference between two binomial distributions. We can find their difference by subtracting their cumulative distribution functions at the point where they cross over each other. To do this, we need to find the point where the first distribution begins to exceed the second distribution. That is, to find the largest q^* such that

$$\left(\frac{1}{2}+\epsilon\right)^{q^*}\left(\frac{1}{2}-\epsilon\right)^{p-1-q^*} - 2^{-p} \leq 0. \tag{3.26}$$

Solving for q^* gives (3.24). Now the cumulative distribution function of a binomial distribution can be expressed in terms of the regularized incomplete beta function $\mathbf{I}$ as follows. For the distribution binomial(r,p), its cumulative distribution function is

$$F(q;p,r) = \mathbf{I}_{1-r}(p-q, q+1). \tag{3.27}$$

Substituting q^* into q, and $1/2$ and m into r, we obtain (3.23). □

An interesting corollary is that as p goes to ∞, $|\omega-\bar{\omega}|_{TV}$ goes to 1 for all $\epsilon > 0$. So each of the constant stochastic symbol sequences correspond to an invariant measure of the map.

3.3.2. Upper Bound. We have seen that for a constant sequence with components not equal to $1/2$, the total variation distance to the invariant distribution $\bar{\omega}$ is 1 from the previous theorem. Hence this result does not give us any information about the upper bound. We would like to bound the distance by the sum of the sequence, and give the following theorem.

THEOREM 3.10. *Suppose δ and $\bar{\delta}$ are two stochastic symbol sequences corresponding to the probability distribution ω and the invariant distribution $\bar{\omega}$, respectively, with $\bar{\delta}$ being (3.18). Suppose*

$$\sum_{i=0}^{\infty}\left|\delta_i - \frac{1}{2}\right| = \frac{M}{2}. \tag{3.28}$$

Then

$$|\omega-\bar{\omega}|_{TV} \leq \begin{cases} \rho - 2^{-\lceil M\rceil} & \text{if } \rho > 1-2^{-\lceil M\rceil} \\ 1-2^{-\lfloor M\rfloor} & \text{otherwise,} \end{cases} \tag{3.29}$$

where $\rho = \frac{1}{2} + \frac{M-\lfloor M\rfloor}{2}$.

Proof. With the convexity inequality (3.19), one can show that for a set of stochastic symbol sequences $(\delta^j)_{j=1}^{\infty}$ and coefficients $(\alpha_j)_{j=1}^{\infty}$ with $\alpha_j \in [0,1]$ and $\sum_{j=1}^{\infty}\alpha_j = 1$, we have

$$\sum_j \alpha_j \left|\psi^{-1}(\delta^j) - \bar{\omega}\right|_{TV} \geq \left|\psi^{-1}\Big(\sum_j \alpha_j\delta^j\Big) - \bar{\omega}\right|_{TV}. \tag{3.30}$$

So for each M, if we can express δ as the convex combination of a set of δ^j, then the total variation distance $|\omega-\bar{\omega}|_{TV}$ can be bounded by the total variation of individual δ^j. Let us define the set $\mathbf{C} = \{\delta \in \Delta \mid \sum_i |\delta_i - \frac{1}{2}| = \frac{M}{2}\}$. It is easy to see that $\mathbf{C} = \operatorname{conv}\mathbf{D}$, the convex hull of $\mathbf{D}$, where $\mathbf{D}$ is defined as

$$\mathbf{D} = \left\{\delta \in \Delta \;\middle|\; \delta \text{ has } \lfloor M\rfloor \text{ 1's, one } M-\lfloor M\rfloor, \text{ and all other elements are } \frac{1}{2}\right\}, \tag{3.31}$$

and $|\psi^{-1}(\delta) - \bar{\omega}(x)|_{TV}$ for each $\delta \in \mathbf{D}$ can be calculated as

$$|\psi^{-1}(\delta) - \bar{\omega}(x)|_{TV} = \begin{cases} \rho - 2^{-\lceil M \rceil} & \text{if } \rho > 1 - 2^{-\lceil M \rceil} \\ 1 - 2^{-\lfloor M \rfloor} & \text{otherwise,} \end{cases} \tag{3.32}$$

where $\rho = \frac{1}{2} + \frac{M - \lfloor M \rfloor}{2}$. We obtain the upper bound by choosing the set of δ^j to be $\mathbf{D}$. □

Unfortunately, the above upper bound is in general not very tight. We state it just to inspire the next theorem, which gives a tighter bound.

THEOREM 3.11. *Suppose δ and $\bar{\delta}$ are two stochastic symbol sequences corresponding to the probability distribution ω and the invariant distribution $\bar{\omega}$, respectively, with $\bar{\delta}$ being (3.18). Suppose*

$$\sum_{i=0}^{\infty} \left| \delta_i - \frac{1}{2} \right| = \frac{M}{2}, \tag{3.33}$$

and $\delta_i \leq M$ for all i. Then

$$|\omega - \bar{\omega}|_{TV} \leq \mathbf{I}_{\frac{1}{2}}(p - q^*, q^* + 1) - \mathbf{I}_{1-M}(p - q^*, q^* + 1) \tag{3.34}$$

with

$$q^* = \left\lfloor p \frac{\log 2 + \log\left(\frac{1}{2} - \epsilon\right)}{\log\left(\frac{1}{2} - \epsilon\right) - \log\left(\frac{1}{2} + \epsilon\right)} \right\rfloor, \tag{3.35}$$

where $\epsilon = |M - \frac{1}{2}|$ and $p = \lceil \frac{M}{2\epsilon} \rceil$.

Proof. Similarly to the previous proof, let

$$\mathbf{D} = \left\{ \delta \in \Delta \;\middle|\; \delta \text{ has } (p-1) \textit{ of } M\text{'s, one } M - (p-1)M, \text{ and all other } \delta_i\text{'s are } \frac{1}{2} \right\}. \tag{3.36}$$

Then the set $\mathbf{C} = \{\delta \in \Delta \mid \sum_i |\delta_i - \frac{1}{2}| = \frac{M}{2}\}$ is the convex hull $\mathbf{C} = \operatorname{conv} \mathbf{D}$. All $\delta \in \mathbf{D}$ have the same value of $|\psi^{-1}(\delta) - \bar{\omega}|_{TV}$ and this is bounded by

$$|\psi^{-1}(\delta) - \bar{\omega}|_{TV} < |\psi^{-1}(\delta^M) - \bar{\omega}|_{TV}, \tag{3.37}$$

and $|\psi^{-1}(\delta^M) - \bar{\omega}|_{TV}$ can be calculated in a similar fashion to that in the lower bound proof by finding the q^* such that the first binomial distribution is less than the second one, i.e., find the smallest q^* such that

$$\left(\frac{1}{2} + \epsilon\right)^{q^*} \left(\frac{1}{2} - \epsilon\right)^{p - q^*} - 2^{-p} \geq 0. \tag{3.38}$$

Once q^* is solved as in (3.35), we have $|\psi^{-1}(\delta^M) - \bar{\omega}|_{TV} = \mathbf{I}_{\frac{1}{2}}(p - q^*, q^* + 1) - \mathbf{I}_{1-M}(p - q^*, q^* + 1)$. Combining this with (3.37), we obtain (3.34). □

The above lower and upper bound theorems can be applied to any given stochastic symbol sequence so long as one can find δ^m and δ^M to bound it. In the following theorem, we consider a scenario in which $p \to \infty$ and $m \to 1/2$ (or $M \to 1/2$), and we calculate where the lower and upper bounds converge to.

THEOREM 3.12. *Suppose $\bar{\omega}$ and ω_{lb} (resp. ω_{ub}) are two probability distributions with stochastic symbol sequences $\bar{\delta}$ and δ^m (resp. δ^M) as defined in (3.18) and (3.22), respectively. If $p \to \infty$ and $m \to \frac{1}{2}$ (resp. $M \to 1/2$), then we have*

$$\begin{aligned} |\omega_{lb} - \bar{\omega}|_{TV} &= \operatorname{erf}\left(\sqrt{\frac{p}{2}}\left(m - \frac{1}{2}\right)\right), \\ \textit{resp. } |\omega_{ub} - \bar{\omega}|_{TV} &= \operatorname{erf}\left(\sqrt{\frac{p}{2}}\left(M - \frac{1}{2}\right)\right). \end{aligned} \tag{3.39}$$

Proof. When $p \to \infty$, the binomial distribution approaches the normal distribution as $\text{Binom}(m,p) \to \mathcal{N}(pm, pm(1-m))$. The term $|\omega_{\text{lb}} - \bar{\omega}|_{TV}$ thus equals the total variation between two normal distributions, which can be evaluated by subtracting their cumulative distribution functions:

$$\begin{aligned} |\omega_{\text{lb}} - \bar{\omega}|_{TV} &= \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x^* - pm}{\sqrt{2pm(1-m)}}\right)\right) - \frac{1}{2}\left(1 + \operatorname{erf}\left(\frac{x^* - p\frac{1}{2}}{\sqrt{2p\frac{1}{2}(1-\frac{1}{2})}}\right)\right) \\ &= \operatorname{erf}\left(\frac{x^* - pm}{\sqrt{2pm(1-m)}}\right) - \operatorname{erf}\left(\frac{x^* - p\frac{1}{2}}{\sqrt{2p\frac{1}{2}(1-\frac{1}{2})}}\right), \end{aligned} \tag{3.40}$$

where x^* is the point at which the two distributions cross. When $m \to \frac{1}{2}$, the denominators in both $\operatorname{erf}(\cdot)$ are the same, and hence the variances of the two normal distributions are the same while the means differ by $p(m - \frac{1}{2})$, and we obtain (3.39). The same argument holds for the ω_{ub} equality. □

4. Main results. In the previous section we only dealt with the relation between fixed ω and $\bar{\omega}$, with no evolution. In this section use our results to bound $|\omega^k - \bar{\omega}|_{TV}$ when ω^k is evolved by the Perron-Frobenius operator $\omega^{k+1} = P\omega^k$ and so the corresponding stochastic symbol sequence δ^k evolves with the shift map σ.

We begin by considering two examples, where given stochastic symbol sequences, we can observe how the corresponding values of $|\omega^k - \bar{\omega}|_{TV}$ change.

EXAMPLE 3. *Suppose the initial probability distribution ω^0 has the stochastic symbol sequence*

$$\psi(\omega^0) = \Big\{ . \underbrace{m\, m \cdots m}_{p \text{ times}} \frac{1}{2}\frac{1}{2}\cdots \Big\}, \tag{4.1}$$

with $m > \frac{1}{2}$, and $\bar{\omega}$ corresponds to (3.18). Now ω^k evolves so that the corresponding symbol sequence δ^k is shifted, with $\delta^{k+1} = \sigma(\delta^k)$. It is then obvious that $|\omega^p - \bar{\omega}|_{TV} = 0$.

However, how sharp is the change from the initial distance to zero? This depends both on p and m. We plot the trajectories of $|\omega^0 - \bar{\omega}|_{TV}$ versus p for $m = \{0.6, 0.7, 0.8, 0.9\}$ in the right plot of Figure 4.1. Note that one can also regard these figures as showing $|\omega^k - \bar{\omega}|_{TV}$ versus k with fixed $p = 300$ for $\psi(\omega^0)$, because the shift operator removes one m in the sequence per iteration.

When $m = 1$, the trajectory is already shown in Figure 2.2 for the tent map. Now from Figure 4.1 we see that even when $m < 1$, we still observe the concave sharp change before the total variation goes to zero.

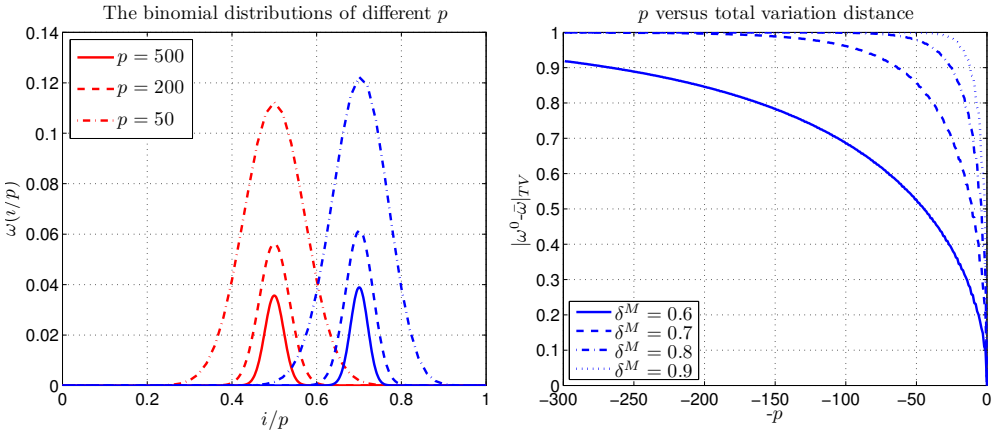

FIG. 4.1. *Left: comparison of the binomial distributions from (4.2). The value of* $|\omega^0 - \bar{\omega}|_{TV}$, *where* ω^0 *has the form (4.1), can be evaluated by finding the overlap area these two binomials. When* p *is large, the overlap area is smaller, so the total variation is smaller. Right: the variation of* $|\omega^0 - \bar{\omega}|_{TV}$ *with* p *for different values of* m. *Evolving the distribution (4.1) is equivalent to changing* p, *so this set of trajectories can be also interpreted as* $|\omega^k - \bar{\omega}|_{TV}$ *for* $k = 300 - p$.

An easier way to see why the concave trajectories occur is to plot the distributions ω *and* $\bar{\omega}$ *in a reduced space. We know that*

$$\begin{aligned} |\omega - \bar{\omega}|_{TV} &= \frac{1}{2} \sum_{s \in \Sigma_p} \left| \prod_{i=0}^{p} \delta_i^{s_i} - \prod_{i=0}^{p} \bar{\delta}_i^{s_i} \right| \\ &= \frac{1}{2} \sum_{q=0}^{p} \binom{p}{q} \left| m^q (1-m)^{p-q} - \frac{1}{2^p} \right| \\ &= \frac{1}{2} \sum_{q=0}^{p} \left| \binom{p}{q} m^q (1-m)^{p-q} - \binom{p}{q} \frac{1}{2^p} \right|. \end{aligned} \tag{4.2}$$

Figure 4.1 (left) plots the two terms $\binom{p}{q} m^q (1-m)^{p-q}$ *and* $\binom{p}{q} \frac{1}{2^p}$ *using* p/q *as the horizontal axis and with* $m = 0.7$. *We see that when* p *is small there is greater overlap, causing the drop in total variation.* ♢

From the above example, we can give the following theorem.

THEOREM 4.1. *For a 1D map* $S : \Lambda \to \Lambda$ *with full shift symbolic dynamics, define the family* $(\Omega, \bar{\omega}, (\omega_n^k)_{k=0}^{\infty})_{n=1}^{\infty}$ *by the initial condition* $\omega_n^0 \in \bar{\Gamma}$ *satisfying*

$$\psi(\omega_n^0) = \{.\underbrace{mm \cdots m}_{n} \frac{1}{2} \frac{1}{2} \cdots\} \tag{4.3}$$

for $0 \le m \le 1$, *and the evolution* $\omega^{k+1} = P\omega^k$ *by the Perron-Frobenius operator* P. *Then this sequence presents a total variation cutoff.*

Proof. Let $t_n = n$. When $k_n > (1+\epsilon)n$, then $|\omega_n^k - \bar{\omega}|_{TV} = 0$. When $k_n < (1-\epsilon)n$,

then $n - k_n > \epsilon n$, and so

$$\begin{aligned} |\omega_n^{k_n} - \bar{\omega}|_{TV} &= |\omega_n^{n-(n-k_n)} - \bar{\omega}|_{TV} \\ &\geq |\omega_n^{n-\lceil \epsilon n \rceil} - \bar{\omega}|_{TV} \\ &= |\omega_p^{p-\lceil \epsilon n \rceil} - \bar{\omega}|_{TV} \text{ for any } p \geq n \\ &\to 1 \text{ as } n \to \infty. \end{aligned}$$

□

Theorem 2.4 is a special case of Theorem 4.1 with $m = 1$.

EXAMPLE 4. *Suppose ω^0 has the stochastic symbol sequence*

$$\psi(\omega^0) = \{.\delta_0\delta_1\delta_2\cdots\}, \tag{4.4}$$

where $\delta_i = \frac{1}{2} + \epsilon r^i$, $\epsilon \in [0, 1/2]$ and $r \in [0, 1]$. We choose it to decay geometrically so we can use the upper bound theorem with fixed p. The area under δ and above $\frac{1}{2}$ can be calculated by the sum of a geometric series, as follows. Let $M^k = \delta_k$; then

$$\frac{M_k}{2} = \frac{\delta_k}{1-r}. \tag{4.5}$$

We thus choose $p = \lfloor \frac{1}{1-r} \rfloor$ for the upper bound Theorem 3.11. Now p is a constant, so the analysis is easier. We use the same p for the lower bound Theorem 3.9, so $m^k = r^p(\delta_k - \frac{1}{2})$. From the upper and lower bound theorems, the total variation distance to the stationary distribution at iteration k can be bounded by

$$|\psi^{-1}((\delta^m)^k) - \bar{\omega}|_{TV} < |\omega^k - \bar{\omega}|_{TV} < |\psi^{-1}((\delta^M)^k) - \bar{\omega}|_{TV}, \tag{4.6}$$

where

$$\begin{aligned} (\delta^m)^k &= \{.\underbrace{m^k m^k \cdots m^k}_{p \text{ times}} \frac{1}{2}\frac{1}{2}\cdots\}, \\ (\delta^M)^k &= \{.\underbrace{M^k M^k \cdots M^k}_{p \text{ times}} \frac{1}{2}\frac{1}{2}\cdots\}. \end{aligned} \tag{4.7}$$

This is valid for each k. The left plot of Figure 4.2 shows how the upper bound distribution moves toward the binomial distribution centered at $p/2$ as k increases, causing the drop in total variation as they overlap. The right plot of Figure 4.2 shows the upper and lower bounds versus iteration number k for $\delta_0 = 1$ and $r = 0.998$. We see the concave shape caused by the comparatively rapid overlapping of the distributions in the left plot.

We stress that finding the actual $|\omega^k - \bar{\omega}|_{TV}$ trajectory is infeasible because we would need to evaluate (3.16) for each ω^k. Even just to find an approximation using (3.17) is very expensive. This demonstrates the value of the upper and lower bound theorems. Even so, we do not get much insight on how the bounds evolve because these two theorems give bounds based on the incomplete beta function. Based on Theorem 3.12, we see that even when we do not have $p \to \infty$ and $m \to 1$, we still have the approximate bounds

$$|\omega^k - \bar{\omega}|_{TV} \lesssim \text{erf}\left(\sqrt{\frac{p}{2}}\left(M^k - \frac{1}{2}\right)\right) = \text{erf}\left(\sqrt{\frac{1}{2(1-r)}}\epsilon^0 r^k\right) \tag{4.8}$$

$$|\omega^k - \bar{\omega}|_{TV} \gtrsim \text{erf}\left(\sqrt{\frac{1}{2(1-r)}}\epsilon^0 r^{\frac{1}{1-r}} r^k\right) = \text{erf}\left(\sqrt{\frac{1}{2(1-r)}}\epsilon^0 e^{-1} r^k\right). \tag{4.9}$$

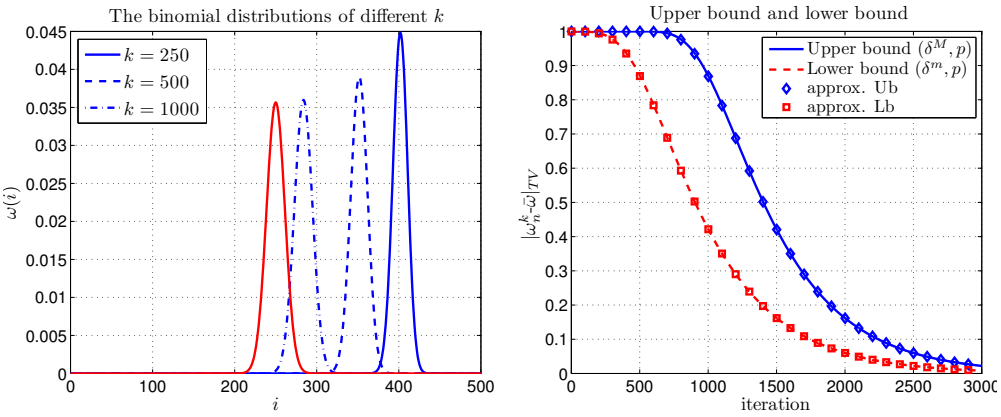

FIG. 4.2. *Left: the binomial distributions from the upper bound for Example 4, which overlap as iteration number k increases, causing a sharp fall in the total variation. The binomials are centered at $p\delta_k = \frac{1}{2} + \epsilon r^k$ and $p/2$. Right: the upper and lower bounds of $|\omega^k - \bar{\omega}|_{TV}$ by Theorems 3.11 and 3.9, as well as their approximations (4.8) and (4.9) derived from Theorem 3.12.*

These approximate upper and lower bounds are also plotted on the right of Figure 4.2, and we see that they are almost identical to the true bounds.

Both (4.9) and (4.8) show the normal shape, just like we saw in Example 1. Since $M^k - \frac{1}{2} = \epsilon^0 r^k$ decays exponentially (with factor r) to 0, the trajectory does not look like the error function itself. Instead, it has a sharper change before the cutoff occurs and decays more mildly after it. ◊

Examples 3 and 4 above show the impact of p and M, which relate to the sum of the stochastic symbol sequence and its largest entry. To summarize, the total variation between the current distribution and the invariant distribution can be characterized by its upper and lower bounds, which correspond to $|\psi^{-1}((\delta^m)^k) - \bar{\omega}|_{TV}$ and $|\psi^{-1}((\delta^M)^k) - \bar{\omega}|_{TV}$. These bounds can be evaluated (and also approximated) and then the shape of $|\omega - \bar{\omega}|_{TV}$ is determined.

4.1. Creating normal cutoffs. We have shown in Example 3 and Theorem 4.1 that evolution of probability distributions by P can produce a cutoff, but we did not establish the precise form of the limiting trajectory as it tended to a step function. In Example 4 we then showed how the limiting trajectories can be almost exactly the normal shape seen in Example 1 and many other finite Markov chain cutoff problems.

Here we go one step further and show how an initial distribution like (4.4) evolved by P can have very similar $|\omega^k - \bar{\omega}|_{TV}$ trajectory to what one sees in a cutoff phenomenon. Now we want to take a further step to show that we can generate a set of initial disitributions ω_n, which are a function of n, such that $\nu_n^k = |\omega_n^k - \bar{\omega}|_{TV}$ actually has precisely the shape (2.3) as n goes to infinity.

THEOREM 4.2. *For a 1D map $S : \Lambda \to \Lambda$ with full shift symbolic dynamics, define the family $(\Omega, \bar{\omega}, (\omega_n^k)_{k=0}^{\infty})_{n=1}^{\infty}$ by the evolution $\omega^{k+1} = P\omega^k$ with the Perron-Frobenius operator P from the initial conditions $\omega_n^0 \in \bar{\Gamma}$ satisfying*

$$\psi(\omega_n^0) = \{.\delta_0\delta_1\delta_2\cdots\}, \tag{4.10}$$

where $\delta_i = \min\{\frac{1}{2} + \epsilon_n r_n^i, 1\}$, *and*

$$\begin{aligned} r_n &= e^{-\frac{2}{n}} \\ \epsilon_n &= \sqrt{\frac{n(1-r_n)}{4}}. \end{aligned} \tag{4.11}$$

Then the family $(\Omega, \bar{\omega}, (\omega_n^k)_{k=0}^{\infty})_{n=1}^{\infty}$ *presents a total variation-cutoff.*

Moreover, letting $k = \frac{1}{4} n \log n + cn$ *for a fixed* $c \in \mathbb{R}$, *then as* $n \to \infty$, *we have*

$$\operatorname{erf}\left(\frac{e^{-2c-1}}{\sqrt{8}}\right) \leq |\omega_n^k - \bar{\omega}|_{TV} \leq \operatorname{erf}\left(\frac{e^{-2c}}{\sqrt{8}}\right). \tag{4.12}$$

Proof. From Example 3 we know that $|\omega_n^k - \bar{\omega}|_{TV}$ is bounded by (4.6). As $n \to \infty$, we have that ϵ_n goes to $\sqrt{\frac{1}{2}}$ and r_n goes to 1. Letting $k > \frac{-\log(2\epsilon_n)}{r_n}$, so that $\epsilon_n r_n^k < 1/2$, and using the results of (4.8) and (4.9), we have

$$\operatorname{erf}\left(\sqrt{\frac{1}{2(1-r_n)}}\epsilon_n e^{-1} r_n^k\right) \leq |\omega_n^k - \bar{\omega}|_{TV} \leq \operatorname{erf}\left(\sqrt{\frac{1}{2(1-r_n)}}\epsilon_n r_n^k\right). \tag{4.13}$$

In this case these bounds are not approximations, because the conditions of Theorem 3.12 are satisfied. Substituting the expressions for ϵ_n and r_n into above the inequality gives

$$\operatorname{erf}\left(\sqrt{\frac{n}{8}} e^{-\frac{2k}{n}-1}\right) \leq |\omega_n^k - \bar{\omega}|_{TV} \leq \operatorname{erf}\left(\sqrt{\frac{n}{8}} e^{-\frac{2k}{n}}\right). \tag{4.14}$$

Letting $k = \frac{1}{4} n \log n + cn$ for $c \in \mathbb{R}$, and substituting into the above expression, we obtain the desired result (4.12), and hence the family presents a cutoff. □

The key idea of the above theorem is that, for a pair (ϵ, r), we know that the stochastic symbol sequence with $\delta_i = \frac{1}{2} + \epsilon r^i$ has a normal-shape cutoff. Hence we equate the coefficients to find (ϵ_n, r_n) as functions of n that have the same $|\omega_n^k - \bar{\omega}|_{TV}$ as the random walk on a n-dimensional hypercube problem. In this case, the ϵ_n we find is larger than $1/2$ when $n > 1$, which is prohibited in the stochastic symbol sequence (it would require $\delta_0 > 1$). However, we set $\delta^i = 1$ for $i < \frac{-\log(2\epsilon_n)}{r_n}$ and then when $k > \frac{-\log(2\epsilon_n)}{r_n}$, the remaining sequence is exponential and Theorem 3.12 is applicable again.

Following the same procedure, we can reproduce any cutoff with a normal shape by a family $\{\Omega, \bar{\omega}, (\omega_n^k)_{k=0}^{\infty}\}_{n=1}^{\infty}$, where $\omega_n^{k+1} = P\omega_n^k$, and P is the Perron-Frobenius for any map $S : \Lambda \to \Lambda$ which has full shift symbolic dynamics.

4.2. What causes cutoffs?. One way to understand the source of cutoff behavior is to regard it as arising from systems which are composed of an increasingly large number of almost independent events. For example, in the hypercube random walk problem, whether each coordinate in the point's position vector is zero or one is almost independent of the other coordinate values, after a few iterations have elapsed. Indeed, assuming that these processes are actually independent (corresponding to the continuous-time random walk on a hypercube problem [7]) simplifies the problem dramatically while still giving the same cutoff behavior.

For chaotic map evolution, we have shown in Lemma 3.4 that the values of the symbols at different positions in the symbol sequence are completely independent,

with appropriately sampled initial conditions. So this process looks like it is composed of an infinite number of independent small processes, and we can choose the initial probability distributions to produce cutoff behavior. It is thus the full shift symbol dynamics of the chaotic map evolution that is key to cutoff in this setting.

5. Conclusion. We have proved that cutoff can be exhibited by the evolution of probability distributions by the Perron-Frobenius operator of 1D chaotic maps with symbolic dynamics. This connects the multistage mixing behavior of chaotic maps with the cutoff phenomenon for finite Markov chain evolution.

The key technique we used was a generalized symbolic dynamics, which we called stochastic symbol sequences. This connected the evolution of probability distributions by the Perron-Frobenius operator with the shift map acting on stochastic symbol sequences, enabling explicit bounds to be computed on the evolution of the total variation distance between an evolving distribution and the invariant distribution.

We showed that a simple class of initial conditions (4.3) results in a cutoff (with the tent map evolution of Example 2 as a special case), and we also showed that a particular choice of initial conditions (4.10) gives cutoff trajectories with normal shape, just as for many finite Markov chain systems.

The relationship between chaos and the cutoff phenomenon has also been observed in numerical studies and experiments. In [12], we provide numerical evidence to show that the variance of functions evolved by the Koopman operator of the diffusive standard map presents a cutoff as the diffusion goes to zero. In [13], a numerical simulation of a chaotic mixing channel also shows cutoffs in mixing trajectories, as do experiments for the same problem[18]. More numerical evidence can be found in the work of Thiffeault and Childress [20] and Tsang et al. [22], though they do not call the phenomenon "cutoff". These studies focus on the evolving variance of a scalar function advected by 2D chaotic maps with small diffusion—a different context from that explored here. However, just as cutoff phenomena are widespread for finite Markov Chains, we believe that cutoffs occur in many settings for chaotic maps.

REFERENCES

[1] D. Aldous and P. Diaconis. Shuffling cards and stopping times. *American Mathematical Monthly*, 93(5):333–348, 1986.

[2] D. Aldous and P. Diaconis. Strong uniform times and finite random walks. *Advances in Applied Mathematics*, 8:69–97, 1990.

[3] T. M. Antonsen, Jr., Z. Fan, E. Ott, and E. Garcia-Lopez. The role of chaotic orbits in the determination of power spectra of passive scalars. *Physical Fluids*, 8(11):3094–3104, 1996.

[4] D. Bayer and P. Diaconis. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, 2(2):294–313, 1992.

[5] G.-Y. Chen. *The Cutoff Phenomenon for Finite Markov Chains*. PhD thesis, Cornell University, 2006.

[6] P. Diaconis. The cutoff phenomena in finite Markov Chains. In *The Proceedings of the National Academy of Sciences*, volume 93, pages 1659–1664, 1996.

[7] P. Diaconis, R. L. Graham, and J. A. Morrison. Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures and Algorithms*, 1(1):51–72, 1990.

[8] P. Diaconis and L. Saloff-Coste. Separation cut-offs for death and birth chain. *Annals of Applied Probability*, 16(4):2098–2122, 2006.

[9] P. Diaconis and M. Shahshahani. Generating a random permutation with random transportations. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.

[10] D. R. Fereday, P. H. Haynes, A. Wonhas, and J. C. Vassilicos. Scalar variance decay in chaotic advection and Batchelor-regime turbulence. *Physical Review E*, 65(3):035301, 2002.

[11] J. Guckenheimer and P. Holmes. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer-Verlag, New York, 1983.

[12] T. Liang and M. West. Numerical evidence of cutoffs in chaotic mixing by the standard map. (preprint), 2011.
[13] T.-C. Liang and M. West. Optimized mixing in microfluidic channels. (preprint), 2011.
[14] J. M. Ottino and S. Wiggins. Designing optimal micromixers. *Science*, 305(5683):485–486, 2004.
[15] J. M. Ottino and S. Wiggins. Introduction: Mixing in microfluidics. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 362(1818):923–935, 2004.
[16] D. Rothstein, E. Henry, and J. P. Gollub. Persistent patterns in transient chaotic fluid mixing. *Letters to Nature*, 401:770–772, 1999.
[17] L. Saloff-Coste. Random walks on finite groups. In *Probability on Discrete Structures*, volume 110 of *Encyclopaedia of Mathematical Sciences*, pages 263–346. Springer, New York, 2004.
[18] A. D. Stroock, S. K. W. Dertinger, A. Ajdari, I. Mezić, H. A. Stone, and G. M. Whitesides. Chaotic mixer for microchannels. *Science*, 295(5555):647–651, 2002.
[19] J.-L. Thiffeault. Scalar decay in chaotic mixing. In *Transport in Geophysical Flows: Ten Years After, Proceedings of the Grand Combin Summer School*, 2004.
[20] J.-L. Thiffeault and S. Childress. Chaotic mixing in a torus map. *CHAOS*, 13(2):502–507, 2003.
[21] L. N. Trefethen and M. Embree. Further examples and applications. In *Spectra and Pseudospectra*, pages 508–525. Princeton University Press, Princeton, 2005.
[22] Y.-K. Tsang, T. M. Antonsen, Jr., and E. Ott. Exponential decay of chaotically advected passive scalars in the zero diffusivity limit. *Physical Review E*, 71:066301, 2005.
[23] G. A. Voth, G. Haller, and J. P. Gollub. Experimental measurements of stretching fields in fluid mixing. *Physical Review Letters*, 88(25), 2002.
[24] S. Wiggins. *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer, New York, 1990.
[25] S. Wiggins and J. M. Ottino. Foundations of chaotic mixing. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 362(1818):937–970, 2004.