Manuscript submitted to Journal of Computational Dynamics Website: http://AIMsciences.org
AIMS' Journals
Volume **X**, Number **0X**, XX **200X** pp. **X–XX**

NUMERICAL EVIDENCE OF CUTOFFS IN CHAOTIC MIXING BY THE STANDARD MAP

T. Liang

Department of Aeronautics and Astronautics,
Stanford University,
Stanford, CA 94305, USA

M. West

Department of Mechanical Science and Engineering,
University of Illinois at Urbana-Champaign,
Urbana, IL 61801, USA

Abstract. We numerically study the decay of the variance of a passive scalar function being mixed by the diffusive standard map. An efficient and parallelizable Markov chain model is used to approximate the Perron-Frobenius and Koopman operators of the chaotic map with diffusion. The limit of zero diffusivity is approximated by constructing a sequence of Markov chains with ever higher resolution (up to 6.4×10^9 states). In the limit of diffusivity going to zero we show that scalar variance exhibits a sharp decay and present numerical evidence to suggest that this is a cutoff in the sense of finite Markov chains.

1. **Introduction.** It is intuitively plausible that complex dynamics causes mixing on the state space. While this observation is crucial in many fields, such as the study of turbulence, here we focus on two well-studied models of this process: advection by a chaotic map and finite Markov chains. We numerically study the chaotic standard map by constructing Markov chain models of the dynamics and show how a connection can be made between the chaotic mixing viewpoint and the theory of cutoff in Markov chains.

1.1. **Mixing by chaotic maps.** The question of how chaotic advection, often with additional diffusion, mixes a passive scalar function has attracted much research effort (see, e.g., [30] for a discussion). Some of the main issues are: how to measure the thoroughness of the mixing, how the mixing process changes qualitatively and quantitatively when the diffusivity is close to zero, and how to enhance the overall mixing process by designing the map which produces chaotic advection. Unfortunately, we have only a partial understanding of most of these topics. In spite of the fact that the detailed mechanism of mixing is unclear, non-trivial mixing processes have been observed in experiments [33, 38] and can be simulated by large-scale computations [25, 37].

The connection between Markov chains and nonlinear dynamical systems has been long appreciated (see, e.g., [12, 13, 20, 21, 22] for one recent use). Markov

2000 *Mathematics Subject Classification.* Primary: 37M25, 65P20; Secondary: 37E30.
Key words and phrases. Chaotic mixing, standard map, cutoff phenomenon.

chain models are simple and parallelizable and they not only capture the multi-stage features of a chaotic mixing process, but also provide a ready mechanism for approaching the zero-diffusivity limit.

A widely observed phenomenon in the chaotic mixing process with small diffusion is the two or three-stage transition [36, 19, 3, 27]. The map does not mix the scalar function with a constant rate in general. When the variance of the scalar function is measured during the mixing process, in general we observe a relatively constant transient initially followed by a very fast decay (together comprising a super-exponential period), with it finally tending to a exponential decay. We are interested in when the transitions happen between the different phases, why they happen, and how to predict the decay rate in the exponential region. A good review and physical interpretation can be found in [35].

Thiffeault and Childress [36] study these properties for a modified Arnold's cat map. Analytical formulas are given to predict the transitions as well as the slopes. Because the linear part of this map has an eigenvalue approximately equal to 2.618, which stretches very fast, and the chaotic part is relatively small, the three phases are separated clearly. The same analytical procedure cannot be applied to, for example, the standard map, although the only difference between the standard map and the modified Arnold's cat map is in the linear part.

As for the exponential decay part, there is still debate about whether the decay rate goes to zero in the zero-diffusivity limit or whether it tends to a constant independent of the diffusivity [35, 37]. Theoretical analysis shows both of these possibilities can occur for different chaotic flows [23].

Difficulties typically arise in studying the above problems numerically, because the small diffusivity usually means that fine grids are required in the solution of the advection-diffusion equation or the simulation of the map. Some studies and numerical results conclude that a proportional relation exists between the stationary decay rate and the diffusivity [6, 32]. However, this is only true for certain diffusivity ranges, as shown by [37], in which the author uses a simple and parallelizable numerical strategy which can simulate on up to a $60\,000 \times 60\,000$ grid to show that the decay rate of a certain chaotic map tends to a constant.

1.2. **Cutoff in Markov chains.** Another approach to mixing can be found in the study of Markov chains, which considers questions such as how many riffle shuffles are required to sufficiently mix 52 cards? This question has been answered by Bayer and Diaconis in [4]. It is quite surprising that the cards are highly ordered in the first several shuffles (6 for 52 cards) and then randomized almost abruptly. This phenomenon has been termed *cutoff*, and was discovered by Aldous, Diaconis, and Shahshahani [2, 1, 17], and formalized by Aldous and Diaconis [14, 2].

The most interesting cases of cutoff 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 [34]. 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 [14, 16] 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 Markov chains that may exhibit cutoff phenomena (namely

discretizations of chaotic maps), and provide a numerical technique for gaining evidence that a cutoff exists.

1.3. **Contributions and organization of this work.** In this paper we focus on the mixing process of the standard map in the near-zero-diffusivity limit. The two main contributions are as follows.

First, we have produced the highest-resolution simulations of the standard map known to date (up to an $80\,000 \times 80\,000$ grid), allowing mixing rates to be computed with very low diffusivity. This is achieved using a simple and parallelizable Markov chain model, based on standard discretization methods, with a variety of diffusion models to test the influence of discretization scheme.

Second, we explicitly connect the multiphase nature of small-diffusivity chaotic mixing to the concept of cutoff in Markov chains, by providing numerical evidence that cutoff seems to occur in standard map mixing. Although there have been some historical differences in emphasis (e.g. total variation versus L^2 norms used to measure mixing), we argue that in many cases the chaotic mixing and Markov chain cutoff problems are essentially similar. Chaotic mixing evolution plots generally use log-scales, while Markov chain cutoffs are typically on linear scales, making published results look different even when the results are closely related. Another approach to this connection is undertaken in [24] using symbolic dynamics for 1D chaotic maps.

This paper is organized as the follows: in Section 2 we briefly review the evolution of functions and measures by maps, which we then discretize with Markov chain models in Section 3. We define cutoff in Section 4 and then provide numerical evidence that the standard map evolution presents a cutoff in Section 5, before giving conclusions in Section 6.

2. **Function Evolution by the Standard Map.** We work on the probability space $(X, \mathcal{A}, \mu)$. We take $S : X \to X$ to be a transformation (or map) that is non-singular and measurable. We choose μ to be the Borel measure. In the measure space $(X, \mathcal{A}, \mu)$ we define the following operators.

Definition 2.1 (Perron-Frobenius operator)**.** The *Perron-Frobenius operator* or *transfer operator* $P_S : L^1(X) \to L^1(X)$ associated with S satisfies

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

for every $\omega \in L^1(X)$ and $A \in \mathcal{A}$.

The Perron-Frobenius operator is linear. Because of our choice of measure space, the Perron-Frobenius operator can be interpreted as a map that evolves probability density functions. Also, suppose that $\bar{\omega}$ is an invariant measure of S, so that

$$\bar{\omega}(S^{-1}(A)) = \bar{\omega}(A) \text{ for all } A \in \mathcal{A}. \tag{2}$$

Then we have

$$P_S\bar{\omega} = \bar{\omega}. \tag{3}$$

Definition 2.2 (Koopman operator)**.** Let $f \in L^\infty(X)$. The operator $U_S : L^\infty(X) \to L^\infty(X)$ defined by

$$U_S f = f \circ S \tag{4}$$

is called the *Koopman operator* or *composition operator* associated to S.

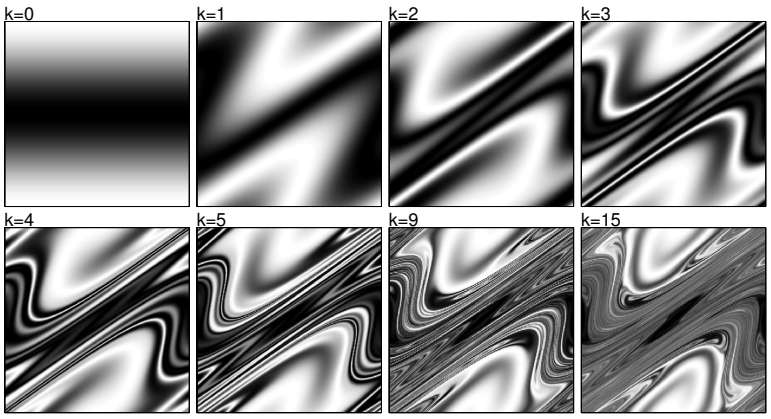

FIGURE 1. Evolution of the standard map with $\epsilon = 0.3$ and initial condition $f^0(x_1, x_2) = \cos(2\pi x_2)$. The simulation is with the numerical discretization $U^{\mathrm{d}}_{S^{-1},n}$ with $n = 500 \times 500$ grid cells. Iterations $k = 0, 1, \ldots, 5$ are shown, as well as iterations $k = 9$ and $k = 15$.

The Koopman operator is linear and is adjoint to Perron-Frobenius operator, which we write as $U_S = P_S^*$. If S is invertible then we have the following actions:

	forward in time	backward in time
probability density ω	P_S	$P_{S^{-1}}$
scalar function f	$U_{S^{-1}} = P^*_{S^{-1}}$	$U_S = P^*_S$

For each probability measure $\omega \in L^1(X)$, the *multiplication operator* $T_\omega : L^\infty(X) \to L^1(X)$ on scalar functions is $(T_\omega f)(x) = \omega(x) f(x)$. If $\bar{\omega}$ is invariant under an invertible S then we have

$$T_{\bar{\omega}} U_{S^{-1}} = P_S T_{\bar{\omega}}. \tag{5}$$

We particularly focus on the evolution of scalar functions forward in time by the map $U_{S^{-1}}$, where the map S of interest is as follows.

Definition 2.3 (Standard Map)**.** The *standard map* $S : [0,1]^2 \to [0,1]^2$ is defined by $S : (x_1, x_2) \mapsto (x_1', x_2')$, with

$$\begin{aligned} x_1' &= x_1 + x_2' \pmod 1 \\ x_2' &= x_2 + \epsilon \sin 2\pi x_1 \pmod 1. \end{aligned} \tag{6}$$

This map is invertible and chaotic [9, 8, 29, 26], where ϵ is a parameter that can be adjusted to change the behavior of the map. Here we focus on how a scalar function is evolved by the Koopman operator of the standard map with the additional diffusion. A simulation of the standard map for the initial condition $f^0(x_1, x_2) = \cos(2\pi x_2)$ is shown in Figure 1, computed using $n = 500 \times 500$ grid cells and the numerical approximation $U^{\mathrm{d}}_{S^{-1},n}$ discussed in Section 3. Because this map is volume preserving, it has a uniform invariant measure and so $U_{S^{-1}} = P_S$ by (5).

3. Markov Chain Models of Function and Measure Evolution.

3.1. **A model reduction view.** In this section we explain the numerical strategy to perform the simulation in detail. It is based on a model reduction view of Perron-Frobenius and Koopman operators, based on [10, 11, 28, 39, 18, 22, 21] (see [5] for an overview in the finite-dimensional setting).

Given the base space $X = [0,1]^2$ of interest, we take a regular grid with n cells $a_1, a_2, \ldots, a_n$. Each cell has area $1/n$ and dimensions $h \times h$, with $h = \sqrt{1/n}$. We are generally interested in n ranging from 500×500 up to $80\,000 \times 80\,000$. We can regard the grid as a discretization X_n of the space X, and we take

$$X_n = \{1, 2, \ldots, n\} \tag{7}$$

to be concrete. For any given n, let $g_n : X \to X_n$ be an *observation map* defined by $g_n(x) = i$ where $x \in a_i$. That is, g_n maps points to the grid cell number that they are in.

We are interested in functions and probability measures on the discretized space X_n and for clarity we write the spaces of these as $L^\infty(X_n)$ and $L^1(X_n)$, although these are both isomorphic to $\mathbb{R}^n$. Functions $f_n \in L^\infty(X_n)$ and probability measures $\omega_n \in L^1(X_n)$ can thus be thought of as vectors in $\mathbb{R}^n$ and we use the standard canonical bases in each case, so the i-th coordinate of f_n is $(f_n)_i = f_n(i)$.

We define the *discretization* $Q_n : L^\infty(X) \to L^\infty(X_n)$ and *reconstructor* $R_n : L^\infty(X_n) \to L^\infty(X)$ to be optimal in the sense

$$Q_n f = \operatorname*{argmin}_{f_n} \left\| f_n \circ g_n - f \right\|_{\bar\omega} \tag{8}$$

$$R_n f_n = \operatorname*{argmin}_{f} \left\| f - f_n \circ g_n \right\|_{\bar\omega} \tag{9}$$

where $\bar\omega$ is an invariant measure of S on X and $\|\cdot\|_{\bar\omega}$ is the $\bar\omega$-weighted 2-norm. In our case these can be explicitly calculated to be

$$(Q_n f)_i = \int_{a_i} f(x)\mu(dx) \tag{10}$$

$$(R_n f_n)(x) = (f_n)_{g_n(x)} \tag{11}$$

Note that $Q_n \circ R_n$ is the identity, but $R_n \circ Q_n$ is not.

For an evolution operator on functions, such as $U_S : L^\infty(X) \to L^\infty(X)$, we define the *optimal prediction* discretization to be a map $U_{S,n} : L^\infty(X_n) \to L^\infty(X_n)$ satisfying

$$U_{S,n} f_n = \operatorname*{argmin}_{f_n'} \left\| f_n' \circ g_n - U_S(f_n \circ g_n) \right\|_{\bar\omega}. \tag{12}$$

That is, the optimal discretization minimizes the one-step prediction error in the given norm. It can be shown [11, 5] that

$$U_{S,n} = Q_n U_S R_n. \tag{13}$$

The discretization $P_{S,n} : L^1(X_n) \to L^1(X_n)$ of $P_S : L^1(X) \to L^1(X)$ is defined so that the adjoint of $P_{S,n}$ is the optimal prediction discretization of the adjoint of P_S. That is, $(P_{S,n})^* = (P_S^*)_n$. But $P_S^* = U_S$, and the adjoint for discretized operators is simply transpose, so

$$P_{S,n} = U_{S,n}^T. \tag{14}$$

It can be shown that the optimal discretizations of $U_{S^{-1}}$ and P_S satisfy

$$T_{\bar\omega_n} U_{S^{-1},n} = P_{S,n} T_{\bar\omega_n}, \tag{15}$$

in analogy to (5), where $\bar{\omega}_n = R_n^* \bar{\omega}$ is an invariant distribution of $P_{S,n}$ (see [5]) and $T_{\bar{\omega}_n} : L^\infty(X_n) \to L^1(X_n)$ is the matrix $T_{\bar{\omega}_n} = \mathrm{diag}(\bar{\omega}_n)$.

By construction, $P_{S,n}$ and $P_{S^{-1},n}$ are Markov matrices, defining Markov chains on the finite state space $\mathbb{R}^n$. In our particular case we can compute the optimal discretizations to be

$$(U_{S,n})_{ij} = (P_{S,n})_{ji} = \frac{\bar{\omega}(S^{-1}(a_j) \cap a_i)}{\bar{\omega}(a_j)} \tag{16}$$

$$(U_{S^{-1},n})_{ij} = (P_{S^{-1},n})_{ji} = \frac{\bar{\omega}(S(a_j) \cap a_i)}{\bar{\omega}(a_j)}. \tag{17}$$

If the map S is volume-preserving (as is the case for the Standard Map), then, as we have remarked above, the uniform measure $\bar{\omega}$ is invariant under P_S and $P_{S^{-1}}$, and so we also have a uniform $\bar{\omega}_n$ that is invariant under $P_{S,n}$ and $P_{S^{-1},n}$. From (15) we then have

$$U_{S^{-1},n} = P_{S,n} \tag{18}$$

and we see that the evolution of functions and probability measures forward in time is identical.

For numerical efficiency, we do not use the exact expressions above for $U_{S,n}$ or $U_{S^{-1},n}$. Instead, we recognize that if $\bar{\omega}$ is smooth, then as $n \to \infty$ and the grid cells become very small, we have

$$(U_{S^{-1},n})_{ij} \approx \frac{\mu(S(a_j) \cap a_i)}{\mu(a_j)}. \tag{19}$$

For the standard map we will use uniform invariant distributions $\bar{\omega}$, so the approximation (19) is in fact exact for this case.

As we wish to scale to extremely high resolution, we approximate this further. Let C_j be the set of four points giving the corners of grid cell j. Then we take an approximation to $U_{S^{-1},n}$ given by

$$(U^{\mathrm{d}}_{S^{-1},n})_{ij} = \frac{1}{4}|S(C_j) \cap a_i|, \tag{20}$$

where $|\cdot|$ indicates cardinality. We are thus giving weight of $\frac{1}{4}$ to entry $(U^{\mathrm{d}}_{S^{-1},n})_{ij}$ for every corner of cell j that maps within cell i. It is easy to see that $U^{\mathrm{d}}_{S^{-1},n}$ is a stochastic matrix (non-negative entries and row-sums of 1), but it will not generally be doubly stochastic (also having column-sums of 1), unlike the exact $U_{S^{-1},n}$.

We present numerical evidence in Section 5.1 to show that the approximation of $U_{S^{-1},n}$ by $U^{\mathrm{d}}_{S^{-1},n}$ is unlikely to introduce any confounding influence into our results, as we add additional diffusion which is larger than the numerical diffusion due to any of these discretizations. We could also adopt other discretization methods, such as the lattice method [31], which approximates $U_{S^{-1}}$ by a permutation matrix, and then adds a smoothing step to make the Markov chain irreducible. Again, however, the addition of additional diffusion below makes the difference between the different discretization negligible, and so using a very computationally efficient method such as (20) is a sensible choice.

In practice we do not form and store $U^{\mathrm{d}}_{S^{-1},n}$, as the storage cost would be prohibitive. Instead we store only a state vector f_n and operate with $U^{\mathrm{d}}_{S^{-1},n}$ in a matrix-free manner. All the numerical studies in this paper were performed on a 72-node cluster with 1 GB RAM per CPU and a gigabit ethernet interconnect. The

largest simulation ($80\,000 \times 80\,000$ grid cells) required 51.2 GB of memory to store a single state vector.

3.2. **Numerical and Physical Diffusion.** Using the numerical strategy from the previous section, we can evolve a function or a probability distribution by the map with some small error. This error manifests as numerical diffusion or smoothing with diffusivity of the order $D \propto h^2 = 1/n$. The effect of numerical diffusion is similar to physical diffusion on large scales, as both cause smoothing, but their behaviors can be quite different on small scales, and small-scale phenomena are key to mixing by chaotic maps. To simulate physical diffusivity D correctly, we need to simulate the map with higher resolution to ensure that numerical diffusion effects are smaller than physical diffusion. We consider two different numerical models of physical diffusion.

First, we define the physical diffusion or spectral filter $F_D^{\mathrm{s}} : L^\infty(X) \to L^\infty(X)$ to be the $t = 1$ evolution of the diffusion equation with diffusivity D acting on the initial function. That is, $F_D^{\mathrm{s}}(f) = u(\cdot, 1)$ where $u(x,t)$ solves $u_t = Du_{xx}$ with $u(x,0) = f(x)$. Equivalently, for our case of $X = [0,1]^2$, we can write $f(x)$ and $F_D^{\mathrm{s}}(f)(x)$ in Fourier sine series of the form

$$f(x) = \sum_{\ell_1,\ell_2=1}^{\infty} A_{\ell_1,\ell_2} \sin(\ell_1 \pi x_1) \sin(\ell_2 \pi x_2) \tag{21}$$

$$F_D^{\mathrm{s}}(f)(x) = \sum_{\ell_1,\ell_2=1}^{\infty} w_{\ell_1,\ell_2} A_{\ell_1,\ell_2} \sin(\ell_1 \pi x_1) \sin(\ell_2 \pi x_2) \tag{22}$$

$$w_{\ell_1,\ell_2} = \exp\left(-4\pi^2(\ell_1^2 + \ell_2^2)D\right), \tag{23}$$

where w_{ℓ_1,ℓ_2} are the specified weights. We write $F_n^{\mathrm{s}} : \mathbb{R}^n \to \mathbb{R}^n$ for the discrete analogue of this, computed using an FFT and the same weighting coefficients with $D = 1/n$, with a final inverse FFT.

Second, we define the finite-difference smoothing filter F_n^{f} based on [37] by

$$(F_n^{\mathrm{f}} f_n)_{(p,q)} = \sum_{|r|,|s|\le 2} C_{|r|} C_{|s|} (f_n)_{(p+r,q+s)} \tag{24}$$

with $C_0 = 1/8, C_1 = 1/4$, and $C_2 = 3/16$ and (p,q) the two-dimensional index in the grid. This creates a large-scale diffusion with diffusivity $D \propto h^2$ several times larger than that due to numerical diffusion [37].

Using these different smoothing approximations, we have the following maps:

$$U_{S^{-1},n}^{\mathrm{d}} \text{ (no extra diffusion)} \tag{25}$$

$$U_{S^{-1},n}^{\mathrm{d+s}} = F_n^{\mathrm{s}} U_{S^{-1},n}^{\mathrm{d}} \text{ (spectral smoothing)} \tag{26}$$

$$U_{S^{-1},n}^{\mathrm{d+f}} = F_n^{\mathrm{f}} U_{S^{-1},n}^{\mathrm{d}} \text{ (finite-difference smoothing)} \tag{27}$$

For the parameter value $\epsilon = 0.3$ in the standard map and the initial condition $f^0(x_1, x_2) = \cos(2\pi x_2)$, Figure 2 shows the spectral content of the iterates f_n^k versus step number k. Here the number of grid cells is $n = 40\,000 \times 40\,000$ and the discretized initial condition is $f_n^0 = Q_n(f^0)$. The left plot in Figure 2 shows iterates under the map $U_{S^{-1},n}^{\mathrm{d}}$ with no additional diffusion (numerical diffusion only) and we observe that the chaotic map rapidly maps low frequency components to high frequency (it also does the converse). At this resolution n, the highest wavenumber that can be meaningfully resolved is around 2×10^4, at which point

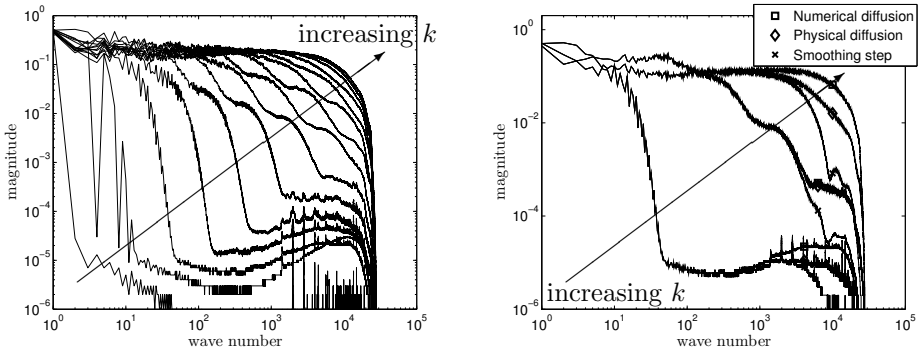

FIGURE 2. Spatial spectral content of iterates of the initial condition $f^0(x_1, x_2) = \cos(2\pi x_2)$ under the standard map with $\epsilon = 0.3$. The number of grid cells is $n = 40\,000 \times 40\,000$. The left plot shows the results using the numerical approximation $U^{\mathrm{d}}_{S^{-1},n}$ with no extra diffusion (numerical diffusion only) for iterates $k = 1$ through 20, with the $k = 1$ line being in the lower-left, and the lines moving towards the upper-right as k increases (as shown by the arrow). The right plot compares the three numerical approximations $U^{\mathrm{d}}_{S^{-1},n}$ (no extra diffusion), $U^{\mathrm{d+s}}_{S^{-1},n}$ (spectral smoothing), and $U^{\mathrm{d+f}}_{S^{-1},n}$ (finite-difference smoothing), for iterates $k = 3$, 7, and 20 (again the arrow shows the direction of increasing k).

the numerical diffusion becomes dominant. The equilibrium distribution is the result of this numerical diffusion balancing the transport to finer scales.

To compare the difference between the operators $U^{\mathrm{d}}_{S^{-1},n}$, $U^{\mathrm{d+f}}_{S^{-1},n}$, and $U^{\mathrm{d+s}}_{S^{-1},n}$ with different diffusion models, we repeated the above simulation for all three operators. The result is plotted in the right of Figure 2 for iterates $k = 3$, 7, and 20. The three trajectories agree well at low spatial frequencies, where the diffusion has little effect, but begin to differ around wavenumber 2×10^3. We observe that the finite difference smoothing on $U^{\mathrm{d+f}}_{S^{-1},n}$ is greater than the spectral smoothing on $U^{\mathrm{d+s}}_{S^{-1},n}$.

The results in Figure 2 show why it is hard to accurately simulate the evolution of functions or measures under chaotic maps such as the standard map. Extremely fine grids are necessary to capture the high-frequency components generated within just a few iterations.

In the limit $n \to \infty$ of infinite resolution, any of the maps $U^{\mathrm{d}}_{S^{-1},n}$, $U^{\mathrm{d+f}}_{S^{-1},n}$, or $U^{\mathrm{d+s}}_{S^{-1},n}$ are good approximations to the Standard Map. From numerical studies in this paper, however, we used only $U^{\mathrm{d}}_{S^{-1},n}$ and $U^{\mathrm{d+f}}_{S^{-1},n}$. This is because they are the cheapest to compute numerically and as $U^{\mathrm{d+s}}_{S^{-1},n}$ has intermediate diffusivity between $U^{\mathrm{d}}_{S^{-1},n}$ and $U^{\mathrm{d+f}}_{S^{-1},n}$, it is a reasonable supposition that because our results hold for $U^{\mathrm{d}}_{S^{-1},n}$ and $U^{\mathrm{d+f}}_{S^{-1},n}$, they will hold for $U^{\mathrm{d+s}}_{S^{-1},n}$ as well. The 2D FFT for the spectral smoothing in $U^{\mathrm{d+s}}_{S^{-1},n}$ is expensive when n is large, as on a parallel cluster the transpose needed to switch between row and column FFTs involves much communication.

4. **Variance Cutoffs in Function Evolution.** In some Markov chains, certain probability distributions converge to an equilibrium via a sharp transition, which becomes sharper for larger chains. This phenomenon is referred to as *cutoff* in the finite Markov chain literature [16, 7]. Here we modify the usual definition slightly to consider function evolution rather than distribution evolution, and to use variance as the measure of distance.

Definition 4.1 (Variance)**.** In a finite probability space X_n, the *variance* of a function $f_n \in L^\infty(X_n)$ with respect to a probability measure $\omega_n \in L^1(X_n)$ is

$$\mathrm{var}_{\omega_n}(f_n) = \sum_{i=1}^{n} (w_n)_i \left((f_n)_i - \sum_{j=1}^{n} (\omega_n)_j (f_n)_j \right)^2 . \tag{28}$$

We always take invariant probability measures $\bar{\omega}_n$ to compute the variance of functions, which will be uniform measures for the standard map. The functions we are interested in are sequences of the form f_n^k, where $f_n^0 = Q_n f^0$ is the discretization of a fixed smooth function $f^0 \in L^\infty(X)$, and $f_n^{k+1} = U_{S^{-1},n} f_n^k$ is the time-evolution under the discretized Koopman operator $U_{S^{-1},n}$ or one of its approximations. It is these function sequences that we are investigating for cutoff, in the following sense.

Definition 4.2 (Cutoff)**.** Take a family $(X_n, \omega_n, (f_n^k)_{k=0}^\infty)_{n=1}^\infty$ of finite probability spaces X_n, probability measures ω_n on X_n, and sequences $(f_n^k)_{k=0}^\infty$ of functions f_n^k on X_n. This family presents a *variance cutoff* if there exist reals M and m and a sequence $(t_n)_{n=1}^\infty$ of positive reals such that, for any $\delta \in (0,1)$, $\gamma > 1$, and sequence $(k_n)_{n=1}^\infty$,

$$\lim_{n\to\infty} \mathrm{var}_{\omega_n}(f_n^{k_n}) = M \text{ if } k_n < (1-\delta)t_n \text{ for all } n \tag{29a}$$

$$\lim_{n\to\infty} \mathrm{var}_{\omega_n}(f_n^{k_n}) = m \text{ if } k_n > (1+\delta)t_n \text{ and } k_n < \gamma t_n \text{ for all } n. \tag{29b}$$

The values t_n are known as *cutoff times*.

This definition of cutoff is essentially the same as that for Markov chains [16], and expresses the idea that the transition between M and m occurs around time t_n, and the time to transition relative to t_n becomes vanishingly small as $n \to \infty$. For evolution by the standard map, we generally have a non-zero final value $m > 0$, as the standard map has non-chaotic regions when $\epsilon > 0$ on which fast mixing does not occur, and on which variance decays arbitrarily slowly as $n \to \infty$. This is also the reason for requiring k_n/t_n to be bounded by a fixed γ, as for any given n it is true that $\mathrm{var}(f_n^k) \to 0$ as $k \to \infty$.

The study of cutoff in the finite Markov chain literature began with studying the total variation distance between an evolving probability measure and its limiting invariant measure (see [14, 16] for a discussion), but then also studied the separation distance, the L^2 distance, and other measures of distance. While variance (28) is capturing a similar idea to the L^2 distance from[16], the main interest in L^2 cutoff for Markov chains has been for initial conditions limiting to delta measures as $n \to \infty$ [7], whereas we take smooth initial conditions. We study variance evolution with smooth initial conditions because it is the most explored case in the chaotic mixing literature (e.g. [27, 36]). This is also the reason we consider the evolution of functions f_n rather than probability measures ω_n, although for volume-preserving maps such as the standard map we have $U_{S^{-1},n} = P_{S,n}$ (see (18)) and so this distinction is not important. The precise statement that we believe is true is as follows.

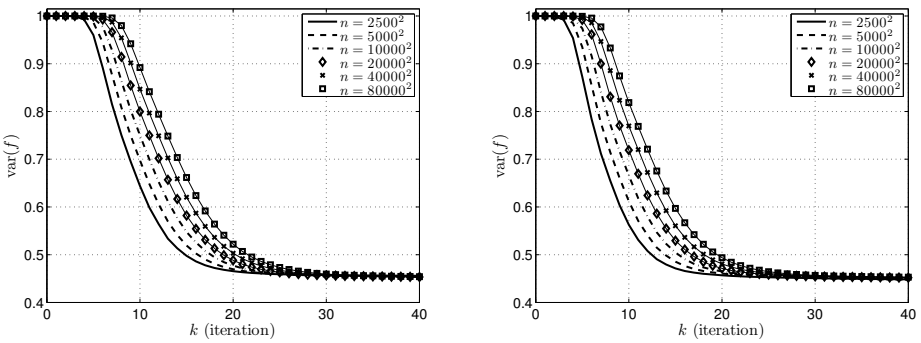

FIGURE 3. The variance of the function evolved by the standard map versus iteration number, starting from the normalization of the initial condition $f^0(x_1, x_2) = \cos(2\pi x_2)$ and with $\epsilon = 0.3$. The standard map is modeled by the Markov chains $U^{\mathrm{d}}_{S^{-1},n}$ (left plot, no extra diffusion) and $U^{\mathrm{d+f}}_{S^{-1},n}$ (right plot, finite difference smoothing added) and the number of grid cells n is varied from 2500×2500 to $80\,000 \times 80\,000$.

Conjecture 1. *Take the fixed smooth function $f^0 \in L^\infty(X)$ given by $f^0(x_1, x_2) = \cos(2\pi x_2)$ as the continuous initial condition. For each $n = 1, 2, \ldots$, let X_n be the regular grid with n grid cells on $X = [0,1]^2$ (see (7)) and ω_n be the uniform probability measure on X_n. Discretize the initial function with Q_n (see (8)) to give $f^0_n = c_n Q_n f^0$, normalized by $c_n \in \mathbb{R}$ so that $\mathrm{var}_{\omega_n}(f^0_n) = 1$. Construct a sequence $(f^k_n)_{k=0}^\infty$ by evolving $f^{k+1}_n = U_{S^{-1},n} f^k_n$, where $U_{S^{-1},n}$ is the discretized Koopman operator (see (12)) for the standard map S (see (6)) with $\epsilon = 0.3$. Then the family $(X_n, \omega_n, (f^k_n)_{k=0}^\infty)_{n=1}^\infty$ presents a variance cutoff.*

The effect of increasing the number of grid cells n is to reduce the numerical (or added) diffusivity. As we see in Section 5, the cutoff behavior seems relatively insensitive to the precise type of diffusion, and it is plausible that cutoffs occur using any of the maps $U_{S^{-1},n}$, $U^{\mathrm{d}}_{S^{-1},n}$, $U^{\mathrm{d+s}}_{S^{-1},n}$, or $U^{\mathrm{d+f}}_{S^{-1},n}$. In addition, as it is the diffusivity going to zero that causes the cutoff, it is likely that the same behavior occurs for the continuous Koopman operator $U_{S^{-1}}$ with smoothing, so we would expect a cutoff for evolution by $F^{\mathrm{s}}_D U_{S^{-1}}$ as $D \to 0$, for example.

5. Numerical results.

5.1. Cutoff phenomenon.

We now present numerical evidence for the central Conjecture 1 of this paper. Namely, that as the number of grid cells n of the Markov chain model of the standard map tends to infinity, the evolution f^k_n of a function tends to a trajectory with cutoff.

To show this, we varied the number of grid cells n from 2500×2500 to $80\,000 \times 80\,000$. At each resolution we took the discretized initial condition $f^0_n = c_n Q_n f^0$ with $f^0(x_1, x_2) = \cos(2\pi x_2)$ and the normalization constant c_n chosen so that $\mathrm{var}(f^0_n) = 1$. We then computed iterates f^k_n using both the Markov chain approximation $U^{\mathrm{d}}_{S^{-1},n}$ with no extra diffusion, and the Markov chain approximation $U^{\mathrm{d+f}}_{S^{-1},n}$ with finite-difference smoothing. The results are plotted in Figure 3.

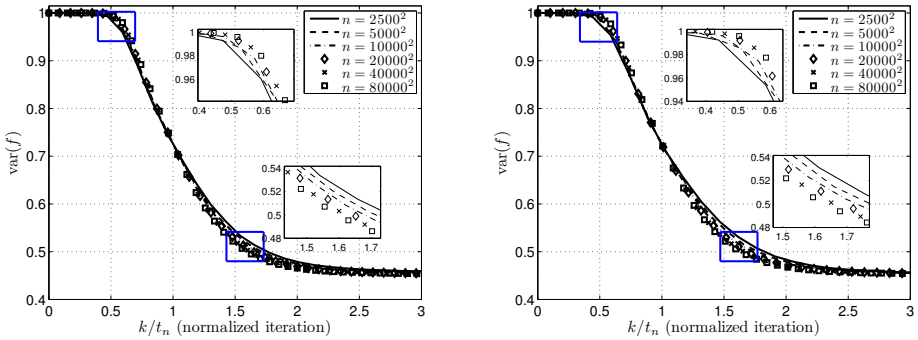

FIGURE 4. The trajectories from Figure 3 with iteration numbers scaled so the linearly interpolated trajectories at $k/t_n = 1$ are equal to $(M+m)/2$. The left plot is for $U^{\mathrm{d}}_{S^{-1},n}$ (no extra diffusion), while the right plot is for $U^{\mathrm{d+f}}_{S^{-1},n}$ (finite difference smoothing). The inset plots show zoomed regions near the corners of the trajectories.

By construction, the variance of f_n^k begins at $M = 1$. It holds near this value for some time, before falling to a limiting value of around $m = 0.4521$ (for $U^{\mathrm{d}}_{S^{-1},n}$) or $m = 0.4498$ (for, $U^{\mathrm{d+f}}_{S^{-1},n}$). The variance does not drop to zero because there are unmixed "islands" in the standard map. The slope of the rapid dropping also becomes slightly milder when n increases. Cutoff occurs in the limit if the "falling time" becomes small relative to the initial "holding time".

To make the study of this question precise in the sense of Definition 4.2, we let t_n equal the point where each linearly interpolated trajectory passes through $(M + m)/2$, and we rescaled each trajectory by scaling t_n to 1. We denote the rescaled trajectories by $\nu_n(r)$, where r represents the normalized iteration k/t_n and we linearly interpolate to make $\nu_n(r)$ continuous for any positive r. The results are plotted in Figure 4.

Although the rescaled trajectories are very similar, we can see that as n gets larger, the rescaled trajectories becomes sharper, suggesting that they may be tending to a cutoff. The left plot in Figure 5 shows the cutoff times t_n versus the grid size on a semi-logarithmic plot, indicating that if there is a cutoff, the cutoff time is inversely proportional to $\log(D)$, where $D \propto 1/n$ is the diffusivity added by the Markov chain approximation.

To quantify the approach of the rescaled trajectories to a sharp cutoff, we define the limiting step function

$$\nu_\infty(r) = \begin{cases} M & \text{if } r < 1, \\ m & \text{otherwise.} \end{cases} \tag{30}$$

We define a 1-norm distance between the rescaled trajectories $\nu_n(r)$ and the step function $\nu_\infty(r)$ by

$$\Delta_n^\ell = \int_0^\ell |\nu_\infty(r) - \nu_n(r)|\, dr, \tag{31}$$

where ℓ is the length of normalized iteration under consideration.

The right plot in Figure 5 shows Δ_n^ℓ with $\ell = 3$ for both $U^{\mathrm{d}}_{S^{-1},n}$ (no extra diffusion) and $U^{\mathrm{d+f}}_{S^{-1},n}$ (finite difference smoothing). In both cases we see a clear

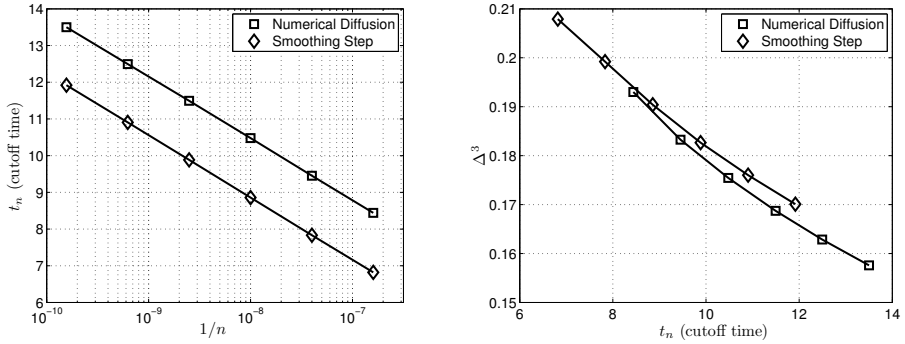

FIGURE 5. Left: the cutoff time t_n versus inverse grid cell number $1/n$, showing that the cutoff time is proportional to $\log(D)$ for diffusivity $D \propto 1/n$. Right: the 1-norm distance Δ_n^3 between the rescaled trajectories and the limiting cutoff step function, suggesting that the rescaled trajectories are exhibiting cutoff behavior. All simulations are for the standard map with parameter $\epsilon = 0.3$, initial condition $f^0(x_1, x_2) = \cos(2\pi x_2)$, and number of grid cells n from 2500×2500 to $80\,000 \times 80\,000$.

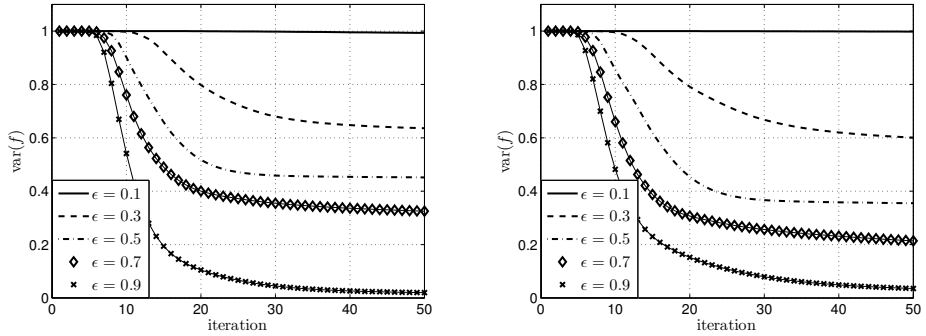

FIGURE 6. The evolution of function variance under the standard map iteration with initial condition $f^0(x_1, x_2) = \cos(2\pi x_1)$ (left) or $f^0(x_1, x_2) = \cos(2\pi x_2)$ (right), and the parameter ϵ varying between 0.1 and 0.9. All computations are performed with $n = 40\,000 \times 40\,000$ grid cells and the discretization $U^{\mathrm{d}}_{S^{-1},n}$ (no extra diffusion).

and consistent trend downwards, suggesting that Δ_n may well be converging to zero, and hence the rescaled trajectories $\nu_n(r)$ may be limiting to the step functions $\nu_\infty(r)$ as the number of grid cell n goes to infinity, and thus the sequences of Markov chains $U^{\mathrm{d}}_{S^{-1},n}$ and $U^{\mathrm{d+f}}_{S^{-1},n}$ may be presenting cutoffs.

5.2. **Effect of chaotic parameter ϵ and initial condition.** All of the above simulations were performed with $\epsilon = 0.3$ in the standard map equation (6) and the initial condition $f^0(x_1, x_2) = \cos(2\pi x_2)$. We consider briefly here how changing ϵ and f^0 might affect our conclusions.

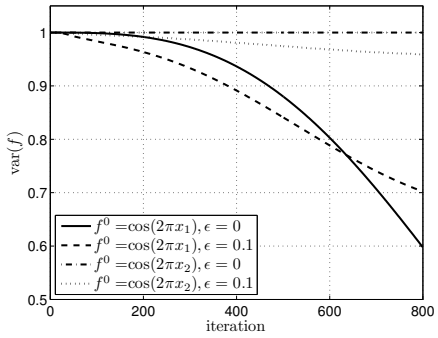

FIGURE 7. More iterations (to $k = 800$) of the evolution of function variance under the standard map iteration with initial condition $f^0(x_1, x_2) = \cos(2\pi x_1)$ or $f^0(x_1, x_2) = \cos(2\pi x_2)$, and parameter $\epsilon = 0$ or $\epsilon = 0.1$. All computations are performed with $n = 40\,000 \times 40\,000$ grid cells and the discretization $U^{\mathrm{d}}_{S^{-1},n}$ (no extra diffusion). Observe that for $f^0(x_1, x_2) = \cos(2\pi x_1)$, the $\epsilon = 0$ trajectory crosses the $\epsilon = 0.1$ trajectory near iteration $k = 627$.

We repeated the simulations with $n = 40\,000 \times 40\,000$ grid cells using the $U^{\mathrm{d}}_{S^{-1},n}$ map (no extra diffusion), taking ϵ ranging from 0.1 to 0.9 and the initial condition f^0 being either $\cos(2\pi x_2)$ or $\cos(2\pi x_1)$. The results are shown in Figure 6. For both initial functions and $\epsilon \geq 0.3$, the trajectories all have similar behavior to that observed earlier and it is likely that cutoff still occurs.

For $\epsilon = 0.1$, the trajectories show no decay in the first 50 iterations for either initial condition. To explore this further, in Figure 7 we plot the results of simulating with $\epsilon = 0$ and 0.1 for both initial functions out to 800 iterations.

We can see that for $f^0(x_1, x_2) = \cos(2\pi x_2)$ and $\epsilon = 0$, the trajectory still remains almost constant 1 for the first 800 iterations, while in for $f^0(x_1, x_2) = \cos(2\pi x_2)$ and $\epsilon = 0.1$ it has a small decrease but we have no evidence to say whether a cutoff may occur.

More interesting behavior is observed for the initial condition $f^0(x_1, x_2) = \cos(2\pi x_1)$. In this case, the trajectory for $\epsilon = 0$ crosses over that for $\epsilon = 0.1$ at around iteration $k = 627$. This phenomenon of the $\epsilon = 0$ map mixing faster than some other trajectories with higher ϵ has also been observed by [27]. This is because when $\epsilon = 0$, there is no unmixed region (no "islands") and so the decay in variance depends only on whether the map is good at mixing the initial functions.

5.3. **Zero-diffusivity limits and choice of norm.** In all the above simulations, we observed how the variance of a scalar function is evolved by the standard map with diffusion, essentially measuring variability by the L^2 norm. There are many other interesting norms that could be considered, however. In the left plot of Figure 8 we show trajectories for the following norms: L^1, L^2, $H^{-0.5}$, H^{-1}, and mix-norm [27]. All simulations were performed using the standard map with $\epsilon = 0.3$ and initial condition $f^0(x_1, x_2) = \cos(2\pi x_1)$ and the discretization $U^{\mathrm{d}}_{S^{-1},n}$ (no extra diffusion) with $n = 500 \times 500$ grid cells. All norms were normalized to equal 1 at iteration $k = 0$. Compared with the earlier simulations, this is a very coarse grid, and so we do not expect to see a very clear cutoff tendency. Nevertheless,

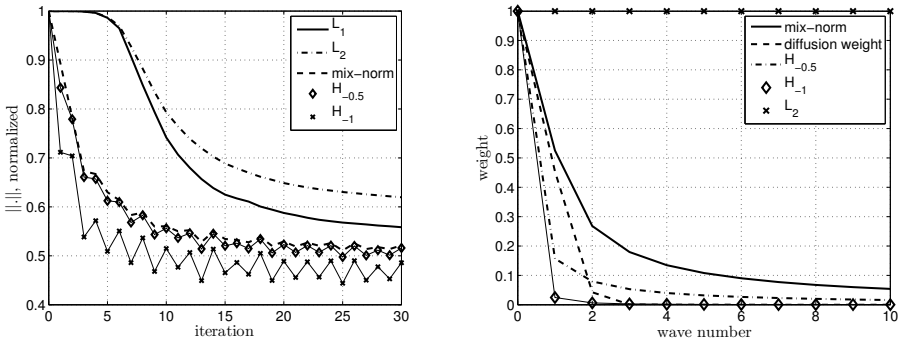

FIGURE 8. Left: Different norms of f_n^k evolved by the standard map with $\epsilon = 0.3$ and initial condition $f^0(x_1, x_2) = \cos(2\pi x_1)$. The discretization is $U^{\mathrm{d}}_{S^{-1},n}$ (no extra diffusion) with $n = 500 \times 500$ grid cells. The following norms of f_n^k are plotted: L^1, L^2, $H^{-0.5}$, H^{-1}, and mix-norm [27]. The norm trajectories form two groups: the L^1 and L^2 norms show the cutoff tendency seen earlier, while the other three norms do not. Right: The weights of L^2, $H^{-0.5}$, H^{-1}, and mix-norm versus wave number in the frequency domain, together with the weights of the Gaussian diffusion operator with diffusivity $D = 0.01$ for comparison. The L^2 norm has a constant weight 1 for all wave numbers. The other three norms have much smaller weights at large wave numbers. We see that computing the $H^{-0.5}$, H^{-1}, or mix-norm is roughly equivalent to computing the L^2 norm of a function smoothed with diffusivity $D = 0.01$.

these norms do form two distinct groups of behavior: for the L^1 and L^2 norms, we observe concave trajectories for the first few iterations, while all the other norms have trajectories that immediately decay to around 0.5 and then oscillate. From this we expect that we will see cutoff in either the L^1 or L^2 norm.

To understand this phenomenon, let us consider the zero-diffusivity case: the standard map is volume preserving, so without diffusion the L^1 and L^2 norms of the scalar function will remain constant under evolution by the Koopman operator of the standard map. The negative Sobolev norms and the mix-norm have diagonal representations in the frequency domain, and so can be plotted as weights for each wavenumber, as shown in the right plot in Figure 8. We also plot the weights of the Gaussian diffusion operator with diffusivity $D = 0.01$ for comparison. As a function is evolved by the Koopman operator it moves upscale in the frequency domain (as shown in Figure 2), and so its magnitude as measured by $H^{-0.5}$, H^{-1}, or the mix-norm will decrease.

Another way to think about this is to note that the $H^{-0.5}$, H^{-1}, and mix-norm of a function f are very roughly equal to the L^2 norm of a smoothing of f, where the smoothing is performed by the diffusion operator with $D \approx 0.01$. This diffusivity is very large and far from entering the range where we are seeing potential cutoff behavior. That is, the norms $H^{-0.5}$, H^{-1}, and mix-norm are explicitly insensitive to high-wavenumber (small-scale) information, but it is precisely the high-wavenumber

behavior of the chaotic map that is responsible for cutoff. It is thus not surprising that these three norms do not exhibit a cutoff.

We stress that there is no one norm that is "better" than another for measuring mixing. They simply give different information about the mixing process. The concave trajectories of L^1 and L^2 norms indicate the "irreversibility" of the chaotic system with small diffusivity: beyond the cutoff time substantial information about the initial system state has been lost. In contrast, the fast decay of the other norms shows the increasing complexity of the function evolved by the chaotic map in the first few iterations, but does not indicate when the initial state information has been lost.

6. **Conclusion.** We have presented the highest-resolution numerical simulations of the standard map known to date, and thereby provided numerical evidence of a cutoff in the sequence of Markov chains generated by approximating the Koopman operator of the standard map. The notion of a cutoff was taken from the study of finite Markov chains and applied to the study of chaotic mixing. We showed that cutoff not only characterizes the behavior of diffusive chaotic maps in the near-zero-diffusivity limit, but it also builds a bridge between the study of large finite Markov chains and chaotic maps.

Many Markov chains with cutoff are known, such as random walks on hypercubes or riffle-shuffling of cards, and it is not yet clear whether such chains can be related to chaos: are they discretizations or approximations of some chaotic map? We offer no good answer to this question so far. However, in [24] another attempt is made: we show that by choosing suitable initial distributions, a 1D chaotic map with symbolic dynamics can have the same limiting behavior as the cutoff of a random walk on n-dimensional hypercube problem [15], and thus it demonstrates another link between these chaotic mixing and cutoff.

The results in the paper have focused on variance as the measure of mixing, but as discussed in Section 5.3, it seems likely that similar results will occur for L^1 distances, and we do not expect cutoffs using negative Sobolov norms. This emphasizes the well-known fact that the details of mixing behavior are highly dependent on the definition of mixing.

We have also considered only finite dimensional Markov chains in the present study, due to the numerical nature of our techniques. From Section 5.1 we saw that the results are insensitive to the details of the discretization and the type of diffusion. From this, we expect that the true Koopman evolution $U_{S^{-1}}$ with diffusion F^{s}_D, giving the map $F^{\mathrm{s}}_D U_{S^{-1}}$, will also present a cutoff as $D \to 0$.

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] C. L. Beck, S. Lall, T. Liang, and M. West. Model reduction, optimal prediction, and the Mori-Zwanzig representation of Markov chains. In *Proceedings of the 48th IEEE Conference on Decision and Control*, pages 3282–3287, 2009.

[6] S. Cerbelli, A. Adrover, and M. Giona. Enhanced diffusion regimes in bounded chaotic flows. *Physics Letters A*, 312(5):355–362, 2003.
[7] G.-Y. Chen. *The Cutoff Phenomenon for Finite Markov Chains*. PhD thesis, Cornell University, 2006.
[8] B. Chirikov and D. Shepelyansky. Chirikov standard map. *Scholarpedia*, 3(3):3550, 2008.
[9] B. V. Chirikov. A universal instability of many-dimensional oscillator systems. *Physics Reports*, 52:263–379, 1979.
[10] A. J. Chorin and O. H. Hald. *Stochastic Tools in Mathematics and Science*. Springer Science and Business, 2006.
[11] A. J. Chorin, O. H. Hald, and R. Kupferman. Optimal prediction and the Mori-Zwanzig representation of irreversible processes. *Proceedings of the National Academy of Sciences*, 97:2968–2973, 2000.
[12] M. Dellnitz and O. Junge. On the approximation of complicated dynamical behavior. *SIAM Journal on Numerical Analysis*, 36(2):491–515, 1999.
[13] M. Dellnitz and O. Junge. Set oriented numerical methods for dynamical systems. In *Handbook of Dynamical Systems II: Towards Applications*, pages 221–264. World Scientific Publishing, 2002.
[14] P. Diaconis. The cutoff phenomena in finite Markov Chains. In *The Proceedings of the National Academy of Sciences*, volume 93, pages 1659–1664, 1996.
[15] 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.
[16] P. Diaconis and L. Saloff-Coste. Separation cut-offs for death and birth chain. *Annals of Applied Probability*, 16(4):2098–2122, 2006.
[17] P. Diaconis and M. Shahshahani. Generating a random permutation with random transportations. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.
[18] D. Evans and G. Morriss. *Statistical mechanics of nonequilibrium liquids*. Cambridge University Press, 2008.
[19] 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.
[20] G. Froyland. Approximating physical invariant measures of mixing dynamical systems in higher dimensions. *Nonlinear Analysis, Theory, Methods and Applications*, 32(7):831–860, 1998.
[21] G. Froyland. Markov modelling for random dynamical systems. In *Proceedings of Equadiff 99, Berlin, Germany*, 1999.
[22] G. Froyland. Extracting dynamical behaviour via Markov models. In *Nonlinear Dynamics and Statistics*, pages 283–324. Birkhäuser, Basel, 2001.
[23] P. H. Haynes and J. Vanneste. What controls the decay of passive scalars in smooth flows? *Physics of Fluids*, 17:097103, 2005.
[24] T.-C. Liang and M. West. Cutoff phenomenon and the mixing properties of chaotic maps. (In preparation), 2008.
[25] T.-C. Liang and M. West. Optimized mixing in microfluidic channels. (In preparation), 2008.
[26] A. J. Lichtenberg and M. A. Lieberman. *Regular and Chaotic Dynamics*. Springer-Verlag, second edition, 1992.
[27] G. Mathew, I. Mezić, and L. Petzold. A multiscale measure for mixing. *Physica D*, 211(1-2):23–46, 2005.
[28] H. Mori. Transport, collective motion, and Brownian motion. *Prog. Theor. Phys*, 33(3):423–455, 1965.
[29] E. Ott. *Chaos in Dynamical Systems*. Cambridge University Press, 2002. ISBN 0521010845.
[30] 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.
[31] R. T. Pierrehumbert. Lattice models of advection-diffusion. *Chaos*, 10(1):61–74, 2000.
[32] A. Pikovsky and O. Popovych. Persistent patterns in deterministic mixing flows. *Europhysics Letters*, 61(5):625–631, 2003.
[33] D. Rothstein, E. Henry, and J. P. Gollub. Persistent patterns in transient chaotic fluid mixing. *Letters to Nature*, 401:770–772, 1999.
[34] 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.

[35] J.-L. Thiffeault. Scalar decay in chaotic mixing. In *Transport in Geophysical Flows: Ten Years After, Proceedings of the Grand Combin Summer School*, 2004.

[36] J.-L. Thiffeault and S. Childress. Chaotic mixing in a torus map. *CHAOS*, 13(2):502–507, 2003.

[37] 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.

[38] G. A. Voth, G. Haller, and J. P. Gollub. Experimental measurements of stretching fields in fluid mixing. *Physical Review Letters*, 88(25), 2002.

[39] R. Zwanzig. Problems in nonlinear transport theory. In L. Garrido, editor, *Systems far from equilibrium*. Springer, Berlin, 1980.