SPEEDING UP LANGEVIN DYNAMICS BY MIXING

ALEXANDER CHRISTIE, YUANYUAN FENG, GAUTAM IYER, AND ALEXEI NOVIKOV

ABSTRACT. We study an overdamped Langevin equation on the d-dimensional torus with stationary distribution proportional to $p = e^{-U/\kappa}$. When U has multiple wells the mixing time of the associated process is exponentially large (of size $e^{O(1/\kappa)}$). We add a drift to the Langevin dynamics (without changing the stationary distribution) and obtain quantitative estimates on the mixing time. We show that an exponentially mixing drift can be rescaled to make the mixing time of the Langevin system arbitrarily small. For numerical purposes, it is useful to keep the size of the imposed drift small, and we show that the smallest allowable rescaling ensures that the mixing time is $O(d/\kappa^2)$, which is an order of magnitude smaller than $e^{O(1/\kappa)}$.

We provide one construction of an exponentially mixing drift, although with rate constants whose κ-dependence is unknown. Heuristics (from discrete time) suggest that κ-dependence of the mixing rate is such that the imposed drift is of size $O(d/\kappa^3)$. The large amplitude of the imposed drift increases the numerical complexity, and thus we expect this method will be most useful in the initial phase of Monte Carlo methods to rapidly explore the state space.

1. Introduction

Sampling from a given target distribution is a problem that arises in many modern applications, such as molecular dynamics [TLF77, QP04, LPV15, MS17, ARNB20], machine learning [AdFDJ03], field theory [BKK+85], Bayesian Statistics and computational physics [GCS+13]. A typical situation of interest is to draw samples from a probability distribution with density proportional to

$$(1.1) \qquad p = e^{-U/\kappa}.$$

Here U is a potential function that is usually regular and explicit, and $\kappa > 0$ is a small parameter.

Even though U is known explicitly, sampling from the above distribution is a numerically challenging problem with a long history [MRR+53, Has70, Nea96, RRT17, CCAY+18, BCP+21, CGLL22]. To briefly explain the difficulties involved, note first that in order to convert p into a probability density function, we need to normalize it by setting

$$(1.2) \qquad \rho_\infty = \frac{p}{Z} = \frac{e^{-U/\kappa}}{Z}, \quad \text{where} \quad Z = Z(\kappa) = \int_{\mathbb{R}^d} p \, dx.$$

Unfortunately, the constant Z is not easy to compute explicitly as numerical integration via quadrature is too expensive in high dimensions (see [Nov16] and references therein).

Moreover, even if the normalization constant Z is known, the majority of the mass of ρ_∞ is typically concentrated in a region with volume $O(\kappa^{d/2})$, where p is relatively larger. In order to effectively sample from ρ_∞, we need to identify this region. There is no obvious way to do this without evaluating p everywhere, a task that is computationally infeasible in high dimensions.

There are many numerical algorithms designed to address these issues. The first such algorithm was the celebrated Metropolis–Hastings algorithm [MRR$^+$53, Has70, LP17] which was designed to sample from the density ρ_∞ without knowledge of the normalization constant Z. Subsequently, numerous methods were developed to improve the convergence rate and address deficiencies in the Metropolis–Hastings algorithm. Some popular methods include Hamiltonian Monte Carlo (HMC), Langevin Monte Carlo, Metropolis Adjusted Langevin algorithm (MALA), and various stochastic gradient methods [AdFDJ03, Dia09, Bet17, LP17, GHKM21, GGZ22].

Of these, one that will be of particular interest to us is the Langevin Monte Carlo method. This method hinges on the fact that ρ_∞ is the stationary distribution of an over-damped Langevin equation, and so one can sample from ρ_∞ by performing Monte Carlo simulations. To elaborate, consider the over-damped Langevin equation

$$(1.3) \qquad dX_t = -\nabla U(X_t)\, dt + \sqrt{2\kappa}\, dW_t \,,$$

where W is a standard d dimensional Brownian motion. It is easy to see that the density ρ_∞ is stationary for the process X. If X is mixing, then the density of X_t for large enough t will be close to ρ_∞, and so performing Monte Carlo simulations on (1.3) will allow us to sample from ρ_∞.

It is well known that if U is strongly convex, then the process X is exponentially mixing [BE85]. More generally, if the stationary distribution satisfies a Poincaré inequality or log-Sobolev inequality (see for instance [Vil09, A.19] and [VW19]), then the process X is exponentially mixing, and one can sample from ρ_∞ by simulating (1.3). This leads to many sampling results such as [DT12, Dal17a, Dal17b, DM17, Che23], with guaranteed bounds on the convergence rate.

We will now restrict our attention to studying (1.3) on the d-dimensional torus $\mathbb{T}^d$ with non-convex U. In this case, a Poincaré inequality holds so the process X is always exponentially mixing. The convergence rate, however, could be extremely slow, and one has to simulate (1.3) for an extremely long time to generate good quality samples. In particular, if the potential U has multiple wells, the *mixing time* of X is $e^{O(1/\kappa)}$, which (for $\kappa \ll 1$) is too large to be practical. Recall the mixing time, denoted by $t_{\mathrm{mix}} = t_{\mathrm{mix}}(\kappa)$ (defined in (2.9) below), is often used to measure the rate at which the distribution of X converges to the stationary distribution [MT06, LP17].

The reason we expect to have $t_{\mathrm{mix}} = e^{O(1/\kappa)}$ is because local minima of the potential are metastable, and trap trajectories of X for exponentially long periods of time. Indeed, if X_0 starts at a local minimum of U, then the exit time from a small neighborhood is known to be $e^{O(1/\kappa)}$ (see [SM79, Sch80]). Thus if ρ_∞ is multimodal, the process X will not mix on time scales smaller than $e^{O(1/\kappa)}$.

The main contribution of this paper is to add a drift to (1.3) in a manner that both avoids the metastable traps at local minima and preserves the stationary distribution ρ_∞. As we will shortly see (Theorem 1.1, below), this can be used

to make the mixing time much smaller than $e^{O(1/\kappa)}$. Explicitly, the modification of (1.3) we consider is

$$(1.4) \qquad dX_t = Av_{At}(X_t)\,dt - \nabla U(X_t)\,dt + \sqrt{2\kappa}\,dW_t\,.$$

Here $A \gg 1$ is a large parameter, and v is a time dependent uniformly C^1 flow such that

$$(1.5) \qquad \kappa\nabla \cdot v - \nabla U \cdot v = 0\,.$$

Note equation (1.5) is equivalent to the condition that $\nabla \cdot (\rho_\infty v) = 0$, which implies that the stationary distribution of (1.4) is still ρ_∞.

It is known that adding any drift to Langevin dynamics (as in (1.4)) always makes it converges to equilibrium faster than directly using (1.3), without any drift (see for example [HHMS93, RBS15, DPZ17, HWG$^+$20]). Prior to our work, the increased convergence rate was obtained by taking $v(x) = J\nabla U(x)$ for an antisymmetric matrix J [LNP13, DLP16, GM16, LS18]. With this approach, the mixing time is still $e^{O(1/\kappa)}$, but with smaller constants than the mixing time of (1.3). Using a different approach, Damak et al. [DFY20] produce a sequence of time independent flows in $\mathbb{R}^2$ which make the mixing time arbitrarily small. Their construction relies on a strong oscillation of stream lines of the imposed drift, and only applies to the two dimensional case with a quadratic potential.

The first result in this paper is to provide a quantitative estimate of the mixing time of X when the deterministic flow of v is *exponentially mixing*. Here we mean exponentially mixing in the sense of dynamical systems, and we recall this notion in Section 2.1, below. We will estimate the mixing time in terms of the *dissipation time*, which roughly speaking, measures the rate at which X_t converges to the stationary distribution in the L^2 norm (the precise definition is in Section 2.2, below).

Theorem 1.1. *Suppose the vector field v satisfies* (1.5) *and generates an exponentially mixing flow. Denote the mixing rate by*

$$(1.6) \qquad h(t) = De^{-\gamma t}\,,$$

where D and γ are constants that may depend on κ. There exists constants C and $A_0 = A_0(\kappa) < \infty$ such that

$$(1.7) \qquad t_{\mathrm{dis}} \leqslant \frac{C\|\nabla v\|_{L^\infty}}{\gamma^2 A}\left(1 + \ln^2\frac{D\gamma^2 A}{\kappa\|\nabla v\|_{L^\infty}}\right),$$

$$(1.8) \qquad t_{\mathrm{mix}} \leqslant Cd\Big(1 + \frac{\|U\|_{\mathrm{osc}}}{\kappa} - \ln(\kappa t_{\mathrm{dis}})\Big)t_{\mathrm{dis}}\,,$$

for all sufficiently small $\kappa > 0$. Here $\|U\|_{\mathrm{osc}} = \max U - \min U$,

$$\|\nabla v\|_{L^\infty} = \Big\|\sum_{i,j}|\partial_i v_j|^2\Big\|_{L^\infty}^{1/2}\,.$$

Remark 1.2. We clarify that the constant C is independent of both the dimension and κ. Clearly both t_{dis} and t_{mix} vanish as $A \to \infty$. However, when A is large, solving (1.4) is computationally expensive, and thus we would like to choose A to be as small as possible. From the proof (see Remark 2.2, below) we will show that A_0 can be chosen according to

$$(1.9) \qquad A_0 = \frac{C'\kappa\|\nabla v\|_{L^\infty}}{\gamma^2}\ln^2\frac{C'D}{\kappa}\,,$$

for some constant C' which is independent of κ. Thus if we choose $A = A_0$ in Theorem 1.1, then the bounds (1.7) and (1.8) reduce to the polynomial bounds

$$(1.10) \qquad t_{\text{dis}} \leqslant \frac{C''}{\kappa} \quad \text{and} \quad t_{\text{mix}} \leqslant \frac{C'' d}{\kappa^2} \,,$$

for some constant C'' that is independent of κ and d. This is an order of magnitude smaller than the mixing time of (1.3) which is $e^{O(1/\kappa)}$ for multi-modal distributions.

We were unable to use the "usual techniques" (e.g. coupling, Cheeger bounds, etc. [LP17, MT06]) to obtain the mixing time bound (1.8). The proof of Theorem 1.1 instead uses a PDE based Fourier splitting method to obtain the dissipation time bound (1.7) (see Theorem 2.1, below), and then estimates the mixing time in terms of the dissipation time (Proposition 2.4, below). Postponing further discussion of these ideas to Section 2, we now explicitly construct velocity fields that are exponentially mixing so that Theorem 1.1 may be applied.

Notice first that a large family of velocity fields satisfying (1.5) can be easily constructed by taking skew gradients. Indeed, if ψ is a periodic stream function, and $i, j \in \{1, \dots, d\}$ with $i \neq j$, then any velocity field v defined by

$$(1.11) \qquad v \stackrel{\text{def}}{=} \frac{1}{p} \nabla_{i,j}^{\perp}(p\psi) = \nabla_{i,j}^{\perp}\psi - \frac{\psi \nabla_{i,j}^{\perp} U}{\kappa} \,,$$

satisfies the measure preserving condition (1.5). Here $\nabla_{i,j}^{\perp}$ is the skew gradient in the x_i-x_j plane, and is defined by

$$(1.12) \qquad \nabla_{i,j}^{\perp}\psi = -\partial_j \psi \, \boldsymbol{e}_i + \partial_i \psi \, \boldsymbol{e}_j \,,$$

where $\boldsymbol{e}_i, \boldsymbol{e}_j$ are the standard i^{th} and j^{th} basis vectors respectively. If ψ is a function of only one coordinate, say $\psi(x) = F(x_i)$, and p is identically constant, then the velocity field v above is simply a shear flow with with magnitude $F'(x_i)$ directed along the jth coordinate axis. If $\psi(x) = F(x_i)$ as above, but p is not identically constant, then we will call v a "modified shear". In this case we note that the velocity field lies in the x_i-x_j plane, but may not necessarily be directed along $\boldsymbol{e}_j$. Moreover, the magnitude now depends on all coordinates, and not just x_j.
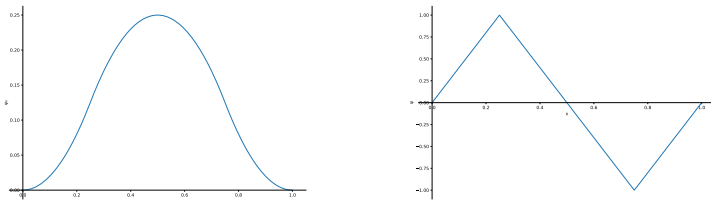

FIGURE 1. The function F (left) and its sawtooth shaped derivative F' (right).

We will now construct exponentially mixing velocity fields using randomly shifted alternating modified shears, in the spirit of Pierrehumbert [Pie94] who used a similar construction to study mixing in fluid dynamics. Explicitly, choose $\alpha_n \in [0, 1]$,

$\beta_n \in [0,1]$ and $i_n, j_n \in \{1, \dots, d\}$ to be uniformly distributed, i.i.d. random variables such that $i_n \neq j_n$. Given a periodic function $F \colon \mathbb{R} \to \mathbb{R}$ define

$$(1.13) \qquad v_t(x) \stackrel{\text{def}}{=} \frac{\beta_n}{p(x)} \nabla^{\perp}_{i_n, j_n} \big(p(x) F(x_{i_n} - \alpha_n) \big), \quad \text{when } t \in [n, n+1).$$

We will either choose $F(x) = \sin(2\pi x)$ (as in [Pie94]), or choose F so that the derivative is a sawtooth shaped function (see Figure 1, or the exact formula (6.1) in Section 6.1, below). We claim that that v is exponentially mixing with probability 1.

Theorem 1.3. *Suppose the potential U is C^2, and F is the function with sawtooth derivative shown in Figure 1. If $d \geqslant 3$, suppose further there exists a small ball $B(\hat{x}, \hat{\varepsilon}) \subseteq \mathbb{T}^d$ such that*

$$(1.14) \qquad U(x) = \sum_{i=1}^{d} U_i(x_i) \quad \text{for all } x \in B(\hat{x}, \hat{\varepsilon}).$$

Then there exists a constant $\gamma = \gamma(\kappa, d) < \infty$, and a finite random variable $D = D(\kappa, d)$ such that almost surely the velocity field (1.13) is exponentially mixing with rate (1.6).

If instead $F(x) = \sin(2\pi x)$, then the same conclusions hold provided we also assume the critical points of U are isolated.

Remark 1.4. While the cosine shears (corresponding to $F(x) = \sin(2\pi x)$) are more stable numerically, there are many common distributions (such as the Rosenbrock distribution [Ros61]) where the critical points of U are not isolated. In this case, we believe the velocity field v is still exponentially mixing for $F(x) = \sin(2\pi x)$, however, certain technical aspects of our proof break down. When F' is a sawtooth-shaped function no assumption on critical points of U is required.

Remark 1.5. Before proceeding further we briefly comment on the situation when the state space is $\mathbb{R}^d$, and not the compact torus. Part of Theorem 1.1 can be generalized to this setting, and we elaborate on this further in Remarks 2.3 and 2.6, below. In order to apply Theorem 1.1, however, we would need to construct flows on $\mathbb{R}^d$ that are exponentially mixing with respect to the density ρ_∞. This is not easy to do, and we are presently not aware of any such examples. Velocity fields of the form (1.11) are not suitable because either the trajectories will reach infinity in finite time or they will be too slow for the mixing effects to be useful. Moreover, the examples of mixing flows we construct (Theorem 1.3) rely heavily on the compactness of the state space, and we are presently not aware of any examples of exponentially mixing flows on $\mathbb{R}^d$ that preserve the stationary distribution ρ_∞.

Unfortunately, D and γ depend on κ and the dimension, and the proof of Theorem 1.3 does not provide any information on the asymptotic behavior of D and γ as $\kappa \to 0$ and $d \to \infty$. We can, however, study a discrete time version of (1.4), and produce exponentially mixing maps for which

$$(1.15) \qquad D = e^{O(1/\kappa)}, \quad \gamma = O(1).$$

(The precise construction is described in Section 5, below.) Suppose, momentarily, that for one of the velocity fields from Theorem 1.3 we still have (1.15). For such velocity fields, we note that $\|\nabla v\|_{L^\infty} = O(\sqrt{d}/\kappa)$. Thus choosing $A = A_0$, where

A_0 is given by(1.9), reduces to choosing

$$A = O\left(\frac{\sqrt{d}}{\kappa^2}\right),$$

in order to obtain the polynomial in $1/\kappa$ mixing time bounds stated in (1.10). Note that with this choice, the drift term in (1.4) is of size $O(d/\kappa^3)$.
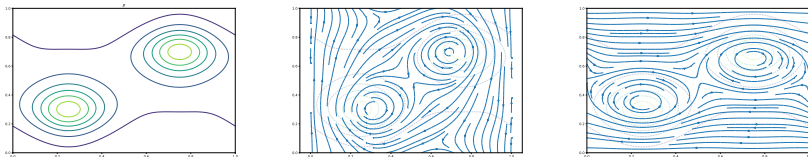

FIGURE 2. Level sets of p (left), and stream plots of v (center, right) corresponding to a modified vertical and horizontal shear respectively.

To illustrate our results numerically, we choose a double well potential U

$$U \overset{\text{def}}{=} (\sin^2(\pi(x_1 - .75)) + \sin^2(\pi(x_2 - .7)))(\sin^2(\pi(x_1 - .75)) + \sin^2(\pi(x_2 + .7))),$$

which has two minima at the points $(0.25, 0.3)$ and $(0.75, 0.7)$. Rather than choosing v according to (1.13), it is numerically more convenient to choose

$$(1.16) \quad \psi = M_t\left(\sin^2(2\pi\omega t)\sin(2\pi(x_1 - W_{1,t}) + \cos^2(2\pi\omega t)\sin(2\pi(x_2 - W_{2,t}))\right),$$

and then define v according to (1.11). Here M_t is a mean reverting Ornstein–Uhlenbeck process, $W_{i,t}$ are independent Brownian motions, and $\omega > 0$ is a parameter. Note when $\omega t \in \pi\mathbb{Z}$, the stream function ψ is only depends on x_2, and when $\omega t \in (\pi + \frac{1}{2})\mathbb{Z}$, the stream function ψ only depends on x_1. Thus this is a time continuous way of choosing the velocity fields defined in (1.11), with ω controlling the frequency at which the fields switch direction. Level lines of the function p (equation (1.1)), and a stream plot of v at times $t = 0$ and $t = (2\pi + 1)/(2\omega)$ are shown in Figure 2. Of particular interest is the fact that v is not 0 at the local minima of U, and this is what allows solutions to (1.4) to quickly escape the metastable traps at critical points.

We now solve equation (1.4) numerically with $\kappa = 1/70$ and choose the initial distribution to be the delta measure located at $(0.75, 0.7)$ (one of the local minima of U). In a short amount of time solutions to both equations fill out a neighborhood of the local minimum that they start at. However, since local minima of U are metastable traps for (1.3), very few realizations of solutions to (1.3) are able to leave this neighborhood. As a result, very few of these points are present in the other local minimum of U located at $(0.25, .3)$ (see Figure 3, left). In contrast, solutions to (1.4) are not trapped at critical points for as long, many realizations of solutions to (1.4) quickly find their way to the other local minimum (see Figure 3, center, right). At the final time in our simulations ($T = 3.14$), the distribution of solutions to (1.4) were close to the stationary distribution, but the distribution of solutions to (1.3) were not.

Despite the mixing time of (1.4) being an order of magnitude smaller than that of (1.3), there are a few issues that increase the complexity when solving (1.4)

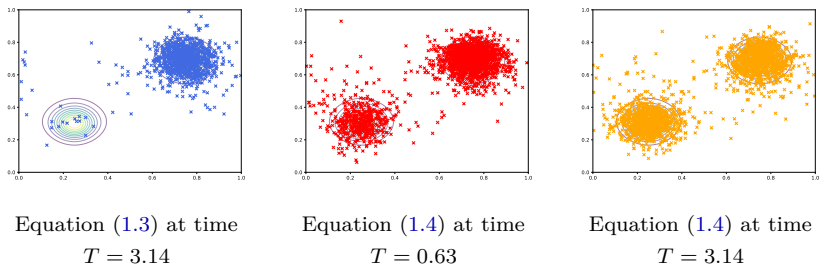| Equation (1.3) at time | Equation (1.4) at time | Equation (1.4) at time |
| $T = 3.14$ | $T = 0.63$ | $T = 3.14$ |

FIGURE 3. Distribution of 2000 realizations of solutions to (1.3) and (1.4) with $\kappa = 1/70$ and a double well potential. Solutions to (1.3) take a very long time to leave a neighborhood of a local minimum of U, where as solutions to (1.4) leave it quickly and approach the stationary distribution much faster.

numerically. Explicitly, solving equation (1.3) is relatively easier as the largest term on the right is of size $O(1)$. Solving equation (1.4), on the other hand, is relatively harder as the added drift has magnitude $O(\sqrt{d}/\kappa)$. The trade-off is that one only needs to solve (1.4) for time $O(d/\kappa^2)$, where as, in order to obtain comparable results with (1.3) one needs to solve it for time $e^{O(1/\kappa)}$. Thus we expect that algorithms using (1.4) will be useful in the initial exploratory phase of Monte Carlo methods, where accuracy is not as important. After rapidly exploring the state space using (1.4), it may be better to use other methods such as [RT96, MCF15, BCVD18, FBPR18, BFR19, LLN19, GHKM21, LSW22].

Finally, we conclude this section by stating a few questions that we are presently unable to address.

(1) The most important question we are unable to address is to rigorously describe the asymptotic behavior of D and γ as $\kappa \to 0$. The heuristics (1.15) (which we can prove for a time discrete example) leads to the polynomial mixing time bounds (1.10), which is an order of magnitude smaller than the mixing time of (1.3). We are presently unaware of techniques that provide any rigorous bounds on D and γ as $\kappa \to 0$.

(2) Since the numerical complexity increases with A, it is useful to find velocity fields v where A_0 can be chosen to be small. From (1.9) this entails finding velocity fields for which $\|\nabla v\|_{L^\infty}$ and D are small, and γ is large. Are there velocity fields v for which one expects $\gamma = O(1)$, and $D = O(1/\kappa)$, as opposed to the bounds in (1.15)?

(3) What happens when we work on $\mathbb{R}^d$ instead of $\mathbb{T}^d$? The current proof techniques will still allow us to prove (1.7). However, we are unable to prove (1.8), or Theorem 1.3 in $\mathbb{R}^d$ as explained in Remark 1.5.

(4) Even though we can prove that the velocity field in (1.13) is almost surely exponentially mixing, it is computationally intensive as it involves changing the velocity field discontinuously at many points in time. It also requires choosing A large, which further increases the computational cost. Numerically we found the continuous time modification (1.16), described above, yielded better results, presumably because of discretization artifacts. In the present work we make no mention of efficient discretizations of (1.4) for the (chaotic) velocity field (1.13)

(or the time continuous version (1.16)). Discretizations of (1.3) and (1.4) have been extensively studied by many authors [JKO98, DPZ17, Wib18, Che23], and an important question is to choose velocity that minimize the computational cost of such schemes.

Plan of this paper. In Section 2 we precisely define the notion of exponentially mixing, and split the proof of Theorem 1.1 into two steps: Obtaining dissipation time bounds (Theorem 2.1), and obtaining mixing time bounds (Proposition 2.4). We prove Theorem 2.1 in Section 3, and prove Proposition 2.4 in Section 4. In Section 5 we study a time discrete version of (1.4) and produce exponentially mixing maps for which the κ-dependence of the mixing rate is known. (This is used to motivate the heuristics (1.15).) In Section 6 we prove Theorem 1.3 and produce (random) velocity fields that are (almost surely) exponentially mixing. Finally, in Appendix A, we show that a family of randomly shifted localized tent shaped shear flows almost surely generates an exponentially mixing flow for the Lebesgue measure. We provide this simpler example since the proof is simpler than the proofs done in Section 6, and does not involve technical calculations checking the Hörmander type conditions.

2. Mixing Time Bounds (Theorem 1.1)

The goal of this section is to fix our notation, precisely recall the notions of mixing rate, mixing time and dissipation time (used in Theorems 1.1 and 1.3), and to state two stronger results that immediately imply Theorem 1.1. The first result (Theorem 2.1, below) obtains a dissipation time bound in terms of the mixing rate of the imposed velocity field v. When the mixing rate is exponential, this quickly reduces to the bound (1.7). The second result (Proposition 2.4, below) bounds the mixing time in terms of the dissipation time, allowing us to deduce (1.8) from (1.7). The heart of the matter lies in the proofs of Theorem 2.1 and Proposition 2.4, which we do in Sections 3 and 4 respectively.

2.1. **Mixing rates.** We begin by fixing our notation and precisely defining the notion of *exponentially mixing*, which was used in the statements of both Theorems 1.1 and 1.3. Throughout this paper we will always assume the potential U is a C^2 periodic function, p is defined by (1.1), ρ_∞ is defined by (1.2), and μ is the probability measure

$$\mu(dx) = \frac{p(x)}{Z}\, dx = \rho_\infty(x)\, dx\,.$$

Define the $L^2(\mu)$ inner-product, $L^2(\mu)$ norm, and $\dot{H}^1(\mu)$ norms by

$$\langle f, g\rangle_\mu = \int_{\mathbb{T}^d} fg\, d\mu\,, \quad \|f\|_{L^2(\mu)}^2 = \langle f, f\rangle_\mu\,, \quad \text{and} \quad \|f\|_{\dot{H}^1(\mu)} = \|\nabla f\|_{L^2(\mu)}\,.$$

respectively, and the corresponding spaces are defined in the usual way. Define

$$\dot{L}^2(\mu) = \left\{ f \in L^2(\mu) \,\Big|\, \int_{\mathbb{T}^d} f\, d\mu = 0 \right\},$$

to be the subspace of μ-mean-zero functions.

Given a (time dependent) C^1 velocity field v, define the flow to be the solution of the ODE

$$(2.1) \qquad \partial_t \Phi_{s,t} = v_t \circ \Phi_{s,t}, \quad \Phi_{s,s}(x) = x.$$

It is easy to verify that Φ preserves the measure μ if and only if v satisfies (1.5). The notion of *mixing* in dynamical systems, requires the flow to spread mass concentrated in a small region to the entire space as $t \to \infty$ (see for instance [KH95, SOW06]). More precisely, a flow is mixing if for every pair of test functions $f, g \in L^2(\mu)$, we have

$$(2.2) \qquad \lim_{t \to \infty} \langle f \circ \Phi_{s,s+t}^{-1}, g \rangle_\mu = \langle f, 1 \rangle_\mu \langle g, 1 \rangle_\mu.$$

Since Φ is invertible, one may equivalently replace $\Phi_{s,s+t}^{-1}$ above with $\Phi_{s,s+t}$.

The *mixing rate* is the rate at which the convergence in (2.2) happens for regular test functions f, g. The standard choice in dynamical systems is to choose f, g to be Hölder continuous, or C^1. However, for our purposes, it is more convenient to use Sobolev regular functions instead. For convenience, we will further assume the test functions are mean-zero so the right-hand side of (2.2) vanishes. We define the flow of v to be mixing with rate h if

$$(2.3) \qquad \sup_{f,g \in \dot{H}^1(\mu), s \in \mathbb{R}} \langle f \circ \Phi_{s,s+t}^{-1}, g \rangle_\mu \leqslant h(t) \|f\|_{\dot{H}^1(\mu)} \|g\|_{\dot{H}^1(\mu)}.$$

The function $h \colon [0, \infty) \to (0, \infty)$ above is called the mixing rate, and it is always assumed to be a continuous decreasing function that vanishes at infinity. The flow of v is said to be *exponentially mixing* if the mixing rate h is in the form (1.6) for some finite constants D, γ, that may depend on κ.

Constructing exponentially mixing flows is not an easy task, and has been studied extensively in the dynamical systems literature [Ano67, Pol85, Dol98, Liv04, BW20, TZ23]. Unfortunately, these results are all on manifolds other than the standard torus, which is not relevant to the scenario studied in the present paper. Several authors have recently constructed (time dependent) examples of exponential mixing on the standard torus [DKK04, ACM19, EZ19, BZ21, BBPS22, BCZG22]. However, all these examples preserve the Lebesgue measure and not the measure μ as we require. We will shortly show (Section 6, below) that the flow defined in (1.13) is exponentially mixing (as stated in Theorem 1.3). Postponing further discussion of this to Section 6, we will now show how such flows can be used to improve the mixing time of (1.4).

2.2. **Dissipation time bounds.** We now recall the notion of *dissipation time*, and provide an upper bound on the dissipation time in terms of the mixing rate of the flow. The results are similar to those in [FI19]. In our context, the added difficulty is that the measures μ concentrate in regions with volume $O(\kappa^{d/2})$, and so we need to track the dependence on both κ and the mixing rate.

Roughly speaking, the *dissipation time* (see [FW03, FI19]) measures the rate at which the distribution of X approaches stationary distribution μ in $L^2(\mu)$, when the initial distribution is also $L^2(\mu)$. Precisely, the *dissipation time* of the process X is defined by

$$t_{\text{dis}} \overset{\text{def}}{=} \inf \left\{ t \geqslant 0 \;\middle|\; \sup_{s \geqslant 0} \|\theta_t\|_{L^2(\mu)} \leqslant \frac{1}{2} \|f\|_{L^2(\mu)}, \; \forall f \in \dot{L}^2(\mu) \right\},$$

where

$$(2.4) \qquad \theta_t(x) = \boldsymbol{E}^{(x,s)} f(X_t) = \int_{\mathbb{T}^d} \rho(x, s; y, t) f(y) \, dy \,,$$

and $\rho(x, s; y, t)$ denotes the transition density of the process X.

The Poincaré inequality (Lemma 4.6, below) quickly implies that

$$(2.5) \qquad t_{\text{dis}} \leqslant \frac{1}{\kappa} \exp\Big(\frac{\|U\|_{\text{osc}}}{2\kappa}\Big) \,,$$

which is too large to be of practical interest. The first result we state is that if v is mixing, then it can be rescaled to ensure t_{dis} is much smaller than the right-hand side of (2.5).

Theorem 2.1. *Let $t_{\text{dis}} = t_{\text{dis}}(A, \kappa, v)$ be the dissipation time of the process X defined by (1.4). For every sufficiently small $\kappa > 0$, there exists $A_0 = A_0(\kappa) < \infty$, independent of the dimension, such that*

$$(2.6) \qquad t_{\text{dis}} = t_{\text{dis}}(\kappa, A) \leqslant \frac{16(1 + \ln 2)}{H(A)} \qquad \text{for all } A > A_0(\kappa) \,.$$

Here $H(A) = H(A, \kappa)$ is defined to be the unique solution of

$$(2.7) \qquad \frac{\kappa}{4H(A)} = h\Big(\frac{1}{16}\Big(\frac{A}{H(A)\|\nabla v\|_{L^\infty}}\Big)^{1/2}\Big) \,.$$

Remark 2.2. During the course of the proof of Theorem 2.1 we will see that A_0 should be chosen so that both

$$(2.8) \qquad A_0 \geqslant 256\Lambda\|\nabla v\|_{L^\infty}\Big(h^{-1}\Big(\frac{\kappa}{4\Lambda}\Big)\Big)^2, \quad \text{and} \quad H(A_0) \geqslant \frac{\kappa}{4h\Big(\frac{1}{32\sqrt{2}\|\nabla v\|_{L^\infty}}\Big)} \,.$$

Here $\Lambda = \Lambda(\kappa, d)$ is a constant that is chosen to ensure a growth condition on eigenvalues of the generator of (1.3) (see (3.23), below). By Weyl's law (specifically from (3.13)) one can check that $\Lambda = O(\kappa)$ as $\kappa \to 0$. Thus, in the case h is given by (1.6), the bound (2.8) reduces to (1.9) stated in Remark 1.2

Remark 2.3. Theorem 2.1 still holds when the state space is $\mathbb{R}^d$, provided the spectrum of the generator of (1.3) is discrete and the eigenvalues grow according to Weyl's law (as stated in Lemma 3.1, below). It is well known (see for instance Chapter 4 in [Pav14]) that both these conditions hold provided the potential U satisfies

$$\lim_{|x| \to +\infty} \Big(\frac{|\nabla U(x)|^2}{2} - \Delta U(x)\Big) = +\infty \,.$$

The main idea behind the proof of Theorem 2.1 is to obtain bounds on the $L^2(\mu)$ decay of solutions to the associated PDE. We do this by a spectral splitting method that is commonly used in the study of such equations. Namely, we divide the analysis into two cases: When θ has most of its energy in large frequencies, the standard energy inequality shows that $\|\theta\|_{L^2(\mu)}$ decays fast. On the other hand, when θ has most of its energy in small frequencies, the mixing caused by the convection term $v \cdot \nabla\theta$ generate high frequencies, which in turn forces $\|\theta\|_{L^2(\mu)}$ to decay fast. When the underlying measure is the Lebesgue measure on $\mathbb{T}^d$ a similar result was proved in [FI19], and our proof follows the same structure.

Once Theorem 2.1 is established, proving the upper bound (1.7) in Theorem 1.1 is simply a matter of choosing h to be the exponential (1.6), and simplifying (2.7). We do this in Section 2.4, below.

Notice that as $A \to \infty$, we must have $H(A) \to \infty$ and hence so t_{dis} can be made arbitrarily small. This, however, is not always computationally advantageous as solving (1.4) when A is large is very computationally intensive. Moreover, making t_{dis} small is not yet sufficient to guarantee solutions to (1.4) escape the metastable traps at local minima of U. Indeed, if X_0 is initially concentrated in a ball $B(x_0, \sqrt{\kappa})$, after time $t \geqslant t_{\text{dis}}$ the process X_t may still be concentrated in a ball of radius $B(x_0, C\sqrt{\kappa})$. Thus, we are not guaranteed X has explored the state space enough to escape metastable traps around local minima of U. We are however close: in the next section we will show that in an additional $O(t_{\text{dis}}/\kappa)$ time, the process X will escape metastable traps be close to mixed.

2.3. **Mixing time bounds.** We will now study the relation between t_{dis} and t_{mix}. Recall [LP17] the *mixing time* of a Markov process is the time taken for its distribution to become sufficiently close (in the total variation norm) to its stationary distribution. In our context, the mixing time of X can be defined by

$$(2.9) \qquad t_{\text{mix}} \overset{\text{def}}{=} \inf\left\{ t \geqslant 0 \,\Big|\, \sup_{x \in \mathbb{T}^d,\, s \geqslant 0} \int_{\mathbb{T}^d} \left| \rho(s, x; s+t, y) - \rho_\infty(y) \right| dy \leqslant \frac{1}{2} \right\},$$

where, as before, ρ is the transition density of X.

The mixing time is a stronger notion than the dissipation time. As mentioned earlier, waiting for time t_{dis} will not ensure X has escaped the metastable traps at local minima of U; however waiting for time t_{mix} will certainly ensure this.

It is easy to see t_{mix} always dominates t_{dis} (see for instance [IZ22, ILN23]). Our interest, however, is to control the mixing time by the dissipation time. The advantage of this is that the dissipation time can be bounded using $L^2(\mu)$ based spectral methods, such as those used in the proof of Theorem 2.1. Thus controlling t_{mix} by t_{dis} will allow us to use Theorem 2.1 to obtain upper bounds on the mixing time. Our next result provides upper (and lower) bounds on the mixing time in terms of the dissipation time.

Proposition 2.4. *There exists a universal (dimension independent) constant C, that is independent of U, v, κ and A, such that*

$$(2.10) \qquad \frac{t_{\text{dis}}}{3} \leqslant t_{\text{mix}} \leqslant Cd\Big(1 + \frac{\|U\|_{\text{osc}}}{\kappa} - \ln(\kappa t_{\text{dis}})\Big) t_{\text{dis}}.$$

Remark 2.5. The Poincaré Inequality (Lemma 4.6, below), will guarantee that $\ln(\kappa t_{\text{dis}}) \leqslant \|U\|_{\text{osc}}/(2\kappa)$, and so the factor on the right of (2.10) is nonnegative.

The reason for the large factor on the right of (2.10) is as follows. Since the noise is regular, the density $\rho(s, x; s+t, y)$ becomes $L^2(\mu)$ for any $t > 0$. However, the $L^2(\mu)$ norm is typically of order $O(1/(\kappa t)^{d/2})$, which is large. Waiting for a large multiple of t_{dis} will now make this small, which is what leads to the large factor on the right of (2.10).

To carry out these details, we prove a stronger $L^1(\mu) \to L^\infty$ bound on the transition density, with constants that are independent of v. The proof follows the structure of similar bounds on the torus with respect to the Lebesgue measure (see [CKRZ08]). The key identity that allows us to make the proof work in our

situation is that the ratio ρ/ρ_∞ satisfies an equation that differs from the Kolmogorov backward equation by only a sign (see Lemma 4.2, below).

Remark 2.6. The lower bound in (2.10) works in a general setting and, in particular, works when the state space is $\mathbb{R}^d$. The upper bound, however, requires certain assumptions and our proof does not presently work when the state space is $\mathbb{R}^d$. Specifically, our proof relies on an $L^1 \to L^\infty$ bound on the transition density (see Lemma 4.1), with constants that are independent of v. This in turn relies on a Nash inequality (Lemma 4.4), which is verifiably false in $\mathbb{R}^d$. Indeed, choosing $f_n = |x|^n$ shows Lemma 4.4 can not hold on $\mathbb{R}^d$ with the potential $U = |x|^2$.

Of course, Proposition 2.4 and the dissipation time bound (1.7) immediately imply the mixing time bound (1.8) stated in Theorem 1.1. We prove Theorem 1.1 in the next section, and postpone the proofs of Theorem 1.3 and Proposition 2.4 to Sections 3 and 4 respectively.

2.4. **Proof of Theorem 1.1.** The proof of Theorem 1.1 now follows by simplifying (2.7) when h is given by (1.6), and using Proposition 2.4. We carry out the details here.

Proof of Theorem 1.1. By Theorem 2.1 we know t_{dis} is bounded by (2.6) where $H(A)$ is defined by (2.7). In order to bound $H(A)$, we make the following simple observation: If T satisfies

$$(2.11) \qquad T = ae^{-b\sqrt{T}},$$

for some constants $a, b > 0$, then we must have

$$(2.12) \qquad T \leqslant \frac{1 + \ln^2(ab^2)}{b^2}.$$

To see this, note that if $T \leqslant 1/b^2$ there is nothing to prove. If $T \geqslant 1/b^2$, then (2.11) implies

$$T = \frac{1}{b^2} \ln^2\left(\frac{a}{T}\right) \leqslant \frac{\ln^2(ab^2)}{b^2},$$

which proves (2.12). Choosing

$$T = \frac{1}{H(A)}, \quad a = \frac{2D}{\kappa}, \quad \text{and} \quad b = \frac{\gamma\sqrt{A}}{16\sqrt{\|\nabla v\|_{L^\infty}}}$$

and using (2.12) in (2.7) immediately implies (1.7) as desired.

The mixing time bound (1.8) follows immediately from (1.7) and Proposition 2.4, concluding the proof. $\qquad\square$

3. **Dissipation time bound (Theorem 2.1)**

In this section we prove Theorem 2.1. The proof is entirely based on PDE techniques. Indeed, the function θ defined by (2.4) is the solution to the Kolmogorov backward equation

$$(3.1) \qquad \partial_t\theta = Av' \cdot \nabla\theta + \mathcal{L}_\kappa\theta, \quad \text{with initial data } \theta_s = f.$$

Here v' is the time changed velocity field

$$(3.2) \qquad v'_t = v_{At},$$

and $\mathcal{L}_\kappa$, defined by

$$\mathcal{L}_\kappa f = -\nabla U \cdot \nabla f + \kappa \Delta f \,,$$

is the generator of (1.3).

The main idea behind the proof is to split the analysis into two cases. When θ has most of its energy in large frequencies, the operator $\mathcal{L}_\kappa$ will provide a strong damping effect, and $\|\theta\|_{L^2(\mu)}$ decays fast. On the other hand, when θ has most of its energy in small frequencies, the mixing caused by the convection term $v' \cdot \nabla\theta$ generate high frequencies, which in turn forces $\|\theta\|_{L^2(\mu)}$ to decay fast.

To carry out the details we need a few spectral properties of the operator $\mathcal{L}_\kappa$, which we collect here for easy reference.

Lemma 3.1. *(1) The operator $-\mathcal{L}_\kappa$ is self-adjoint and nonnegative with respect to the inner-product $\langle \cdot, \cdot \rangle_\mu$.*
(2) For all $f, g \in H^1(\mathbb{T}^d, \mu)$ we have

$$-\langle \mathcal{L}_\kappa f, g \rangle_\mu = \kappa \langle \nabla f, \nabla g \rangle_\mu \,. \tag{3.3}$$

(3) The spectrum of $-\mathcal{L}_\kappa$ is discrete, and the smallest eigenvalue on $\dot{L}^2(\mu)$ is strictly positive. Moreover, if $0 < \lambda_0 \leqslant \lambda_1 \leqslant \lambda_2 \ldots$ are the eigenvalues of $-\mathcal{L}_\kappa$ in ascending order, then

$$\lambda_n \xrightarrow{n \to \infty} \infty \qquad and \qquad \frac{\lambda_{n+1}}{\lambda_n} \xrightarrow{n \to \infty} 1 \,,$$

Lemma 3.1 directly follows from Weyl's law [HS21], and is presented later in this section. We now state two lemmas which show fast energy decay both when θ has mainly high frequencies, and when it does not.

Lemma 3.2. *Solutions to (3.1) satisfy the energy inequality*

$$\partial_t \|\theta\|_{L^2(\mu)}^2 = -2\langle \mathcal{L}_\kappa \theta, \theta \rangle_\mu = -2\kappa \|\nabla\theta\|_{L^2(\mu)}^2 \,. \tag{3.4}$$

Consequently, if for all $t \in [0, T]$ we have

$$\kappa \|\nabla\theta_t\|_{L^2(\mu)}^2 \geqslant \lambda \|\theta_t\|_{L^2(\mu)}^2 \,,$$

for some constant $\lambda > 0$. Then for all $t \in [0, T]$ we must have

$$\|\theta_t\|_{L^2(\mu)}^2 \leqslant \exp\left(-2\lambda t\right)\|\theta_0\|_{L^2(\mu)} \,. \tag{3.5}$$

Lemma 3.3. *Let λ_N be the largest eigenvalue of $\mathcal{L}_\kappa$ such that $\lambda_N \leqslant H(A)$. If*

$$\kappa \|\nabla\theta_s\|_{L^2(\mu)}^2 \leqslant \lambda_N \|\theta_s\|_{L^2(\mu)}^2 \,, \tag{3.6}$$

then

$$\|\theta_{s+t_0}\|_{L^2(\mu)}^2 \leqslant \exp\left(-\frac{H(A)t_0}{8}\right)\|\theta_s\|_{L^2(\mu)}^2 \,, \tag{3.7}$$

where t_0 is given by

$$t_0 \stackrel{\text{def}}{=} \frac{2}{A} \, h^{-1}\left(\frac{\kappa}{4\lambda_N}\right) \,. \tag{3.8}$$

Since the proof of Lemma 3.2 is short, we present it first.

Proof of Lemma 3.2. Equation (3.4) follows by multiplying (3.1) by θ, integrating by parts, and using (1.5) and (3.3). Equation (3.5) follows immediately from (3.4) and Grönwall's lemma. □

The proof of Lemma 3.3 is more involved and relies on the mixing properties of v. The main idea is that when the spectrum of θ_s is concentrated in low frequencies, then it is close to the solution of the transport equation,

$$\partial_t \phi = A v' \cdot \nabla \phi, \quad \phi(0) = \theta_0.$$

The mixing assumption on v guarantees that the transport equation moves energy to high frequencies. These high frequencies are then dissipated faster by $\mathcal{L}_\kappa$, leading to the faster decay stated in Lemma 3.3. Postponing the proof of Lemma 3.3 to Section 3.2, we now prove Theorem 2.1.

Proof of Theorem 2.1. We claim that any solution of (3.1) with $\theta_0 \in \dot{L}^2(\mu)$ satisfies

$$(3.9) \qquad \|\theta_{s+t}\|_{L^2(\mu)} \leqslant \exp\Big(-\frac{H(A)t}{16} + 1\Big)\|\theta_s\|_{L^2(\mu)},$$

which immediately implies (2.6).

To prove (3.9), we may without loss of generality assume $s = 0$. Choose λ_N as in Lemma 3.3, choose $\lambda = \lambda_N$, and repeatedly apply Lemmas 3.2 and 3.3 to obtain an increasing sequence of times $t'_k \to \infty$ such that $t'_{k+1} - t'_k \leqslant t_0$ and

$$(3.10) \qquad \|\theta_{t'_{k+1}}\|^2_{L^2(\mu)} \leqslant \exp\Big(-(t'_{k+1} - t'_k)\min\Big\{\lambda_N, \frac{H(A)}{8}\Big\}\Big)\|\theta_{t'_k}\|^2_{L^2(\mu)},$$

By Lemma 3.1 there exists $\Lambda = \Lambda(\kappa)$ such that

$$\lambda_{n+1} \leqslant 8\lambda_n \quad \text{whenever } \lambda_{n+1} \geqslant \Lambda.$$

Let A_0 be defined by (2.8), and note that for $A \geqslant A_0$ we have $H(A) \geqslant \Lambda$. This implies

$$\frac{H(A)}{8} \leqslant \lambda_N \leqslant H(A),$$

and hence (3.10) implies

$$(3.11) \qquad \|\theta_{t'_{k+1}}\|^2_{L^2(\mu)} \leqslant \exp\Big(-(t'_{k+1} - t'_k)\frac{H(A)}{8}\Big)\|\theta_{t'_k}\|^2_{L^2(\mu)}.$$

By construction of t'_k, for any $t \geqslant 0$ there exists $k \in \mathbb{N}$ such that $t - t_0 \leqslant t'_k \leqslant t$. Iterating (3.11) shows

$$(3.12) \quad \|\theta_t\|_{L^2(\mu)} \leqslant \exp\Big(-\frac{H(A)t'_k}{16}\Big)\|\theta_0\|_{L^2(\mu)} \leqslant \exp\Big(-\frac{H(A)(t - t_0)}{16}\Big)\|\theta_0\|_{L^2(\mu)}.$$

By choice of t_0 and λ_N, we note

$$H(A)t_0 = \frac{2}{A}h^{-1}\Big(\frac{\kappa}{4\lambda_N}\Big)H(A) \leqslant \frac{H(A)}{8\sqrt{A\lambda_N}\|\nabla v\|_{L^\infty}}$$

$$\leqslant \Big(\frac{H(A)}{8A\|\nabla v\|_{L^\infty}}\Big)^{1/2} = \frac{1}{32\sqrt{2}\|\nabla v\|_{L^\infty}h^{-1}\big(\frac{\kappa}{4H(A)}\big)} \leqslant 1.$$

Using this in (3.12) implies (3.9) as desired. $\qquad\qquad\square$

The remainder of this section is devoted to proving Lemmas 3.1 and 3.3.

3.1. **Spectral bounds on $\mathcal{L}_\kappa$ (Lemma 3.1).** The operator $\mathcal{L}_\kappa$ can be conjugated to a Schrödinger operator and well-known results spectral results (e.g. [HS21]) for Schrödinger operators will imply Lemma 3.1.

Proof of Lemma 3.1. The first two assertions of Lemma 3.1 are direct computations. Indeed,

$$-\mathcal{L}_\kappa = -\kappa\Delta + \nabla U \cdot \nabla = -\kappa e^{U/\kappa}\nabla \cdot (e^{-U/\kappa}\nabla)\,,$$

and hence for all $f, g \in H^1(\mathbb{T}^d, \mu)$, we have

$$-\langle \mathcal{L}_\kappa f, g\rangle_\mu = -\kappa\int_{\mathbb{T}^d} e^{U/\kappa}\nabla \cdot (e^{-U/\kappa}\nabla f)g\,\rho_\infty\,dx = \frac{-\kappa}{Z}\int_{\mathbb{T}^d}\nabla\cdot(e^{-U/\kappa}\nabla f)g\,dx$$

$$= \frac{\kappa}{Z}\int e^{-U/\kappa}\nabla f\cdot\nabla g\,dx = \kappa\langle\nabla f, \nabla g\rangle_\mu\,.$$

This immediately implies the first two assertions.

To study the spectrum, let $L^2 = L^2(\mathbb{T}^d)$ denote the space of all square-integrable functions with respect to the Lebesgue measure, and $\langle\cdot,\cdot\rangle$ the associated inner-product. Define the operator $\mathcal{U}\colon L^2(\mu)\to L^2$ by

$$\mathcal{U}f = \frac{1}{\sqrt{Z}}e^{-U/2\kappa}f\,.$$

Clearly $\langle f, g\rangle_\mu = \langle f, g\rangle$, and so $\mathcal{U}$ is an isometry. Define the operator $\mathcal{H}\colon L^2\to L^2$ by $\mathcal{H}\stackrel{\text{def}}{=}\mathcal{U}\mathcal{L}_\kappa\mathcal{U}^{-1}$. We compute

$$-\mathcal{H}f = -\kappa\Delta f + \Big(\frac{1}{4}|\nabla U|^2 - \frac{1}{2}\Delta U\Big)f\,.$$

Thus $\mathcal{L}_\kappa$ is unitarily equivalent to the operator $\mathcal{H}$, and hence the operators $\mathcal{L}$ and $\mathcal{U}$ have the same spectrum.

The operator $\mathcal{H}$ is a Schrödinger operator and has been extensively studied. In particular, the eigenvalues of $\mathcal{H}$ satisfy Weyl's law [HS21] (see also [Ray54, Theorem VI]), which states that

$$(3.13) \qquad N(\lambda)\stackrel{\text{def}}{=}\sum_{\lambda_n<\lambda}1 = \frac{\omega_d}{(2\pi)^d}\Big(\frac{\lambda}{\kappa}\Big)^{\frac{d}{2}} + O\Big(\frac{\lambda}{\kappa}\Big)^{\frac{d-1}{2}}\,,$$

asymptotically, as $\lambda\to\infty$. Here ω_d is the volume of the unit ball in $\mathbb{R}^d$. This immediately implies the third assertion in Lemma 3.1, finishing the proof. $\qquad\square$

3.2. **Low frequency energy decay (Lemma 3.3).** To prove Lemma 3.3, we will first show (Lemma 3.4, below) that when (3.6) holds, θ is sufficiently close to solutions to the transport equation (3.14). By the mixing assumption on v this will move energy to high frequencies, which will then be dissipated faster by the diffusion operator $\mathcal{L}_\kappa$.

Lemma 3.4. *Let θ be the solution of (3.1) with initial data $\theta_0 \in \dot{L}^2(\mu)$, and let ϕ be a solution of the transport equation*

$$(3.14) \qquad \partial_t\phi = Av'\cdot\nabla\phi\,, \quad \phi(0) = \theta_0\,,$$

with the same initial data. For any $t \geqslant 0$ we have

$$\|\theta - \phi\|_{L^2(\mu)}^2 \leqslant \sqrt{2\kappa t}\|\theta_0\|_{L^2(\mu)}\Big(2A\|\nabla v\|_{L^\infty}\int_0^t\|\nabla\theta\|_{L^2(\mu)}^2\,ds + \|\nabla\theta_0\|_{L^2(\mu)}^2\Big)^{1/2}\,.$$

Proof. Multiplying (3.1) by $-\mathcal{L}_\kappa\theta$ and integrating over space gives

$$\frac{\kappa}{2}\partial_t\langle\nabla\theta,\nabla\theta\rangle_\mu = -\frac{1}{2}\partial_t\langle\theta,\mathcal{L}_\kappa\theta\rangle_\mu = -\langle\partial_t\theta,\mathcal{L}_\kappa\theta\rangle_\mu$$

$$(3.15) \qquad\qquad = -\|\mathcal{L}_\kappa\theta\|^2_{L^2(\mu)} - A\langle v'\cdot\nabla\theta,\mathcal{L}_\kappa\theta\rangle_\mu\,.$$

For the last term, we note

$$\langle v'\cdot\nabla\theta,\mathcal{L}_\kappa\theta\rangle_\mu = \frac{\kappa}{Z}\int_{\mathbb{T}^d} v'\cdot\nabla\theta\nabla\cdot(e^{-U/\kappa}\nabla\theta)\,dx$$

$$= -\frac{\kappa}{Z}\int_{\mathbb{T}^d}\nabla(v'\cdot\nabla\theta)\cdot\nabla\theta e^{-U/\kappa}\,dx$$

$$= \frac{\kappa}{2Z}\int_{\mathbb{T}^d} v'e^{-U/\kappa}\cdot\nabla(|\nabla\theta|^2)\,dx + \kappa\int_{\mathbb{T}^d}(\nabla\theta\cdot\nabla)v'\cdot\nabla\theta\,d\mu$$

$$= \kappa\int_{\mathbb{T}^d}(\nabla\theta\cdot\nabla)v'\cdot\nabla\theta\,d\mu\,,$$

since v' satisfies (1.5). Consequently,

$$\left|\langle v'\cdot\nabla\theta,\mathcal{L}_\kappa\theta\rangle_\mu\right| \leqslant \kappa\int_{\mathbb{T}^d}|\nabla v'||\nabla\theta|^2\,d\mu \leqslant \kappa\|\nabla v\|_{L^\infty}\|\nabla\theta\|^2_{L^2(\mu)}\,,$$

and substituting this in (3.15), we have

$$(3.16) \qquad \|\mathcal{L}_\kappa\theta\|^2_{L^2(\mu)} \leqslant A\kappa\|\nabla v\|_{L^\infty}\|\nabla\theta\|^2_{L^2(\mu)} - \frac{\kappa}{2}\partial_t\langle\nabla\theta,\nabla\theta\rangle_\mu\,.$$

On the other hand, $\|\theta - \phi\|_{L^2(\mu)}$ satisfies

$$\partial_t\|\theta - \phi\|^2_{L^2(\mu)} = 2\langle\mathcal{L}_\kappa\theta,\theta-\phi\rangle_\mu \leqslant -2\langle\mathcal{L}_\kappa\theta,\phi\rangle_\mu \leqslant 2\|\mathcal{L}_\kappa\theta\|_{L^2(\mu)}\|\phi\|_{L^2(\mu)}$$

$$(3.17) \qquad\qquad \leqslant 2\|\mathcal{L}_\kappa\theta\|_{L^2(\mu)}\|\theta_0\|_{L^2(\mu)}\,.$$

To obtain the last equality above we used the fact that $\|\phi_t\|_{L^2(\mu)} = \|\phi_0\|_{L^2(\mu)}$, which is true because v' satisfies (1.5). Integrating (3.17) and applying (3.16), we have

$$\|\theta-\phi\|^2_{L^2(\mu)} \leqslant 2\|\theta_0\|_{L^2(\mu)}\int_0^t\|\mathcal{L}_\kappa\theta\|_{L^2(\mu)}\,ds$$

$$\leqslant 2\sqrt{t}\|\theta_0\|_{L^2(\mu)}\left(\int_0^t A\kappa\|\nabla v\|_{L^\infty}\|\nabla\theta\|^2_{L^2(\mu)} - \frac{\kappa}{2}\partial_t\langle\nabla\theta,\nabla\theta\rangle_\mu\,ds\right)^{1/2}$$

$$\leqslant \sqrt{2\kappa t}\|\theta_0\|_{L^2(\mu)}\left(2A\|\nabla v\|_{L^\infty}\int_0^t\|\nabla\theta\|^2_{L^2(\mu)}\,ds + \|\nabla\theta_0\|^2_{L^2(\mu)}\right)^{1/2},$$

concluding the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\Box$

We now use Lemma 3.4 to prove Lemma 3.3.

Proof of Lemma 3.3. For simplicity, and without loss of generality, we assume $s = 0$. We claim our choice of λ_N and t_0 will guarantee

$$(3.18) \qquad \kappa\int_0^{t_0}\|\nabla\theta\|^2_{L^2(\mu)}\,ds \geqslant \frac{\lambda_N t_0}{8}\|\theta_0\|^2_{L^2(\mu)}\,.$$

To prove this, assume, for sake of contradiction, that

$$(3.19) \qquad \kappa\int_0^{t_0}\|\nabla\theta\|^2_{L^2(\mu)}\,ds < \frac{\lambda_N t_0}{8}\|\theta_0\|^2_{L^2(\mu)}\,.$$

Let P_N be the orthogonal projection from $\dot{L}^2(\mu)$ to the space spanned by the first N eigenfunctions $\{e_1, \ldots, e_N\}$, recall (3.3) and notice

$$\int_0^{t_0} \|\nabla \theta\|_{L^2(\mu)}^2 \, ds \geqslant \frac{\lambda_N}{\kappa} \int_{\frac{t_0}{2}}^{t_0} \|(I - P_N)\theta\|_{L^2(\mu)}^2 \, ds$$

$$\geqslant \frac{\lambda_N}{2\kappa} \int_{\frac{t_0}{2}}^{t_0} \|(I - P_N)\phi\|_{L^2(\mu)}^2 \, ds - \frac{\lambda_N}{\kappa} \int_{\frac{t_0}{2}}^{t_0} \|(I - P_N)(\theta - \phi)\|_{L^2(\mu)}^2 \, ds$$

$$(3.20) \quad \geqslant \frac{\lambda_N t_0}{4\kappa} \|\theta_0\|_{L^2(\mu)}^2 - \frac{\lambda_N}{2\kappa} \int_{\frac{t_0}{2}}^{t_0} \|P_N \phi\|_{L^2(\mu)}^2 \, ds - \frac{\lambda_N}{\kappa} \int_{\frac{t_0}{2}}^{t_0} \|\theta - \phi\|_{L^2(\mu)}^2 \, ds.$$

We will now bound each of the negative terms on the right of (3.20). For the second term in (3.20), we first note that the mixing rate of the rescaled velocity field $Av' = Av_{At}$ is $h(At)$. Thus, by (2.3),

$$\|P_N \phi_s\|_{L^2(\mu)}^2 = \langle \phi_s, P_N \phi_s \rangle_{L^2(\mu)} = \langle \theta_0 \circ \Phi_{0,s}^{-1}, P_N \phi_s \rangle_{L^2(\mu)}$$

$$\leqslant h(As) \|\theta_0\|_{\dot{H}^1(\mu)} \|P_N \phi_s\|_{\dot{H}^1(\mu)} \leqslant \sqrt{\lambda_N/\kappa} \, h(As) \|\theta_0\|_{\dot{H}^1(\mu)} \|\phi_s\|_{L^2(\mu)}$$

$$= \sqrt{\lambda_N/\kappa} \, h(As) \|\theta_0\|_{L^2} \|\theta_0\|_{\dot{H}^1(\mu)}.$$

Combining this with (3.6) implies

$$(3.21) \quad \int_{\frac{t_0}{2}}^{t_0} \|P_N \phi\|_{L^2(\mu)}^2 \, ds \leqslant \sqrt{\lambda_N/\kappa} \int_{\frac{t_0}{2}}^{t_0} h(As) \|\theta_0\|_{\dot{H}^1(\mu)} \|\theta_0\|_{L^2(\mu)} \, ds$$

$$\leqslant \frac{\lambda_N t_0}{2\kappa} h\left(\frac{At_0}{2}\right) \|\theta_0\|_{L^2(\mu)}^2.$$

For the last term in (3.20), applying Lemma 3.4 and inequality (3.19), we note

$$\int_{\frac{t_0}{2}}^{t_0} \|\theta - \phi\|_{L^2(\mu)}^2 \, ds$$

$$\leqslant \int_{\frac{t_0}{2}}^{t_0} \sqrt{2\kappa\tau} \|\theta_0\|_{L^2(\mu)} \left(2A \|\nabla v\|_{L^\infty} \int_0^\tau \|\nabla \theta\|_{L^2(\mu)}^2 \, ds + \|\nabla \theta_0\|_{L^2(\mu)}^2\right)^{1/2} d\tau$$

$$\leqslant \frac{\sqrt{2\kappa} t_0^{3/2} \|\theta_0\|_{L^2(\mu)}}{2} \left(\frac{A \lambda_N t_0 \|\nabla v\|_{L^\infty}}{4\kappa} \|\theta_0\|_{L^2(\mu)}^2 + \|\nabla \theta_0\|_{L^2(\mu)}^2\right)^{1/2}.$$

Using (3.6) this gives

$$\int_{\frac{t_0}{2}}^{t_0} \|\theta - \phi\|_{L^2(\mu)}^2 \, ds \leqslant \frac{\sqrt{2\kappa} t_0^{3/2} \|\theta_0\|_{L^2(\mu)}^2}{2} \left(\frac{A \lambda_N t_0 \|\nabla v\|_{L^\infty}}{4\kappa} + \frac{\lambda_N}{\kappa}\right)^{1/2}$$

$$(3.22) \quad \leqslant \frac{\sqrt{2} t_0^{3/2} \|\theta_0\|_{L^2(\mu)}^2}{2} \left(\frac{A \lambda_N t_0 \|\nabla v\|_{L^\infty}}{2}\right)^{1/2}.$$

To obtain the last inequality above, we used the fact that by the choice of t_0 in (3.8) guarantees $At_0 \gg 1$.

Using (3.21) and (3.22) in (3.20) gives

$$\kappa \int_0^{t_0} \|\nabla \theta\|_{L^2(\mu)}^2 \, ds = \frac{\lambda_N t_0}{4} \|\theta_0\|_{L^2(\mu)}^2 \left(1 - \frac{\lambda_N}{\kappa} h\left(\frac{At_0}{2}\right) - 2t_0 \sqrt{A \lambda_N \|\nabla v\|_{L^\infty}}\right)$$

$$= \frac{\lambda_N t_0}{4} \|\theta_0\|_{L^2(\mu)}^2 \left(1 - \frac{1}{4} - 4h^{-1}\left(\frac{\kappa}{4\lambda_N}\right) \sqrt{\frac{\lambda_N \|\nabla v\|_{L^\infty}}{A}}\right),$$

where we used (3.8) to obtain the last equality. Since the function

$$\lambda \mapsto \sqrt{\lambda} h^{-1}\left(\frac{\kappa}{4\lambda}\right)$$

is increasing, the definition of $H(A)$ (equation (2.7)) and the fact that $\lambda_N \leqslant H(A)$ imply

$$\kappa \int_0^{t_0} \|\nabla\theta\|_{L^2(\mu)}^2 \, ds = \frac{\lambda_N t_0}{4} \|\theta_0\|_{L^2(\mu)}^2 \left(1 - \frac{1}{4} - \frac{1}{4}\right) \geqslant \frac{\lambda_N t_0}{8} \|\theta_0\|_{L^2(\mu)}^2 \,.$$

This contradicts (3.19), and hence concludes the proof of (3.18).

Finally, to obtain (3.7) we use Lemma 3.1 to find $\Lambda = \Lambda(\kappa)$ such that

$$(3.23) \qquad\qquad \lambda_{n+1} \leqslant 2\lambda_n \quad \text{whenever } \lambda_{n+1} \geqslant \Lambda \,.$$

If A_0 is chosen according to (2.8), then we will have $H(A) \geqslant \Lambda$, and hence

$$\frac{H(A)}{2} \leqslant \lambda_N \leqslant H(A) \,.$$

Combining this with (3.18) we obtain

$$\kappa \int_0^{t_0} \|\nabla\theta\|_{L^2(\mu)}^2 \, ds \geqslant \frac{H(A) t_0}{16} \|\theta_0\|_{L^2(\mu)}^2 \,.$$

Now the energy equality (3.4) implies (3.7) as desired. $\qquad\qquad\square$

4. Bounding t_{mix} in terms of t_{dis} (Proposition 2.4)

4.1. **The lower bound.** It is natural to expect that the mixing time controls the dissipation time in a general setting, and a similar result appeared recently in [IZ22]. Roughly speaking, to bound the mixing time, we need to start X with *any* initial distribution and show that the distribution of X becomes close to the invariant distribution in the total variation norm. To bound the dissipation time, we only need to consider L^2 initial distribution and bound the distance to the invariant distribution in a weaker sense. As a result, the lower bound in Proposition 2.4 is true in a more general setting, and the proof we present doesn't rely on the specific structure of (1.4).

Proof of the lower bound in Proposition 2.4. Let $f \in \dot{L}^2(\mu)$, and define

$$\theta_t(x) = \boldsymbol{E}^{(x,s)} f(X_{s+t}) = \int_{\mathbb{T}^d} \rho(x, s; y, s+t) f(y) \, dy \,.$$

In order to prove the lower bound in (2.10), we need to show that for any $t \geqslant 3t_{\mathrm{mix}}$, we have

$$(4.1) \qquad\qquad \|\theta_t\|_{L^2(\mu)} \leqslant \frac{1}{2} \|f\|_{L^2(\mu)} \,.$$

To prove this, we may without loss of generality assume $s = 0$. For notational convenience, we write

$$(4.2) \qquad\qquad \rho_t(x, y) = \rho(x, 0; y, t) \,.$$

Since $f \in \dot{L}^2(\mu)$ we note

$$\theta_t(x) = \int_{\mathbb{T}^d} \rho_t(x, y) f(y) \, dy = \int_{\mathbb{T}^d} \left(\rho_t(x, y) - \rho_\infty(y)\right) f(y) \, dy$$

and hence

$$\|\theta_t\|_{L^2(\mu)}^2 = \int_{\mathbb{T}^d} \left(\int_{\mathbb{T}^d} (\rho_t(x,y) - \rho_\infty(y))\, f(y)\, dy \right)^2 \rho_\infty(x)\, dx$$

$$\leqslant \int_{\mathbb{T}^d} \left(\int_{\mathbb{T}^d} |\rho_t(x,y) - \rho_\infty(y)| f(y)^2\, dy \int_{\mathbb{T}^d} |\rho_t(x,y) - \rho_\infty(y)|\, dy \right) \rho_\infty(x)\, dx\,.$$

Since t_{mix} is the mixing time, for any $t \geqslant n\, t_{\mathrm{mix}}$ we must have

$$\sup_{x \in \mathbb{T}^d} \int_{\mathbb{T}^d} |\rho_t(x,y) - \rho_\infty(y)|\, dy \leqslant \frac{1}{2^n}\,.$$

This implies

$$\|\theta_t\|_{L^2(\mu)}^2 \leqslant \frac{1}{2^n} \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} |\rho_t(x,y) - \rho_\infty(y)| f(y)^2 \rho_\infty(x)\, dy\, dx$$

(4.3)
$$\leqslant \frac{1}{2^n} \int_{\mathbb{T}^d} \int_{\mathbb{T}^d} (\rho_t(x,y) + \rho_\infty(y))\, f(y)^2 \rho_\infty(x)\, dx\, dy\,.$$

Since ρ_∞ is the density of the invariant measure, we know

$$\int_{\mathbb{T}^d} \rho_t(x,y)\rho_\infty(x)\, dx = \rho_\infty(y) \qquad \text{and} \qquad \int_{\mathbb{T}^d} \rho_\infty(x)\, dx = 1\,.$$

Using this in (4.3) implies

$$\|\theta_t\|_{L^2(\mu)}^2 \leqslant 2^{1-n} \|f\|_{L^2(\mu)}^2\,,$$

and choosing $n = 3$ implies (4.1) as desired. $\qquad\square$

4.2. **The upper bound.** To control the mixing time by the dissipation time we need to use the regularizing effects of the noise. More precisely, given any initial distribution, the noise regularizes it and the density becomes square-integrable, but with a large L^2 norm. Now waiting some multiple of the dissipation time will ensure mixing.

We implement the above idea by starting with an $L^1 \to L^\infty$ bound on the transition density ρ. This is the analog of the well-known drift independent $L^1 \to L^\infty$ estimates in [CKRZ08] in the case where the underlying measure is μ instead of the Lebesgue measure.

Lemma 4.1. *When $d \geqslant 3$, for every $x \in \mathbb{T}^d$, $s > s' \geqslant 0$, and $t > 0$ we have*

(4.4)
$$\left\| \frac{\rho_{s+t}^{(x,s')}}{\rho_\infty} - 1 \right\|_{L^\infty} \leqslant \frac{C_1(d)}{(\kappa t)^{d/2}} \exp\left(\frac{2d\|U\|_{\mathrm{osc}}}{\kappa} \right) \left\| \frac{\rho_s^{(x,s')}}{\rho_\infty} - 1 \right\|_{L^1(\mu)},$$

where $\rho_t^{(x,s)}$ denotes the transition density $\rho(x,s;\cdot,t)$, and C_1 is a dimensional constant that can be bounded by

(4.5)
$$C_1(d) \leqslant C 2^d\,,$$

where C is a universal constant independent of d. When $d = 2$, the inequality (4.12) needs to be replaced by

(4.6)
$$\left\| \frac{\rho_{s+t}^{(x,s')}}{\rho_\infty} - 1 \right\|_{L^\infty} \leqslant \frac{C_1'(\varepsilon)}{(\kappa t)^{1+\varepsilon}} \exp\left(\frac{(4+4\varepsilon)\|U\|_{\mathrm{osc}}}{\kappa} \right) \left\| \frac{\rho_s^{(x,s')}}{\rho_\infty} - 1 \right\|_{L^1(\mu)},$$

where $\varepsilon > 0$, and $C_1'(\varepsilon)$ is an ε-dependent constant.

Momentarily postponing the proof of Lemma 4.1, we now prove the upper bound in Proposition 2.4 and control the mixing time in terms of the dissipation time.

Proof of the upper bound in Proposition 2.4. For simplicity, and without loss of generality, we will again assume $s = 0$, and abbreviate the transition density as in (4.2). We will also assume $d \geqslant 3$. The proof when $d = 2$ is similar, and follows by choosing $\varepsilon > 0$ and replacing our use of (4.4) with (4.6).

When $d \geqslant 3$, inequality (4.4) implies

$$\|\rho_t(x, \cdot) - \rho_\infty\|_{L^1} = \int_{\mathbb{T}^d} |\rho_t(x, y) - \rho_\infty(y)|\, dy \leqslant \left\| \frac{\rho_t(x, \cdot)}{\rho_\infty} - 1 \right\|_{L^\infty}$$

$$\leqslant C_1 \left(\frac{4}{\kappa t} \right)^{\frac{d}{2}} \exp\left(\frac{2d\|U\|_{\mathrm{osc}}}{\kappa} \right) \left\| \frac{\rho_{3t/4}(x, \cdot)}{\rho_\infty} - 1 \right\|_{L^1(\mu)}$$

$$\leqslant C_1 \left(\frac{4}{\kappa t} \right)^{\frac{d}{2}} \exp\left(\frac{2d\|U\|_{\mathrm{osc}}}{\kappa} \right) \left\| \frac{\rho_{3t/4}(x, \cdot)}{\rho_\infty} - 1 \right\|_{L^2(\mu)}.$$

Here $C_1 = C_1(d)$ is the constant form Lemma 4.1.

The above implies that for any $t \geqslant 4nt_{\mathrm{dis}}$, we have

$$\|\rho_t(x, \cdot) - \rho_\infty\|_{L^1}$$

$$\leqslant C_1 \left(\frac{4}{\kappa t_{\mathrm{dis}}} \right)^{\frac{d}{2}} \exp\left(\frac{2d\|U\|_{\mathrm{osc}}}{\kappa} - n \right) \left\| \frac{\rho_{t/2}(x, \cdot)}{\rho_\infty} - 1 \right\|_{L^2(\mu)}$$

$$\leqslant C_1^2 \left(\frac{4}{\kappa t_{\mathrm{dis}}} \right)^{d} \exp\left(\frac{(4d\|U\|_{\mathrm{osc}})}{\kappa} - n \right) \left\| \frac{\rho_{t/4}(x, \cdot)}{\rho_\infty} - 1 \right\|_{L^1(\mu)}$$

$$\leqslant 2C_1^2 \left(\frac{4}{\kappa t_{\mathrm{dis}}} \right)^{d} \exp\left(\frac{4d\|U\|_{\mathrm{osc}}}{\kappa} - n \right).$$

Choosing

(4.7) $$n = \left\lceil \frac{4d\|U\|_{\mathrm{osc}}}{\kappa} + (2d + 2)\log 2 + 2\ln C_1 - d\ln(\kappa t_{\mathrm{dis}}) \right\rceil$$

yields

$$\|\rho_t(x, \cdot) - \rho_\infty\|_{L^1} \leqslant \frac{1}{2}, \quad \text{at } t = 4nt_{\mathrm{dis}}.$$

Using (4.5) we note that n in (4.7) can be bounded by

$$n \leqslant Cd\left(1 + \frac{\|U\|_{\mathrm{osc}}}{\kappa} - \ln(\kappa t_{\mathrm{dis}}) \right),$$

for some dimension independent constant C, and this implies the upper bound in (2.10) as desired. $\qquad\square$

It remains to prove Lemma 4.1, which we do in the next sub-section.

4.3. The $L^1 \to L^\infty$ bound on the transition density (Lemma 4.1).

To prove Lemma 4.1, we first compute an evolution equation for the ratio $\rho_t^{(x,s)}/\rho_\infty$. Recall that in the variables y, t, the transition density ρ is a solution to the forward equation

(4.8) $$\partial_t \rho = -\nabla \cdot_y (Av'\rho) + \mathcal{L}_{\kappa,y}^* \rho,$$

where v' is the time rescaled velocity field (3.2), $\mathcal{L}_\kappa^*$ is defined by

(4.9) $$\mathcal{L}_\kappa^* f = \nabla \cdot (f\nabla U) + \kappa\Delta f.$$

We clarify that and the notation $\mathcal{L}_{\kappa,y}$ refers to the fact that $\mathcal{L}_\kappa$ uses derivatives with respect to variable y in (4.8). While the operator $\mathcal{L}_\kappa$ is self adjoint with respect to the $L^2(\mu)$ inner-product, it is not self-adjoint with respect to the standard L^2 inner-product. Indeed, the adjoint of $\mathcal{L}_\kappa$ with respect to the standard L^2 inner-product is precisely $\mathcal{L}_\kappa^*$.

Lemma 4.2. *Let φ be a solution of the forward equation*

$$(4.10) \qquad \partial_t \varphi = -\nabla \cdot (Av'\varphi) + \mathcal{L}_\kappa^* \varphi \,,$$

where we recall $\mathcal{L}_\kappa^*$ is defined in equation (4.9). The function θ, defined by

$$\theta_t(x) \stackrel{\text{def}}{=} \frac{\varphi_t(x)}{\rho_\infty(x)} \,,$$

is a solution of the equation

$$(4.11) \qquad \partial_t \theta + Av' \cdot \nabla \theta - \mathcal{L}_\kappa \theta = 0 \,.$$

Remark. The equation (4.11) differs from the backward equation (3.1) only in the sign of the convection term $Av' \cdot \nabla \theta$.

Proof. The proof is a direct calculation. Substituting $\varphi = \rho_\infty \theta$ in (4.10) yields

$$\rho_\infty \partial_t \theta = \theta \Big(\mathcal{L}_\kappa^* \rho_\infty - \nabla \cdot (Av'\rho_\infty) \Big)$$
$$+ \rho_\infty \Big((\nabla U + Av') \cdot \nabla \theta + \kappa \Delta \theta \Big) + 2\kappa \nabla \rho_\infty \cdot \nabla \theta \,.$$

Using (1.5), and the fact that

$$\mathcal{L}_\kappa^* \rho_\infty = 0 \,, \qquad \kappa \nabla \rho_\infty = -\rho_\infty \nabla U \,,$$

we obtain

$$\rho_\infty \partial_t \theta = \rho_\infty \big(-Av' \cdot \nabla \theta + \mathcal{L}_\kappa \theta \big) \,.$$

Since $\rho_\infty > 0$, this implies (4.11) concluding the proof. $\qquad\square$

The next lemma we need is an $L^1 \to L^\infty$ bound on the semigroup operator of (3.1). This is the analog of the results in [CKRZ08, Zla10, IXZ21, FKR06] when the underlying measure is μ and not the Lebesgue measure.

Lemma 4.3. *When $d \geqslant 3$, every solution to (4.11) with μ-mean zero initial data satisfies*

$$(4.12) \qquad \|\theta_{s+t}\|_{L^\infty} \leqslant \frac{C_1(d)}{(\kappa t)^{d/2}} \exp\Big(\frac{2d\|U\|_{\text{osc}}}{\kappa} \Big) \|\theta_s\|_{L^1(\mu)} \,.$$

where $C_1(d)$ is as defined in (4.5).
When $d = 2$, the inequality (4.12) needs to be replaced by

$$(4.13) \qquad \|\theta_{s+t}\|_{L^\infty} \leqslant \frac{C_1'(\varepsilon)}{(\kappa t)^{d/2+\varepsilon}} \exp\Big(\frac{(2d+4\varepsilon)\|U\|_{\text{osc}}}{\kappa} \Big) \|\theta_s\|_{L^1(\mu)} \,,$$

where $\varepsilon > 0$ and $C_1'(\varepsilon)$ is an ε-dependent constant.

Of course Lemma 4.2 and 4.3 immediately imply Lemma 4.1.

Proof of Lemma 4.1. For any fixed $x \in \mathbb{T}^d$, $s' \geqslant 0$, we know that the transition density $\rho(x, s'; y, t)$ satisfies the forward equation (4.10) in the variables y, t. Thus, by Lemma 4.2, the function θ defined by

$$\theta_t(y) \stackrel{\text{def}}{=} \frac{\rho(x, s'; y, t)}{\rho_\infty(y)} - 1\,,$$

satisfies equation (4.11). Clearly θ has μ-mean zero. Also, for any for any $s > s'$, $\theta_s \in L^1(\mu)$, and so Lemma 4.3 applies. The bounds (4.4) and (4.6) follow immediately from (4.12) and (4.13) respectively. $\qquad\square$

It remains to prove Lemma 4.3. For this we will need a Nash inequality with respect to the measure μ.

Lemma 4.4 (Nash Inequality)**.** *For $d \geqslant 3$ and any μ-mean zero function f we have*

$$(4.14) \qquad\qquad C_2 \|\nabla f\|^2_{L^2(\mu)} \geqslant \frac{\|f\|^{2+\frac{4}{d}}_{L^2(\mu)}}{\|f\|^{\frac{4}{d}}_{L^1(\mu)}}\,,$$

where $C_2 = C_2(d, U, \kappa)$ is a dimensional constant that can be bounded by

$$C_2 \leqslant 2^{2+\frac{8}{d}} C_d^2 \exp\left(\frac{4\|U\|_{\text{osc}}}{\kappa}\right)$$

where

$$(4.15) \qquad\qquad C_d \stackrel{\text{def}}{=} \frac{1}{\sqrt{\pi d(d-2)}}\left(\frac{\Gamma(d)}{\Gamma(\frac{d}{2})}\right)^{1/d}.$$

When $d = 2$, the inequality (4.14) needs to be replaced by

$$(4.16) \qquad\qquad C_2' \|\nabla f\|^2_{L^2(\mu)} \geqslant \frac{\|f\|^{4-\delta}_{L^2(\mu)}}{\|f\|^{2-\delta}_{L^1(\mu)}}\,,$$

where $\delta \in (0, 2)$ is arbitrary, and $C_2' = C_2'(\delta, U, \kappa)$ can be bounded by

$$C_2' \leqslant C_2''(\delta) \exp\left(\frac{4\|U\|_{\text{osc}}}{\kappa}\right),$$

for some δ-dependent constant C_2''.

Momentarily postponing the proof of Lemma 4.4, we finish the proof of Lemma 4.3.

Proof of Lemma 4.3. Multiplying equation (4.11) by $\rho_\infty \theta$ and integrating gives

$$(4.17) \qquad\qquad \partial_t \|\theta\|^2_{L^2(\mu)} = -2\kappa \|\nabla\theta\|^2_{L^2(\mu)} \leqslant -\frac{2\kappa}{C_2} \frac{\|\theta\|^{\frac{2(d+2)}{d}}_{L^2(\mu)}}{\|\theta\|^{\frac{4}{d}}_{L^1(\mu)}}.$$

We claim $\|\theta_t\|_{L^1(\mu)} \leqslant \|\theta_0\|_{L^1(\mu)}$. To see this, let θ_t^+ and θ_t^- be solutions of (4.11) with initial data $\max\{\theta_0, 0\}$ and $-\min\{\theta_0, 0\}$ respectively. By the comparison principle, we know $\theta_t^\pm \geqslant 0$, and by linearity $\theta_t = \theta_{t,+} - \theta_{t,-}$. This implies

$$\|\theta_t\|_{L^1(\mu)} \leqslant \int_{\mathbb{T}^d} (\theta_t^+ + \theta_t^-)\,d\mu = \int_{\mathbb{T}^d} (\theta_0^+ + \theta_0^-)\,d\mu = \|\theta_0\|_{L^1(\mu)}\,.$$

Using this in (4.17) yields

$$\partial_t \|\theta_t\|^2_{L^2(\mu)} \leqslant -\frac{2\kappa}{C_2} \frac{\|\theta_t\|^{\frac{2(d+2)}{d}}_{L^2(\mu)}}{\|\theta_0\|^{\frac{4}{d}}_{L^1(\mu)}} \,.$$

This is a differential inequality for $\|\theta_t\|^2_{L^2(\mu)}$, which can be solved to give

$$(4.18) \qquad \|\theta_t\|_{L^2(\mu)} \leqslant \frac{(d\,C_2)^{d/4}}{(4\kappa t)^{d/4}} \|\theta_0\|_{L^1(\mu)} \,.$$

Now let $\mathcal{P}_{s,t}(v')$ denote the solution operator to (4.11) (i.e. the function ϑ_t defined by $\vartheta_t \stackrel{\text{def}}{=} \mathcal{P}_{s,t}(v')(f)$ solves (4.11) with initial data $\vartheta_s = f$). From (4.18), we see

$$\|\mathcal{P}_{s,s+t}(v')\|_{\dot{L}^1(\mu)\to\dot{L}^2(\mu)} \leqslant \frac{(d\,C_2)^{d/4}}{(4\kappa t)^{d/4}}$$

Moreover, since v' satisfies (1.5) we see that

$$(\mathcal{P}_{s,t}(v'))^* = \mathcal{P}_{s,t}(-v') \,,$$

where $\mathcal{P}_t(v')^*$ denotes the adjoint of $\mathcal{P}_t v'$ with respect to the $L^2(\mu)$ inner-product. Consequently,

$$\begin{aligned}
\|\mathcal{P}_{s,s+2t}(v')\|_{\dot{L}^1\to\dot{L}^\infty} &\leqslant \|\mathcal{P}_{s+t,s+2t}(v')\|_{\dot{L}^1\to\dot{L}^2}\|\mathcal{P}_{s,s+t}(v')\|_{\dot{L}^2\to\dot{L}^\infty} \\
&= \|\mathcal{P}_{s+t,s+2t}(v')\|_{\dot{L}^1\to\dot{L}^2}\|(\mathcal{P}_{s,s+t}(v'))^*\|_{\dot{L}^1\to\dot{L}^2} \\
&\leqslant \frac{(d\,C_2)^{d/2}}{(4\kappa t)^{d/2}} \,.
\end{aligned}$$

This in turn implies

$$\|\theta(t)\|_{L^\infty} \leqslant \frac{(d\,C_2)^{d/2}}{(\kappa t)^{d/2}} \|\theta_0\|_{L^1} \,,$$

Recalling the definition of C_2, we actually have

$$(d\,C_2)^{d/2} \leqslant C 2^d \exp\Big(\frac{2d\|U\|_{\text{osc}}}{\kappa}\Big) \,,$$

where C is some universal constant independent of d. which concludes the proof when $d \geqslant 3$.

The proof when $d = 2$ is similar and only involves using (4.16) instead of (4.14). $\qquad\square$

4.4. **The Nash and Poincaré inequalities.** We conclude this section by proving the Nash (Lemma 4.4) and Poincaré inequalities.

When $d \geqslant 3$, recall the standard Nash inequality states

$$(4.19) \qquad C_d^2 \|\nabla f\|^2_{L^2} \geqslant \frac{\|f - f_0\|^{\frac{2(d+2)}{d}}_{L^2}}{\|f - f_0\|^{\frac{4}{d}}_{L^1}} \,,$$

where C_d ia as defined in (4.15), and $f_0 = \int_{\mathbb{T}^d} f\,dx$.

The Nash inequality above can be deduced from the Sobolev inequality and interpolation. Indeed, the Sobolev inequality (see for instance [Lie83]), implies

$$(4.20) \qquad C_d \|\nabla f\|_{L^2} \geqslant \|f - f_0\|_{L^{2^*}} \,, \qquad \text{where} \qquad 2^* = \frac{2d}{d - 2} \,.$$

Since $d \geqslant 3$ note $2^* > 2$, and hence the interpolation inequality gives

$$\|f - f_0\|_{L^2} \leqslant \|f - f_0\|_{L^{2^*}}^{\frac{d}{d+2}} \|f - f_0\|_{L^1}^{\frac{2}{d+2}} .$$

Combined with (4.20) this implies (4.19) as claimed.

To prove the Nash inequality (4.14) with respect to the measure μ, we first need an elementary result controlling the $L^p(\mu)$ difference to the mean when the underlying measure μ is changed.

Lemma 4.5. *Let $\tilde{\rho}_\infty$ be a probability density function on $\mathbb{T}^d$, and let $\tilde{\mu}$ be the probability measure such that $d\tilde{\mu} = \tilde{\rho}_\infty \, dx$. Suppose there exists a constants B_1, B_2 such that*

$$\frac{1}{B_1} \tilde{\rho}_\infty(x) \leqslant \rho_\infty(x) \leqslant B_2 \tilde{\rho}_\infty(x) \quad \text{for all } x \in \mathbb{T}^d .$$

Then, for any $p \in [1, \infty)$, and any $f \in L^p(\mu)$, we have

$$\frac{1}{2B_1^{1/p}} \|f - \tilde{f}\|_{L^p(\tilde{\mu})} \leqslant \|f - \bar{f}\|_{L^p(\mu)} \leqslant 2B_2^{1/p} \|f - \tilde{f}\|_{L^p(\tilde{\mu})} .$$

Here

$$\bar{f} = \int_{\mathbb{T}^d} f \, d\mu \qquad \text{and} \qquad \tilde{f} = \int_{\mathbb{T}^d} f \, d\tilde{\mu} ,$$

are the means of f with respect to the measures μ and $\tilde{\mu}$ respectively.

Proof. From the triangle inequality, we note

$$\|f - \bar{f}\|_{L^p(\mu)} \leqslant \|f - \tilde{f}\|_{L^p(\mu)} + |\bar{f} - \tilde{f}| \leqslant B_2^{1/p} \|f - \tilde{f}\|_{L^p(\tilde{\mu})} + \left| \int_{\mathbb{T}^d} (f - \tilde{f}) \, d\mu \right|$$

$$\leqslant B_2^{1/p} \|f - \tilde{f}\|_{L^p(\tilde{\mu})} + \|f - \tilde{f}\|_{L^1(\mu)} \leqslant 2B_2^{1/p} \|f - \tilde{f}\|_{L^p(\tilde{\mu})} .$$

The proof of the lower bound is similar. $\qquad\square$

We now prove Lemma 4.4.

Proof of Lemma 4.4. To prove (4.14) we note Lemma 4.5 implies

$$C_d^2 \|\nabla f\|_{L^2(\mu)}^2 \geqslant C_d^2 \min(\rho_\infty) \|\nabla f\|_{L^2}^2 \geqslant \frac{\|f - \bar{f}\|_{L^2(\mu)}^{\frac{2(d+2)}{d}}}{2^{2+\frac{8}{d}} B \|f - \bar{f}\|_{L^1(\mu)}^{\frac{4}{d}}} ,$$

where $\bar{f} = \int_{\mathbb{T}^d} f \, d\mu$ and

$$B = B(U, \kappa) = \frac{\max(\rho_\infty)^{1+\frac{2}{d}}}{\min(\rho_\infty)^{1+\frac{4}{d}}} \leqslant \left(\frac{\max(\rho_\infty)}{\min(\rho_\infty)} \right)^{1+\frac{4}{d}} \leqslant \exp\left(\left(1 + \frac{4}{d}\right) \frac{\|U\|_{\text{osc}}}{\kappa} \right) .$$

This finishes the proof when $d \geqslant 3$.

When $d = 2$, the Sobolev inequality in (4.20) becomes

$$C' \|\nabla f\|_{L^2} \geqslant \|f - f_0\|_{L^p} ,$$

for any $2 < p < \infty$. And the interpolation inequality gives

$$\|f - f_0\|_{L^2} \leqslant \|f - f_0\|_{L^p}^{\frac{p}{2p-2}} \|f - f_0\|_{L^1}^{\frac{p-2}{2p-2}} ,$$

which further yields

$$C'^2 \|\nabla f\|_{L^2}^2 \geqslant \frac{\|f - f_0\|_{L^2}^{4 - \frac{4}{p}}}{\|f - f_0\|_{L^1}^{2 - \frac{4}{p}}} .$$

We then apply Lemma 4.5 and get the desired result. $\qquad\square$

Finally, for completeness, we conclude this section by stating the Poincaré inequality for the measure μ. We do not use this in the proof of our main result, but only use it in Remark 2.5 to comment that right hand side of (2.10) is nonnegative.

Lemma 4.6 (Poincaré Inequality). *Let λ_0 be the smallest eigenvalue of $-\mathcal{L}_\kappa$ on $\dot{L}^2(\mu)$. Then λ_0 is bounded below by*

$$(4.21) \qquad \lambda_0 \geqslant 2\pi \exp\Big(\frac{-\|U\|_{\mathrm{osc}}}{2\kappa}\Big).$$

Moreover, for all $f \in H^1 \cap \dot{L}^2(\mu)$, we have

$$(4.22) \qquad \lambda_0\|f\|_{L^2(\mu)}^2 \leqslant \|\nabla f\|_{L^2(\mu)}^2.$$

Proof. Using (3.3) and standard spectral theory, we know

$$\lambda_0 = \inf_{f \in \dot{L}^2(\mu)-0} \frac{\|\nabla f\|_{L^2(\mu)}}{\|f\|_{L^2(\mu)}},$$

which immediately implies (4.22). To obtain (4.21), let L_{nc}^2 denote the set of all non-constant L^2 functions, and note that the above implies

$$\lambda_0 = \inf_{f \in L_{\mathrm{nc}}^2(\mu)} \sup_{c \in \mathbb{R}} \frac{\|\nabla f\|_{L^2(\mu)}}{\|f-c\|_{L^2(\mu)}} \geqslant \exp\Big(\frac{-\|U\|_{\mathrm{osc}}}{2\kappa}\Big) \inf_{f \in L_{\mathrm{nc}}^2} \sup_{c \in \mathbb{R}} \frac{\|\nabla f\|_{L^2}}{\|f-c\|_{L^2}}$$

$$= \exp\Big(\frac{-\|U\|_{\mathrm{osc}}}{2\kappa}\Big) \inf_{f \in \dot{L}^2-0} \frac{\|\nabla f\|_{L^2}}{\|f\|_{L^2}} = 2\pi \exp\Big(\frac{-\|U\|_{\mathrm{osc}}}{2\kappa}\Big). \qquad\square$$

5. **Explicit asymptotics in discrete time**

In this this section we consider a discrete time version of (1.4). Namely, we will run equation (1.3) (without the drift) for time $1/A$; and then we will run the flow Av' (without noise) for time $1/A$. Running the flow (without noise) corresponds to applying the μ-measure preserving diffeomorphism $\Phi_{s,s+1}$, defined in (2.1). If instead of applying $\Phi_{s,s+1}$ (the flow map of a velocity field), we apply an arbitrary μ-measure preserving diffeomorphism, then we provide an example which is exponentially mixing with rate (1.6), where the behavior of both D and γ is known as $\kappa \to 0$. This is what leads to the heuristics (1.15) described in Section 1.

Explicitly, suppose $\Phi \colon \mathbb{T}^d \to \mathbb{T}^d$ is a μ-measure preserving diffeomorphism, and $\rho_t^X(x,y)$ is the transition density of the solution to (1.3). Let Y be the Markov process with transition probability

$$\boldsymbol{P}(Y_{n+1} \in dy \mid Y_n = x) = \rho_{1/A}^X(x, \Phi^{-1}(y)).$$

Alternately, one can (equivalently) define Y_{n+1} by letting Z be the solution of (1.3) with initial data $Z_0 = Y_n$, and then defining

$$Y_{n+1} = \Phi(Z_{1/A}).$$

With this notation Y_n above serves as a proxy for X_t (the solution to (1.4)) where n and t are related through

$$n = At.$$

The notions of mixing, dissipation time, etc. in discrete time are defined analogously to those in the continuous-time setting. To differentiate from the continuous time versions, in the discrete-time setting we will use N_{dis} and N_{mix} to denote the

dissipation and mixing times respectively. The main results in this section are the following.

Proposition 5.1. *(1) There exists a μ-measure preserving, exponentially mixing diffeomorphism Φ whose mixing rate is*

$$h(n) = De^{-\gamma n}, \tag{5.1}$$

where where $D = D(d, \kappa)$, but γ is independent of both κ and d.
(2) If further U is in the form

$$U(x) = \sum_{i=1}^{d} \tilde{U}(x_i) \quad \text{for some} \quad \tilde{U} \in C^2(\mathbb{T}), \tag{5.2}$$

then D can be bounded by

$$D \leqslant de^{O(1/\kappa)} \quad \text{as } \kappa \to 0. \tag{5.3}$$

Proposition 5.2. *If Φ is mixing with rate h, then the mixing time and dissipation time of Y are bounded above by*

$$N_{\text{mix}} \leqslant Cd\Big(1 + \frac{\|U\|_{\text{osc}}}{\kappa} - \ln\Big(\frac{\kappa N_{\text{dis}}}{A}\Big)\Big)N_{\text{dis}}, \qquad N_{\text{dis}} \leqslant \frac{CA}{H(A)}.$$

Here

$$H(A) \stackrel{\text{def}}{=} \sup\Big\{\lambda \,\Big|\, h\Big(\frac{\sqrt{A}}{2\sqrt{\lambda}}\Big) \leqslant \frac{\kappa}{2\lambda}\Big\}.$$

The proof of Proposition 5.2 follows the same method as Proposition 2.1, with the continuous-time energy decay replaced with the time discrete analogs (see Lemmas 3.1 and 3.2 in [FI19]). For brevity, we do not present it here. We will prove Proposition 5.1 below.

The main idea behind the proof of Proposition 5.1 is to construct μ-exponentially mixing diffeomorphisms as conjugates of Lebesgue exponentially mixing diffeomorphisms. There are many examples of Lebesgue exponentially mixing diffeomorphisms, such as the baker's map or toral automorphisms [KH95, SOW06]. In the time inhomogeneous case, they can also be constructed as flow maps of alternating shear flows as in Section A or [BCZG22]. To prove Proposition 5.1, however, we will need a Lebesgue exponentially mixing map with mixing rate that is independent of the dimension. While many of the examples mentioned above likely have a mixing rate that can be bounded independent of the dimension, it is easiest to prove this for an explicit toral automorphism. Once this has been established, a direct calculation will show that the pre-factor D in (5.1) may depend on κ, but the exponential rate γ does not.

Lemma 5.3. *There exists a diffeomorphism $\Psi: \mathbb{T}^d \to \mathbb{T}^d$ which is Lebesgue exponentially mixing with a rate that is independent of the dimension.*

Proof. We will choose Ψ to be a toral automorphism. Recall, given any $A \in SL(d, \mathbb{Z})$, a *toral automorphism* with matrix A is the map $\Psi: \mathbb{T}^d \to \mathbb{T}^d$ defined by

$$\Psi(x) = Ax \pmod{\mathbb{Z}^d}.$$

The mixing properties of these maps are well known (see for instance [Lin82, FW03, FI19]). In particular, if all eigenvalues of A are irrational and at least one lies

outside the unit disk, then Ψ is Lebesgue exponentially mixing. To prove this, note that a Fourier series expansion immediately shows

$$(5.4) \qquad \langle f \circ \Psi^n, g \rangle \leqslant \sup_{k \in \mathbb{Z}^d - 0} \frac{1}{|A^n k||k|^2} \|f\|_{\dot{H}^1} \|g\|_{\dot{H}^2} ,$$

for all test functions $f \in \dot{H}^1$, $g \in \dot{H}^2$ (see for instance equation (4.7) in [FI19]). Now using Diophantine approximation results one can show

$$(5.5) \qquad |A^n k||k^{d-1}| \geqslant \frac{|\lambda_1|^n}{C_d} ,$$

where λ_1 is an eigenvalue of A with the largest modulus, and C_d is a dimensional constant (see for instance the inequality immediately after (4.8) in [FI19]). This can be used to show Ψ is Lebesgue exponentially mixing, however, constants appearing in the mixing rate will depend on the dimension.

We will now choose A in a specific form that will ensure the mixing rate is independent of the dimension. Let

$$A_1 = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \qquad \text{and} \qquad A_2 = \begin{pmatrix} 2 & -1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}$$

and choose $A \in SL(d, \mathbb{Z})$ to be any block diagonal matrix with only 2×2 blocks A_2, or 3×3 blocks A_3 on the diagonal. One can directly check that both A_2 and A_3 are ergodic toral automorphisms. Since the domain of A_m is $\mathbb{T}^m$, for $m \in \{2,3\}$, the lower bound (5.5) becomes

$$|A_m^n k||k|^2 \geqslant |A_m^n k||k|^{m-1} \geqslant \frac{\lambda^n}{C} , \quad \text{for all } m \in \{2,3\}, \ k \in \mathbb{Z}^m - 0 ,$$

where

$$\lambda = \min_{m \in \{2,3\}} \max\{|\mu| \mid \mu \text{ is an eigenvalue of } A_m\} > 1 ,$$

and C only depends on A_2, A_3 (and hence is independent of d).

Now for $k \in \mathbb{Z}^d$, write $k = (k_1, \ldots, k_{d'})$ where each $k_i \in \mathbb{Z}^{m_i}$, $m_i \in \{2,3\}$, corresponds to the block diagonal structure of A. If $k \neq 0$, at least one k_i must be non-zero, and hence

$$|A^n k||k|^2 \geqslant |A_{m_i}^n k_i||k_i|^2 \geqslant \frac{\lambda^n}{C} .$$

Combined with (5.4) this immediately implies

$$\|f \circ \Psi^n\|_{H^{-2}} \leqslant C\lambda^{-n} \|f\|_{\dot{H}^1} .$$

By Hölder's inequality

$$\langle f \circ \Psi, g \rangle \leqslant \|f \circ \Psi\|_{H^{-2}}^{1/2} \|f \circ \Psi\|_{L^2}^{1/2} \|g\|_{\dot{H}^1} \leqslant \sqrt{C}\lambda^{-n/2} \|f\|_{\dot{H}^1}^{1/2} \|f\|_{L^2} \|g\|_{\dot{H}^1}$$

$$\leqslant \sqrt{C}\lambda^{-n/2} \|f\|_{\dot{H}^1} \|g\|_{\dot{H}^1} ,$$

showing Ψ is Lebesgue exponentially mixing with rate $\sqrt{C}\lambda^{-n/2}$. Since C and $\lambda > 1$ are independent of d, this concludes the proof. $\qquad \square$

Proof of Proposition 5.1. Let $\Psi \colon \mathbb{T}^d \to \mathbb{T}^d$ be the Lebesgue exponentially mixing diffeomorphism from Lemma 5.3. Notice that this is completely independent of κ, and neither D', γ, nor Ψ depend on κ.

Let $\psi\colon \mathbb{T}^d \to \mathbb{T}^d$ be a diffeomorphism such that the push forward of the measure μ under ψ is the Lebesgue measure (i.e. for all Borel sets A we have $\mu(\psi^{-1}(A)) = m(A)$, where m is the Lebesgue measure). One can, for instance, prove the existence of such a map using optimal transport. We claim that

$$\Phi = \psi^{-1}\Psi\psi \, ,$$

is a μ-exponentially mixing map with exponential rate γ. To see this, we note first that clearly Φ preserves the measure μ. Moreover, for any pair of test functions $f, g \in \dot{H}^1(\mu)$, we have

$$\langle f \circ \Phi^n, g \rangle_\mu = \langle f \circ (\psi^{-1}\Psi^n\psi), g \rangle_\mu = \langle (f \circ \psi^{-1})\Psi^n, g \circ \psi^{-1} \rangle \, .$$

Since f, g have μ-mean zero, the functions $f \circ \psi^{-1}$ and $g \circ \psi^{-1}$ must be Lebesgue mean-zero. Since Ψ is Lebesgue exponentially mixing, this implies

$$\langle (f \circ \psi^{-1})\Psi^n, g \circ \psi^{-1} \rangle \leqslant D' e^{-\gamma n} \|f \circ \psi^{-1}\|_{\dot{H}^1} \|g \circ \psi^{-1}\|_{\dot{H}^1}$$
$$\leqslant D' \|\nabla \psi^{-1}\|_{L^\infty}^2 e^{-\gamma n} \|f\|_{\dot{H}^1(\mu)} \|g\|_{\dot{H}^1(\mu)} \, ,$$

where we clarify that

$$\|\nabla \psi^{-1}\|_{L^\infty} = \left\| \sum_{i,j} |\partial_i \psi_j^{-1}|^2 \right\|_{L^\infty}^{1/2} \, .$$

Hence Φ is μ-exponentially mixing with rate

$$(5.6) \qquad\qquad h(n) = D' \|\nabla \psi^{-1}\|_{L^\infty}^2 e^{-\gamma n} \, . \qquad\qquad \square$$

This proves the first assertion of Proposition 5.1.

To prove the second assertion, note from (5.6), that the pre-factor D is bounded by

$$D \leqslant D' \|\nabla \psi^{-1}\|_{L^\infty} \, ,$$

where ψ is any diffeomorphism that pushes forward μ onto the Lebesgue measure. When $d = 1$ such maps are characterized by

$$\partial_x \psi^{-1} = \frac{e^{-U/\kappa}}{Z} \, .$$

When $d > 1$ and U is in the form (5.2), we can construct ψ using the one dimensional maps described above. Explicitly, define $\tilde\varphi\colon \mathbb{R} \to \mathbb{R}$ by

$$\tilde\varphi(x) = \frac{1}{\tilde Z} \int_0^x e^{-\tilde U(y)/\kappa} \, dy \, , \quad \text{where} \quad \tilde Z \stackrel{\text{def}}{=} \int_0^1 e^{-U(y)/\kappa} \, dy \, .$$

Since U is 1-periodic we note $\tilde\varphi(x+1) = 1 + \tilde\varphi(x)$, and hence $\tilde\varphi$ can be viewed as a map on the one dimensional torus. We now define $\psi^{-1}\colon \mathbb{T}^d \to \mathbb{T}^d$ by

$$\psi^{-1}(x) = (\tilde\varphi(x_1), \tilde\varphi(x_2), \ldots, \tilde\varphi(x_d)) \, .$$

Clearly the push forward of μ under ψ is the Lebesgue measure, and hence

$$D \leqslant D' \|\nabla \psi^{-1}\|_{L^\infty} \leqslant dD' \sup_{x \in [0,1]} \exp\left(\frac{1}{\kappa} \int_0^1 (\tilde U(y) - \tilde U(x)) \, dy \right) \, .$$

This proves (5.3), concluding the proof.

6. **Proof of exponential mixing of sawtooth shears (Theorem 1.3).**

The objective of this section is to show that the modified shears in (1.13) are exponentially mixing with probability 1, as stated in Theorem 1.3. The proof involves the analysis of geometric ergodicity of a pair of trajectories of the velocity field v. This study was initiated in [BS88] and further developed in [DKK04, Theorem 4], [BBPS22, Theorem 1.3], and [BCZG22, Theorem 1.1]. Among these results our proof is closest to Theorem 1.3 in [BBPS22] and differs from Theorem 1.3 in [BBPS22] only in one aspect. The proof in [BBPS22] uses Hörmander's condition to obtain irreducibility and a positive Lyapunov exponent of underlying Markov processes. We cannot use Hörmander's condition in our context. Instead, we use the Rashevsky–Chow Theorem [Sac22, Theroem 5] (see Theorem 6.5, below) to obtain the same results.

6.1. **Modified Sawtooth Shears in Two Dimensions.** We will first prove that the modified, randomly shifted, sawtooth shears are almost surely exponentially mixing in two dimensions. Following this we will prove the remaining conclusions of Theorem 1.3.

We begin by writing down the function F with the sawtooth shaped derivative (shown in Figure 1). Define

$$(6.1) \qquad F(x) = F_0(x) \stackrel{\text{def}}{=} \begin{cases} 2x^2 & x \in [0, \frac{1}{4}] \\ -2(x - \frac{1}{4})(x - \frac{3}{4}) + \frac{1}{8} & x \in [\frac{1}{4}, \frac{3}{4}] \\ 2(x-1)^2 & x \in [\frac{3}{4}, 1], \end{cases}$$

and extended periodically to $x \in \mathbb{R}$ (see Figure 1, right). For $\alpha \in [0, 1]$ we define F_α by

$$(6.2) \qquad F_\alpha(x) \stackrel{\text{def}}{=} F_0(x - \alpha).$$

Given F_α, we define the associated velocity fields v_α using (1.11) by replacing F with F_α. Explicitly, we define

$$(6.3) \qquad v_{\alpha,1} = \frac{1}{p}\nabla^\perp(pF_\alpha(x_1)) = \frac{1}{\kappa}\begin{pmatrix} U_{x_2}F_\alpha(x_1) \\ \kappa F'_\alpha(x_1) - U_{x_1}F_\alpha(x_1) \end{pmatrix},$$

$$(6.4) \qquad v_{\alpha,2} = \frac{1}{p}\nabla^\perp(pF_\alpha(x_2)) = \frac{1}{\kappa}\begin{pmatrix} U_{x_2}F_\alpha(x_2) - \kappa F'_\alpha(x_2) \\ -U_{x_1}F_\alpha(x_2) \end{pmatrix},$$

where $\nabla^\perp = \nabla^\perp_{1,2}$ is the skew gradient in two dimensions. Stream plots of these flow are shown in Figure 2.

For notational convenience, define $V_n = v_n$, where v defined by (1.13). Note

$$(6.5) \qquad V_n = \beta_n v_{\alpha_n, i_n},$$

where (α_n, β_n, i_n) are i.i.d. random variables that are uniformly distributed on the parameter space $[0, 1] \times [0, 1] \times \{1, 2\}$. We will show that the paths of the random flow obtained by composing the time 1 flows of each of the vector fields $V_1, \ldots, V_n$ are almost surely exponentially mixing.

6.2. **Conditions Guaranteeing Exponential Mixing.** To prove that the paths of the random flow above is exponentially mixing, it will be convenient to use Theorem 1.4 from [BCZG22]. For clarity of presentation we introduce the required preliminaries and restate this result below.

Consider the Markov process X defined by

$$(6.6) \qquad X_{n+1} = \varphi^{V_{n+1}}(X_n).$$

Here the notation φ^v denotes time 1 flow map of the vector field v, and the vector fields V_n were defined in (6.5). Define the random flow ϕ_n by

$$(6.7) \qquad \phi_n \overset{\text{def}}{=} \varphi^{V_n} \circ \varphi^{V_{n-1}} \circ \cdots \varphi^{V_1}.$$

Let $\mathbb{P}^1$ be the projective space of $\mathbb{R}^2$, that is the collection of lines in $\mathbb{R}^2$ that pass through the origin. For initial condition $(x, u) \in \mathbb{T}^2 \times \mathbb{P}^1$ the projective process on $\mathbb{T}^2 \times \mathbb{P}^1$ is defined by

$$(6.8) \qquad (X_n, U_n) = (\phi_n(x), D_x \phi_n u).$$

Given an initial condition $(x, g) \in \mathbb{T}^2 \times SL_2(\mathbb{R})$ the rescaled derivative process on $\mathbb{T}^2 \times SL_2(\mathbb{R})$ is defined by

$$(6.9) \qquad (X_n, A_n) = \left(\phi_n(x), \frac{D_x \phi_n g}{(\det D_x \phi_n)^{1/2}} \right).$$

For initial condition $(x, y) \in \mathcal{D}^c \overset{\text{def}}{=} \mathbb{T}^2 \times \mathbb{T}^2 - \{(x, x)\}$ the two point process (X_n, Y_n) on $\mathcal{D}^c$ is defined by

$$(6.10) \qquad (X_n, Y_n) = (\phi_n(x), \phi_n(y)).$$

Remark 6.1. Since each velocity field is Lipschitz, each flow is Lipschitz by Grönwall's inequality. That is,

$$(6.11) \qquad |\varphi_t^v(x) - \varphi_t^v(y)| \leqslant |x - y| e^{t C_v}$$

when x and y are sufficiently close. Thus, each flow is differentiable almost everywhere and the processes, (6.8) and (6.9), are well defined.

Theorem 1.4 in [BCZG22] can now be stated as follows.

Proposition 6.2 (Theorem 1.4 in [BCZG22]). *Assume that the single point process* (6.6) *and the projective process* (6.8) *are uniformly geometrically ergodic. Suppose the two point process* (6.10) *is $\mathcal{V}$-geometrically ergodic with respect to a Lyapunov function* (6.13) *given by*

$$(6.12) \qquad \mathcal{V}(x, y) = d(x, y)^{-p} \chi(x, y) \text{ for some small } p > 0,$$

and a continuous function $\chi(x, y)$ which is bounded both from above and away from 0. Further assume all three processes are aperiodic, and the single point process has a positive Lyapunov exponent. Let μ be the unique invariant measure of the single point process. If $\mathcal{V} \in L^1(\mathbb{T}^4, \mu \times \mu)$ then there exists a deterministic constant $\gamma > 0$ and a random constant $D(\omega)$ such that for all mean-zero functions $f, g \in H^1(\mathbb{T}^2, \mu)$ we have that

$$\langle f \circ \phi_n, g \rangle_\mu \leqslant D(\omega) e^{-\gamma n} \|f\|_{\dot{H}^1} \|g\|_{\dot{H}^1}$$

almost surely.

Recall a Lyapunov function for a Markov process with state space E is a function

$$(6.13) \qquad \mathcal{V} : E \to [1, \infty) \text{ such that } \boldsymbol{E}^x[\mathcal{V}(X_1)] \leqslant \lambda V(x) + C$$

for some constants $0 < \lambda < 1$, $C > 0$. If the process is aperiodic, irreducible and there exists a Lyapunov function $\mathcal{V}$ with compact sub-level sets then he process is

$\mathcal{V}$-geometrically ergodic [MT09, Theorem 15.0.1]. That is, there exists $\gamma > 0$ and a unique invariant distribution, μ, such that

$$(6.14) \qquad \|P^n(x, \cdot) - \mu\|_{TV} \leqslant \mathcal{V}(x)e^{-n\gamma} \text{ for all } x \in E.$$

If $\mathcal{V}$ in (6.14) is constant, the process is said to be uniformly geometrically ergodic. A Lyapunov exponent is the rate at which tangent vectors are stretched or shrunk under the iterates of a dynamical system. A general theory of Lyapunov exponents for random dynamical systems can be found in e.g. [Arn98, Chapter 3].

Remark 6.3. The state space of the two point process is not compact. Thus, showing geometric ergodicity is not immediate. We take the Lyapunov function of the two point process to be a specific perturbation of the principle eigenfunction of the projective process, see [BBPS22, Section 5]. In this case equation (6.13) states that two particles which are close together move away from each other on average. For more information on geometric ergodicity of Markov processes with non-compact state spaces see [MT09, Chapter 15].

Note that the processes (6.6), (6.8), and (6.10) are all aperiodic. Indeed, for all z in the processes' state space and every $\varepsilon > 0$, the flow of any vector field v with sufficiently small amplitude will stay inside an ε-ball centered at its initial position. This will ensure $\boldsymbol{P}(z, B(z, \varepsilon)) > 0$, showing that the processes (6.6), (6.8), and (6.10) are all aperiodic. Therefore to prove a sequence of randomly shifted modified sawtooth shears are mixing we must show the following.

(1) The processes (6.6) and (6.8) are uniformly geometrically ergodic.
(2) The existence of a positive Lyapunov exponent for (6.6).
(3) Existence of a Lyapunov function as in equation (6.12).
(4) The process (6.10) is $\mathcal{V}$-geometrically ergodic with respect to a function of the form (6.12).

To prove the first item it suffices to prove the processes, (6.6) and (6.8), are irreducible and Feller (see for example Theorem 15.0.1 in [MT09]). We prove irreducibility in Lemma 6.6 using the Rachevsky–Chow theorem, and prove the Feller condition in Lemma 6.11. The second item follows from irreducibility, and is shown in Lemma 6.10 below. The third item follows from [BBPS22] Section 5, since the projective process is irreducible and uniformly geometrically ergodic. Finally, the fourth item follows from section 6 in [BBPS22], and the fact that both the two point process (6.6) and the projective process (6.8) are irreducible and Feller.

Remark 6.4. In [BCZG22], the authors needed one more condition to show $\mathcal{V}$-geometric ergodicity. The condition was on the existence of an open so called small set [MT09, Chapter 5]. Theorem 5.5.7 and Proposition 6.2.8 in [MT09] imply that compact sets are small when the process is irreducible, aperiodic, and Feller. Thus, any small enough open ball is an open small set.

6.3. **Irreducibility.** In order to prove irreducibility, we will use a controllability result of Rachevsky [Ras38] and Chow [Cho40], which shows that if a collection of vector fields satisfies a Hörmander condition, then any two points are connected by a composition of flows.

Theorem 6.5 (Rashevsky–Chow (Theorem 5 in [Sac22]))**.** *Let M be a smooth connected manifold, and $\mathcal{F}$ a collection of vector fields on M. Suppose that for every $x \in M$, the Lie algebra $\mathrm{Lie}_x(\mathcal{F})$ spans the tangent space T_xM. Then for every $x, y \in$*

M there exists a sequence of times $t_1, t_2, \ldots t_n$ and vector fields $v_1, v_2, \ldots v_n \in \mathcal{F}$ with flows φ^i such that

$$\varphi_{t_n}^n(\varphi_{t_{n-1}}^{n-1}(\cdots(\varphi_{t_1}^1(x))\cdots)) = y.$$

Recall that Lie algebra $\mathrm{Lie}_x \mathcal{F}$ is defined by

$$\mathrm{Lie}_x \mathcal{F} = \mathrm{span}\left(\bigcup_n \{v(x) \mid v \in \mathcal{F}_n\}\right),$$

where $\mathcal{F}_0 = \{v\}_{v \in \mathcal{F}}$, and $\mathcal{F}_n$ is defined inductively by

$$\mathcal{F}_{n+1} \stackrel{\mathrm{def}}{=} \mathcal{F}_n \cup \{[v_1, v_2] \mid v_1 \in \mathcal{F}_n, \ v_2 \in \mathcal{F}\}.$$

Here the notation $[v_1, v_2]$ denotes the Lie bracket of two vector fields v_1, v_2 on a smooth manifold M. Recall the Lie bracket is the derivative of v_2 along the flow of v_1, and can be computed by

$$[v_1, v_2] = \mathrm{D}_{v_1} v_2 - \mathrm{D}_{v_2} v_1\,,$$

where D_v is the directional derivative in the direction of v.

We will now use Theorem 6.5 to prove irreducibility.

Lemma 6.6. *The single point process* (6.6), *the projective process* (6.8), *the rescaled derivative process* (6.9), *and the two point process* (6.10) *are all irreducible.*

Proof. To show the irreducibility of the single-point process we need to demonstrate that the Lie algebra generated by vector fields $v_{\alpha,1}$ and $v_{\alpha,2}$ (equations (6.3) and (6.4), respectively), is two-dimensional for every point $x \in \mathbb{T}^2$. It is straightforward to verify directly, and we do not explicitly do that. Instead, we conclude it from the irreducibility of the two-point process. The irreducibility of the projective process follows from that of the rescaled derivative process. Thus we only need to show irreducibility of the two-point process and the rescaled derivative process.

The proofs of irreducibility of the rescaled derivative process and the two-point process are similar, and we consider the two-point process first. We need to show that for a dense connected subset of $\mathcal{D}^c$, the corresponding Lie algebra generated by the vector fields

$$\{(v_{\alpha,i}^T, v_{\alpha,i}^T)^T\}$$

has dimension 4. We will choose this dense connected subset to be the set M defined by

$$(6.15) \qquad M \stackrel{\mathrm{def}}{=} \mathcal{D}^c - \{(x, y) \in \mathbb{T}^2 \times \mathbb{T}^2 \mid x_1 = y_1 \text{ or } x_2 = y_2\}.$$

Fix a pair $(x, y) \in M$, $x = (x_1, x_2)$, $y = (y_1, y_2)$. For our x_1 and y_1 we can choose $\alpha \in [0, 1]$ so that the stream function (6.2) satisfies

$$F_\alpha(x_1) = -2(x_1 - \alpha)\left(x_1 - \left(\alpha + \frac{1}{4}\right)\right) + \frac{1}{8}\,,$$
$$F_\alpha(y_1) = 2(y_2 - \alpha)^2\,,$$

respectively. This gives that $(v_{\alpha,1}^T(x), v_{\alpha,1}^T(y))$ is a quadratic polynomial in α. That is,

$$(6.16) \qquad \begin{pmatrix} v_{\alpha,1}(x) \\ v_{\alpha,1}(y) \end{pmatrix} = w_0(x, y) + a w_1(x, y) + a^2 w_2(x, y)\,,$$

for some vector fields w_0, w_1, w_2 on $\mathbb{T}^2 \times \mathbb{T}^2$. This implies that the span of $(v_{\alpha,1}(x), v_{\alpha,1}(y))^T$ is contained in

$$\text{Span}\{w_0(x,y),\ w_1(x,y),\ w_2(x,y)\}.$$

By taking $\tilde{\alpha} \equiv \alpha + 1/2$ (mod 1), we have that $(v_{\alpha,1}(x), v_{\alpha,1}(y))^T$ is also a quadratic polynomial in $\tilde{\alpha}$. Thus we write

$$\begin{pmatrix} v_{\alpha,1}(x) \\ v_{\alpha,1}(y) \end{pmatrix} = \tilde{w}_0(x,y) + \tilde{a}\tilde{w}_1(x,y) + \tilde{a}^2\tilde{w}_2(x,y),$$

for some vector fields $\tilde{w}_0$, $\tilde{w}_1$ and $\tilde{w}_2$ on $\mathbb{T}^2 \times \mathbb{T}^2$. Thus, the 6 vector fields

$$w_0(x,y),\ w_1(x,y)\ ,w_2(x,y),\ \tilde{w}_0(x,y)\ \tilde{w}_1(x,y),\ \tilde{w}_2(x,y)$$

are all contained in the Lie algebra at (x,y). Similarly, for our x_2 and y_2 we can choose another $\alpha \in [0,1]$, and obtain another 6 vector fields from $v_{\alpha,2}$. We can directly compute and check that the span of these 12 vector fields has dimension 4 if $x_1 \neq y_1$ and $x_2 \neq y_2$. Since the calculation is somewhat tedious to verify, we do this symbolically and the code verifying this can be found [CFIN23].

Also note that the set M (defined in (6.15)) is connected. Thus we can apply the Rachevsky-Chow Theorem 6.5 and conclude irreducibility of the two-point process.

We now show the rescaled derivative process is irreducible. The corresponding Lie algebra is in the tangent space at $(x, A) \in \mathbb{T}^2 \times SL_2\mathbb{R}$. Let $A_t(v,x)$ be the derivative matrix of the flow v at x and time t rescaled by its determinant, and a_1, ..., a_4 be the individual entries of this matrix. That is, define

$$(6.17) \qquad A_t(v,x) \overset{\text{def}}{=} \frac{D_x\varphi_t^v}{\sqrt{\det D_x\varphi_t^v}} = \begin{pmatrix} a_1 & a_2 \\ a_3 & a_4 \end{pmatrix}.$$

We consider the vector fields of the form

$$(6.18) \qquad \tilde{v}_{\alpha,1} = \begin{pmatrix} v_{\alpha,1} \\ \partial_t a_1 \\ \partial_t a_2 \\ \partial_t a_3 \\ \partial_t a_4 \end{pmatrix}, \quad \tilde{v}_{\alpha,2} = \begin{pmatrix} v_{\alpha,2} \\ \partial_t a_1 \\ \partial_t a_2 \\ \partial_t a_3 \\ \partial_t a_4 \end{pmatrix}.$$

That is, the vector fields have the first two coordinates from the vectors $v_{\alpha,1}$ or $v_{\alpha,2}$ and last four coordinates as the time derivatives of the corresponding matrix A_t.

We now show the collection of these vector fields satisfies the conditions of Theorem 6.5. For this, it suffices to show that for every $x \in \mathbb{T}^2$, the Lie algebra at (x, Id) has dimension 5. Since the rescaled derivative process is a matrix flow, given any other $g \in SL_2(\mathbb{R})$, the Lie algebra also has dimension 5.

Since each v is piecewise-smooth, we compute the Lie bracket on each smooth piece. By selecting α_1 and α_2 such that

$$F_{\alpha_1}(x_1) = -2(x_1 - \alpha_1)\left(x_1 - \left(\alpha_1 + \frac{1}{4}\right)\right) + \frac{1}{8},$$

$$F_{\alpha_2}(x_2) = -2(x_2 - \alpha_2)\left(x_2 - \left(\alpha_2 + \frac{1}{4}\right)\right) + \frac{1}{8},$$

we get 12 vectors that are contained in the Lie algebra at (x, Id) from the same methodology as that done in the case of the two point process.

It turns out that the span of these 12 vector fields is contained in the span of

$$(6.19) \quad \begin{pmatrix} \kappa \\ 0 \\ 0 \\ 0 \\ \partial_1 U \\ \partial_1 U \end{pmatrix}, \begin{pmatrix} 0 \\ \kappa \\ 0 \\ 0 \\ -\partial_2 U \\ -\partial_2 U \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 2 \\ 1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \partial_1 U \partial_2 U + \kappa \partial_{1,1} U \\ \partial_1 U \partial_2 U + \kappa \partial_{1,1} U \end{pmatrix}.$$

While this can also be directly checked, the computation is somewhat tedious. Thus we include code verifying this symbolically [CFIN23].

The span of the five vectors shown in (6.19) may be less than five-dimensional. If, however, the 5th vector is non-zero, the span is five-dimensional. It is not easy to verify that the the 5th vector is non-zero on a connected dense set. Therefore, we use different vectors to obtain the full rank.

Computing the Lie bracket of $\tilde{v}_{\alpha_1,1}$ and $\tilde{v}_{\alpha_2,2}$, defined in (6.18), we obtain a quadratic polynomial in α_1 and α_2. Similar to (6.16), we now obtain a collection of 9 vectors that lie in our Lie Algebra. Using a computer to perform symbolic Gauss-Jordan elimination, we can use these 9 vectors and (6.19) to obtain another set of 9 vectors, such that their first 4 coordinates are 0. (The code for this can also be found [CFIN23].) The resulting vectors, and the last vector in (6.19) yields the following non-degeneracy condition: the dimension of the Lie algebra at (x, Id) is 5 if at least one of the following five inequalities hold

$$0 \neq \partial_{1,2} U$$
$$0 \neq \partial_1 U \partial_2 U + \kappa \partial_{1,2} U$$
$$0 \neq \partial_1 U \partial_2 U + (\partial_1 U)^2 + 3\kappa \partial_{1,2} U$$
$$0 \neq 16\kappa + \partial_{2,2} U + \partial_{1,1} U - 3\partial_{1,2} U$$
$$0 \neq \kappa \partial_{2,2} U + \kappa \partial_{1,1} U + 4 (\partial_1 U)^2 + \kappa \partial_{1,2} U.$$

We claim that for any $\kappa \neq 0$, at least one of the above non-degeneracy conditions must hold. Indeed if all the above non-degeneracy conditions fail, then we must have

$$\Delta U(x) = 0 \quad \text{and} \quad \Delta U(x) = -16\kappa,$$

which is impossible. Thus, the Lie algebra at (x, Id) has dimension 5 and we can apply the Rachevsky–Chow Theorem. This completes the proof of irreducibility of the rescaled derivative process (6.9). □

Remark 6.7. Note that we demonstrated controllability of the two point process only on a connected dense set of $\mathcal{D}^c$, but controllability of the rescaled derivative process was shown on the entire set $\mathbb{T}^2 \times SL_2(\mathbb{R})$.

6.4. **Positivity of top Lyapunov Exponent.** Controllability of the rescaled derivative process implies that the top Lyapunov exponent is positive. This follows from the following version of Furstenberg's criterion.

Theorem 6.8 (Furstenberg's criterion)**.** *For a sequence of elementary events $\underline{\omega} = (\omega_1, \omega_2, \dots) \in \Omega^{\mathbb{N}}$, continuous random dynamical system $f = f_\omega(x)$, $x \in X$ and measurable $A : \Omega \times X \to \mathrm{SL}_d(\mathbb{R})$ consider a linear cocycle*

$$f_{\underline{\omega}}^n = f_{\omega_n} \circ \cdots \circ f_{\omega_1} \text{ and } A_{\underline{\omega},x}^n = A_{w_n, f_{\underline{\omega}}^{n-1}(x)} \circ \dots A_{\omega_1, x}.$$

Suppose $A_{\underline{\omega},x}^n$ is integrable, and π is the invariant measure of the Markov process corresponding to f. Then $A_{\underline{\omega},x}^n$ has two Lyapunov exponents. Let $\lambda^+ \geqslant 0$ be the larger of the two. Then λ^+ can only be 0 if there exists a family of probability measures $\{\nu_x\}_{x\in\mathrm{supp}\{\pi\}}$ on $\mathbb{P}^1$ varying measurably in x such that for every $x \in \mathrm{Support}(\pi)$, $n \geqslant 1$, and all pairs

$$(6.20) \qquad\qquad (y, A) \in \mathrm{Support}(\mathrm{Law}(f_{\underline{\omega}}^n(x), A_{\underline{\omega},x}^n))$$

we have that the pushforward of ν_y satisfies

$$(6.21) \qquad\qquad \left(A^T\right)_* \nu_y = \nu_x.$$

Remark 6.9. There are many versions of the Furstenberg's criterion. We rescale the derivative process to use this version of Furstenberg's criterion. The version of Furstenberg's criterion given by Theorem 6.8 is a combination of Proposition 2 and Theorem 3 in [Led86] with one difference. This version has an extra conclusion that Equation (6.21) holds for all elements in the support of the n-step transition probabilities. The conclusion follows from arguments done in Bougerol's paper [Bou88], and was explicitly proven by A. Blumenthal et al. [BCZG22, Proposition 2.9].

Let us recall definitions of notions that arise in Theorem 6.8(refer to [Arn98, Chapter 3]). Briefly recall that a *continuous random dynamical system* (with independent increments in our case) on a metric space X, over a probability space $(\Omega, \mathcal{G}, \boldsymbol{P})$, is a mapping $(\omega, x) \mapsto f_\omega(x)$ such that for every fixed $x \in X$ the function $f_\omega(x)$ is a random variable in ω, and for every fixed $\omega \in \Omega$ the function $f_\omega(x)$ is continuous in x. Such random dynamical systems *correspond* to Markov processes with transition probabilities

$$P(x, A) = \boldsymbol{P}(f_\omega(x) \in A).$$

A cocycle is *integrable* if

$$\int \left(\ln^+\|A_{\omega,x}\| + \ln^+\|A_{\omega,x}^{-1}\|\right) \, d\boldsymbol{P}(\omega)\, d\pi(x).$$

Note that the rescaled derivative process (6.9) is a continuous random dynamical system with integrable linear cocycle, and so we can apply Theorem 6.8 to show that the single point process (6.6) has a positive Lyapunov exponent.

Lemma 6.10. *The top Lyapunov exponent of the single point process (6.6) is positive.*

Proof. Suppose towards contradiction that the top Lyapunov exponent of the single point process (6.6) is 0. This implies that the top Lyapunov exponent of the linear cocycle A_n (defined in (6.9)) is 0. Indeed, since U is bounded the measure μ is equivalent to the Lebesgue measure on $\mathbb{T}^2$. So the μ-measure preserving map ϕ_n satisfies

$$0 < C_1 \leqslant \det D_x\phi_n(x) \leqslant C_2 < \infty,$$

uniformly in $n \in \mathbb{N}$, $x \in \mathbb{T}^2$ and the parameters $(\alpha, \beta, i) \in [0,1] \times [0,1] \times \{0,1\}$. This implies that the top Lyapunov exponent of the single point process is the same as the top Lyapunov exponent of the rescaled derivative process (6.9). Indeed, for any $v \in \mathbb{R}^2$ we have

$$\lim_{n\to\infty} \frac{1}{n} \ln\|A_n v\| = \lim_{n\to\infty} \frac{1}{n} \left(\ln\|D_x\phi_n v\| - \ln\left(\det D_x\phi_n\right)\right) = \lim_{n\to\infty} \frac{1}{n} \ln\|D_x\phi_n v\|,$$

almost surely with respect to $\boldsymbol{P} \times \mu$.

Now assume that the top Lyapunov exponent of the linear cocycle A_n is 0. Suppose ν_x is a family of probability measures produced by Theorem 6.8. Identify $\mathbb{P}^1$ with S^1 by considering the elements of elements of $\mathbb{P}^1$ as unit vectors in $\mathbb{R}^2$. Let $\varepsilon > 0$ be arbitrary and fix $x \in \mathbb{T}^2$. Further suppose that $\mathcal{U} \subset \mathbb{P}^1$, $\mathcal{U} = (a - \varepsilon, a + \varepsilon)$ such that $\nu_x(\mathcal{U}) > 0$. We can select $B_k \in SL_2(\mathbb{R})$ such that B_k contracts $\mathcal{U}$ into $\mathcal{U}$ by mapping $a - \varepsilon$ to $a - \varepsilon/k$, and $a + \varepsilon$ to $a + \varepsilon/k$. In other words, we claim we can find $B_k \in SL_2(\mathbb{R})$ such that

$$(6.22) \qquad B_k \mathcal{U} \subset \mathcal{U}, \quad a \in B_k \mathcal{U}, \quad \text{and} \quad \mathrm{Leb}\,(B_k \mathcal{U}) < \frac{\varepsilon}{k},$$

for an arbitrary $k \in \mathbb{N}$. Set $A_k = \left(B_k^T\right)^{-1}$. By controllability of the rescaled derivative process (see Remark 6.7), we can reach (x, A_k) from (x, Id) in finite time. Therefore, (x, A_k) satisfies condition (6.20) and so equation (6.21) for each B_k becomes

$$\nu_x\,(\mathcal{U}) = \nu_x\,(B_k \mathcal{U})\,.$$

Since k was arbitrary

$$\nu_x(\{a\}) = \nu_x(\mathcal{U}) > 0.$$

Select any $c \in \mathcal{U}$, $c \neq a$. Analogous to (6.22), we can select C_k such that

$$C_k \mathcal{U} \subset \mathcal{U}, \quad c \in C_k \mathcal{U}, \quad \text{and} \quad \mathrm{Leb}\,(C_k \mathcal{U}) < \frac{\varepsilon}{k}\,.$$

Thus, again $\nu_x(\{c\}) = \nu_x(\mathcal{U}) > 0$, but $\nu_x(\mathcal{U}) \geqslant \nu_x(\{a\}) + \nu_x(\{c\})$. So we have reached a contradiction and the top Lyapunov exponent of the rescaled derivative process is positive. This implies the top Lyapunov exponent of the single point process is positive. $\qquad \square$

6.5. **Feller Property.** To prove that the modified, randomly shifted, sawtooth shears are exponentially mixing it remains to show that that the single point, projective and two point processes are Feller. Recall, a process is Feller if the function $x \mapsto \boldsymbol{E}^x[f(X_1)]$ is continuous for every bounded continuous function f.

Lemma 6.11. *The single point process* (6.6), *projective process* (6.8), *and two point process* (6.10) *are Feller.*

Remark 6.12. No randomness is needed to show that the single point and two-point processes are Feller. We will, however, use randomness to show that the projective process is Feller. We will capitalize on the fact that the parameter α in (6.2) is uniformly distributed on $[0, 1]$.

Proof of Lemma 6.11. The velocity fields given in (6.3) and (6.4) are uniformly Lipschitz. Thus, the two point process satisfies a bound similar to equation (6.11) in Remark 6.1. Therefore, it is Feller. Similarly the single point process is Feller.

We will now show that the projective process is Feller. Note that $v_{\alpha,i}(x)$, and, therefore, the time 1 flows $\varphi^{\beta v_{\alpha,i}}(x)$ and its derivatives $D_x \varphi^{\beta v_{\alpha,i}}(x)$ are smooth on $\mathbb{T}^2$ except for three lines: $x_i = \alpha$, $x_i = \alpha + 1/4$, and $x_i = \alpha + 3/4$. In order to compute the expectation

$$(6.23) \qquad \boldsymbol{E}^{(x,u)}[f(\phi_1(x), D_x\phi_1 u)] = \boldsymbol{E}^{(x,u)}[f(\varphi^{V_1}(x), D_x\varphi^{V_1} u)],$$

we need to fix $(x, u) \in \mathbb{T}^2 \times \mathbb{P}^1$ and integrate over the parameter space, $(\alpha, \beta, i) \in [0, 1] \times [0, 1] \times \{1, 2\}$. The function $f(\phi_1(x), D_x\phi_1(x)u)$ is uniformly bounded. It is also continuous everywhere except $x_i = \alpha$, $x_i = \alpha + 1/4$, and $x_i = \alpha + 3/4$.

Therefore its integral with respect to the parameter α is continuous. Thus the expectation (6.23) is continuous with respect to x and u.

We now do the formal proof that the projective process is Feller. Fix $f \in \mathcal{C}(\mathbb{T}^2 \times \mathbb{P}^1)$ and let $\varepsilon > 0$ be arbitrary. Let $d(\cdot, \cdot)$ be the product distance on $\mathbb{T}^2 \times \mathbb{P}^1$. First select $\delta_1 > 0$ such that $6C\delta_1 < \varepsilon/2$, where C is such that $f(z) \leqslant C$ for all $z \in \mathbb{T}^2 \times \mathbb{P}^1$. Now select $0 < \delta_2 < \delta_1$ so that if $(x, u), (y, u') \in \mathbb{T}^2 \times \mathbb{P}^1$ are such that $d((x, u), (y, u')) < \delta_2$, and x and y are in the same smooth piece of $\beta v_{\alpha,i}$ we have that

$$(6.24) \qquad |f(\phi_t^{\beta v_{\alpha,i}}(x), D_x\phi_t^{\beta v_{\alpha,i}}u) - f(\phi_t^{\beta v_{\alpha,i}}(y), D_x\phi_t^{\beta v_{\alpha,i}}u')| < \frac{\varepsilon}{2}$$

for all $\alpha, \beta \in [0, 1]$ and $i \in \{1, 2\}$. Let (x, u), $(y, u') \in \mathbb{T}^2 \times \mathbb{P}^1$ be such that $d((x, u), (y, u')) < \delta_2$ and

$$A = \{(\alpha, \beta, i) \in [0, 1] \times [0, 1] \times \{0, 1\} \mid \text{equation (6.24) does not hold}\}.$$

Note that $\boldsymbol{P}(A) < 6\delta_1$ since there are both vertical and horizontal modified shear flows, and each modified shear flow has 3 different smooth pieces. This gives

$$\left| \boldsymbol{E}^{(x,v)}[f(X_1, V_1)] - \boldsymbol{E}^{(y,v')}[f(X_1, V_1)] \right|$$
$$\leqslant \int_A |f(\phi_1(x), D_x\phi_1(x)v) - f(\phi_1(y), D_x\phi_1(y)v')| \, d\boldsymbol{P}$$
$$+ \int_{A^c} |f(\phi_1(x), D_x\phi_1(x)v) - f(\phi_1(y), D_x\phi_1(y)v')| \, d\boldsymbol{P}$$
$$< \frac{\varepsilon}{2} + 6C\delta_1 < \varepsilon.$$

Thus, the projective process is Feller. $\qquad\square$

6.6. **Proof of Theorem 1.3.**

Proof of Theorem 1.3 with a sawtooth shear profile in two dimensions. The processes (6.6), (6.8), and (6.10) are all irreducible by Lemma 6.6. The three processes are Feller by Lemma 6.11. Thus, both the derivative and single point processes are uniformly geometrically ergodic. Furthermore, the single point process has a positive Lyapunov exponent by Lemma 6.10. Finally, the two point process is $\mathcal{V}$-geometrically ergodic with respect to a function of the form (6.12). This follows from section 6 in [BBPS22], and the fact that the two point process (6.6) is irreducible and Feller. Thus, by Proposition 6.2 the modified, randomly shifted sawtooth shears are almost surely mixing. $\qquad\square$

We now consider modified, randomly shifted, sine shears. In this case we take the stream function F_0 in (6.1) as

$$F_0(x) = \sin(2\pi x),$$

and define

$$(6.25) \qquad\qquad F_\alpha(x) = F_0(x - \alpha) \text{ for } \alpha \in [0, 1].$$

The vector fields (6.3) and (6.4) are defined using the stream functions given by (6.25). All other definitions are the same.

Proof of Theorem 1.3 with a sine shear profile in two dimensions. After we prove that both the two point process and rescaled derivative process are irreducible, the proof with a sine stream function is identical to the proof with a tent stream function. We will now show that the two point process is irreducible. As in Lemma 6.6 we must show that the Lie algebra generated by the vector fields

$$\{(v_{\alpha,i}^T(x), v_{\alpha,i}^T(y))^T\}$$

has dimension 4 on a dense connected set. We will show that the Lie algebra has dimension 4 on the set

$$M \overset{\text{def}}{=} \mathcal{D}^c - S,$$

where

$$S = \{x_1 = y_1 \text{ or } x_2 = y_2\} \cup \{x \text{ or } y \text{ is a critical point of } U\} \subset \mathbb{T}^2 \times \mathbb{T}^2.$$

By selecting $\alpha = x_2$ and $\alpha = x_1$ we obtain the 2 linearly independent vectors,

$$(6.26) \quad s_2 \overset{\text{def}}{=} -\kappa \begin{pmatrix} v_{x_2,2}(x) \\ v_{x_2,2}(y) \end{pmatrix} = \begin{pmatrix} 2\pi\kappa \\ 0 \\ 2\pi\kappa\cos(2\pi(x_2 - y_2)) + \sin(2\pi(x_2 - y_2))\partial_2 U(y) \\ -\sin(2\pi(x_2 - y_2))\partial_1 U(y) \end{pmatrix}$$

and

$$(6.27) \quad s_1 \overset{\text{def}}{=} \kappa \begin{pmatrix} v_{x_1,1}(x) \\ v_{x_1,1}(y) \end{pmatrix} = \begin{pmatrix} 0 \\ 2\pi\kappa \\ \sin(2\pi(y_1 - x_1))\partial_2 U(y) \\ 2\Pi\kappa\cos(2\pi(y_1 - x_1)) - \sin(2\pi(y_1 - x_1))\partial_1 U(y) \end{pmatrix}.$$

Using a computer to perform symbolic Gauss-Jordan elimination with s_2 and s_1 from (6.27) and (6.26) we can reduce the vector

$$\left[\begin{pmatrix} v_{\alpha,1} \\ v_{\alpha,1} \end{pmatrix}, \begin{pmatrix} v_{\beta,2} \\ v_{\beta,2} \end{pmatrix} \right](x, y)$$

to a vector L_1 with 0 in the first two coordinates. Using trigonometric identities we can write L_1 as a linear combination of trigonometric products in the form

$$\begin{aligned}
L_1 = {}& \sin(2\pi(\alpha - x_1))\sin(2\pi(\beta - x_2))\ell_1 + \sin(2\pi(\alpha - x_1))\cos(2\pi(\beta - x_2))\ell_2 \\
& + \cos(2\pi(\alpha - x_1))\sin(2\pi(\beta - x_2))\ell_3 + \cos(2\pi(\alpha - x_1))\cos(2\pi(\beta - x_2))\ell_4 \\
& + \sin(2\pi(\alpha - y_1))\sin(2\pi(\beta - y_2))\ell_5 + \sin(2\pi(\alpha - y_1))\cos(2\pi(\beta - y_2))\ell_6 \\
(6.28) \quad & + \cos(2\pi(\alpha - y_1))\sin(2\pi(\beta - y_2))\ell_7 + \cos(2\pi(\alpha - y_1))\cos(2\pi(\beta - y_2))\ell_8,
\end{aligned}$$

for some explicit vectors ℓ_i, $i = 1, 2, \ldots, 8$. Similarly we can reduce

$$\begin{pmatrix} v_{\alpha,1} \\ v_{\alpha,1} \end{pmatrix}, \begin{pmatrix} v_{\alpha,2} \\ v_{\alpha,2} \end{pmatrix}, \left[\begin{pmatrix} v_{\alpha,1} \\ v_{\alpha,1} \end{pmatrix}, \begin{pmatrix} v_{\beta,1} \\ v_{\beta,1} \end{pmatrix} \right](x, y), \text{ and } \left[\begin{pmatrix} v_{\alpha,2} \\ v_{\alpha,2} \end{pmatrix}, \begin{pmatrix} v_{\beta,2} \\ v_{\beta,2} \end{pmatrix} \right](x, y)$$

to the vectors

$$\begin{aligned}
L_2 &= \sin(2\pi(\alpha - x_1))\ell_9, \quad L_3 = \sin(2\pi(\alpha - x_2))\ell_{10}, \\
L_4 &= \sin(2\pi(\alpha - \beta)\ell_{11}, \quad L_5 = \sin(2\pi(\alpha - \beta))\ell_{12},
\end{aligned}$$

respectively, for some explicit vectors $\ell_9, \ldots, \ell_{12}$, which all have a 0 in their first two coordinates.

On the set M since $x_1 \neq y_1$ and $x_2 \neq y_2$, the trigonometric products in (6.28) are linearly independent unless

$$x_1 = y_1 + \frac{1}{2} \quad \text{and} \quad x_2 = y_2 + \frac{1}{2}.$$

We first consider the case when $x_1 \neq y_1 + 1/2$ or $x_2 \neq y_2 + 1/2$. In this case we can use different choices of α, β in (6.28) and take linear combinations to ensure each ℓ_i is contained in the Lie algebra at (x, y). In particular, the vectors ℓ_5 and ℓ_8 are contained in the Lie algebra, and we symbolically compute

(6.29)
$$\frac{\ell_5}{4\pi^2\kappa^2} = \begin{pmatrix} 0 \\ 0 \\ \partial_1 U(y) \\ -\partial_2 U(y) \end{pmatrix}, \qquad \frac{\ell_8}{4\pi^2\kappa^2} = \begin{pmatrix} 0 \\ 0 \\ -\partial_2 U(y) \\ \partial_1 U(y) \end{pmatrix}.$$

Thus the Lie algebra has dimension 4 unless

$$\partial_1 U(y) = -\partial_2 U(y) \quad \text{or} \quad \partial_1 U(y) = \partial_2 U(y).$$

Symbolic computations (see [CFIN23]) show that in either case, the vector

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}$$

is contained in the span of the remaining ℓ_i. Since in M both $\partial_1 U(y)$ and $\partial_2 U(y)$ can not vanish, using ℓ_5 or ℓ_8 from (6.29) shows that the Lie algebra has dimension 4 as desired.

Now consider the case where $x_1 = y_1 + 1/2$ and $x_2 = y_2 + 1/2$. We rewrite (6.28) as

$$L_1 = \sin(2\pi(\alpha - x_1))\sin(2\pi(\beta - x_2))\ell_1' + \sin(2\pi(\alpha - x_1))\cos(2\pi(\beta - x_2))\ell_2'$$
$$+ \cos(2\pi(\alpha - x_1))\sin(2\pi(\beta - x_2))\ell_3' + \cos(2\pi(\alpha - x_1))\cos(2\pi(\beta - x_2))\ell_4',$$

for some explicit vectors $\ell_1', \ldots, \ell_4'$. We symbolically compute and verify

$$\begin{pmatrix} 0 \\ 0 \\ 16\pi^3\kappa^2 \\ 0 \end{pmatrix} = \ell_{11}\kappa^2 + \frac{\ell_3'}{\kappa} \quad \text{and} \quad \begin{pmatrix} 0 \\ 0 \\ 0 \\ 16\pi^3\kappa^2 \end{pmatrix} = \ell_{12}\kappa^2 - \frac{\ell_2'}{\kappa},$$

which implies that the Lie algebra is 4 dimensional on M if $x_1 = y_1 + 1/2$ and $x_2 = y_2 + 1/2$. Thus, in either case the Lie algebra is 4 dimensional, and hence the two point process is irreducible.

We now show that the rescaled derivative process (6.9) is irreducible by showing that the Lie algebra at (x, Id) is 5 dimensional on the set

$$M' \overset{\text{def}}{=} \{x \in \mathbb{T}^2 \mid \nabla U(x) \neq 0\}.$$

Since M' is connected, this will imply irreducibility of the rescaled derivative process exactly as in the proof of Lemma 6.6.

We consider the vector fields (6.18) with a sine stream function. That is, vectors with first two coordinates coming from the coordinates of (6.3) or (6.4) with a sine

stream function, and last four coordinates coming from the corresponding matrix, $A_t(v, x)$ (defined in (6.17)). By choosing $\alpha = x_2$, and $\alpha = x_1$ in (6.18) we compute

$$
\tilde{s}_2 \overset{\text{def}}{=} -2\kappa \tilde{v}_{x_2, 2} = \begin{pmatrix} 4\pi\kappa \\ 0 \\ -2\pi\partial_1 U(x) \\ -4\pi\partial_2 U(x) \\ 0 \\ 2\pi\partial_1 U(x) \end{pmatrix} \quad \text{and} \quad \tilde{s}_1 \overset{\text{def}}{=} 2\kappa \tilde{v}_{x_1, 1} = \begin{pmatrix} 0 \\ 2\pi\kappa \\ \pi\left(\partial_2 U(x) - \partial_1 U(x)\right) \\ 0 \\ -4\pi\partial_1 U(x) \\ -2\pi\left(\partial_2 U(x) - \partial_1 U(x)\right) \end{pmatrix}.
$$

By performing symbolic Gauss-Jordan elimination on the vectors $\tilde{v}_{\alpha,1}$, $\tilde{v}_{\alpha,2}$, $\tilde{s}_1$ and $\tilde{s}_2$, yields the vectors

$$
\tilde{L}_1 = \sin(2\pi(\alpha - x_1))\tilde{\ell}_1 \quad \text{and} \quad \tilde{L}_2 = \sin(2\pi(\alpha - x_1))\tilde{\ell}_2
$$

for some explicit vectors $\tilde{\ell}_1$ and $\tilde{\ell}_2$ whose first two coordinates are both 0. We symbolically compute

$$
\tilde{s}_3 \overset{\text{def}}{=} 2(\tilde{\ell}_1 - \tilde{\ell}_2) = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 2 \\ 2 \\ 1 \end{pmatrix}.
$$

Using the $\tilde{s}_i$ we symbolically compute and reduce the vectors

$$
[\tilde{v}_{\alpha,1}, \tilde{v}_{\beta,2}], [\tilde{v}_{\alpha,1}, \tilde{v}_{\beta,1}], \text{ and } [\tilde{v}_{\alpha,2}, \tilde{v}_{\beta,2}]
$$

to the vectors

$$
\tilde{L}_3 = \sin(2\pi(\alpha - x_1))\sin(2\pi(\beta - x_2))\tilde{\ell}_3 + \sin(2\pi(\alpha - x_1))\cos(2\pi(\beta - x_2))\tilde{\ell}_4
$$
$$
+ \cos(2\pi(\alpha - x_1))\sin(2\pi(\beta - x_2))\tilde{\ell}_5 + \cos(2\pi(\alpha - x_1))\cos(2\pi(\beta - x_2))\tilde{\ell}_6\,,
$$
$$
\tilde{L}_4 = \sin(2\pi(\alpha - \beta))\tilde{\ell}_7\,, \qquad \text{and} \qquad \tilde{L}_5 = \sin(2\pi(\alpha - \beta))\tilde{\ell}_8\,,
$$

respectively, for some explicit vectors $\tilde{\ell}_3, \ldots, \tilde{\ell}_8$, which all have zeros in the first three coordinates. By choosing $\alpha = x_1 + 1/4, x_1 - 1/4, x_2, x_1, x_2 - 1/4$, and taking Lie brackets, we obtain three more vectors

$$
\tilde{v}_{x_1 - 1/4, 1}\,, \quad [\tilde{v}_{x_1 - 1/4, 1}, \tilde{v}_{x_2, 2}]\,, \quad [\tilde{v}_{x_1 - 1/4, 1}, \tilde{v}_{x_2 - 1/4, 2}] \quad \text{and} \quad [\tilde{v}_{x_1, 1}, \tilde{v}_{x_2, 2}]\,.
$$

Again using the $\tilde{s}_i$ and performing symbolic Gauss-Jordan elimination we obtain three new vectors, $\tilde{\ell}_9$, $\tilde{\ell}_{10}$, $\tilde{\ell}_{11}$, and $\tilde{\ell}_{12}$ which have zeros in the first three coordinates. We now symbolically compute and check that the span of $\text{span}\{\tilde{\ell}_1, \ldots, \tilde{\ell}_{12}\}$ also contains the vectors

$$
\tilde{\ell}_{13} \overset{\text{def}}{=} \begin{pmatrix} 0 \\ 0 \\ 0 \\ \frac{1}{2}\left(8\pi^2\kappa^2 + \partial_2 U(x)^2 - 6\partial_1 U(x)\partial_2 U(x) + 5\partial_1 U(x)^2\right) \\ 2\partial_1 U(x)\left(\partial_1 U(x) - \partial_2 U(x)\right) \\ \partial_1 U(x)\left(\partial_1 U(x) - \partial_2 U(x)\right) \end{pmatrix}
$$

and

$$\tilde{\ell}_{14} \overset{\text{def}}{=} \begin{pmatrix} 0 \\ 0 \\ 0 \\ \tilde{\ell}_{14,4} \\ \partial_1 U(x)\partial_2 U(x) + \kappa\partial_{1,2}U(x) \\ \partial_1 U(x)\partial_2 U(x) + \kappa\partial_{1,2}U(x) \end{pmatrix},$$

where

$$\tilde{\ell}_{14,4} = \frac{1}{2}\left(-4\pi^2\kappa^2 - \partial_2 U(x)^2 - \kappa\Delta U(x) + 2\partial_2 U(x)\partial_1 U(x) - \partial_1 U(x)^2 + 2\kappa\partial_{1,2}U(x)\right).$$

Clearly, the two vectors $\tilde{\ell}_{13}$, $\tilde{\ell}_{14}$ are linearly independent unless

$$\partial_1 U(x)\partial_2 U(x) + \kappa\partial_{1,2}U(x) = 0 \text{ or } \partial_1 U(x)\left(\partial_1 U(x) - \partial_2 U(x)\right) = 0.$$

In either case we symbolically compute and explicitly verify that the span of the $\tilde{\ell}_i$ is 2 dimensional (see [CFIN23] for details). This shows that the Lie algebra is 5 dimensional, and hence the rescaled derivative process is irreducible. $\qquad\square$

Remark 6.13. If the potential is constant then the vectors (6.29) do not prove irreducibility of the two point process. However, similar symbolic computations can still be used to show that the two point process is irreducible.

When $d > 2$, we perform the 2-dimensional dynamics on pairs of coordinates. That is, for a pair of indices (i,j), $i < j$ we consider the velocity fields

$$v_{\alpha,(i,j),1} = \frac{1}{p}\nabla_{i,j}^{\perp}(pF_{\alpha}(x_i)),$$

$$v_{\alpha,(i,j),2} = \frac{1}{p}\nabla_{i,j}^{\perp}(pF_{\alpha}(x_j)),$$

where $\nabla_{i,j}^{\perp}$ is the 2-dimensional perpendicular gradient on the coordinate pair (i,j) and 0 in the other coordinates. We define

$$V_n = \beta_n v_{\alpha_n,\xi_n,i_n},$$

where ξ_n are i.i.d. uniform random variables on the set of pairs of coordinates, and (α_n,β_n,i_n) are again i.i.d. random variables uniformly distributed on the parameter space $[0,1] \times [0,1] \times \{1,2\}$. All of the associated processes are defined identically to their 2-dimension counterparts (see (6.6), (6.8), and (6.9)).

Proof of Theorem 1.3 in dimension $d \geqslant 3$. The proof in high dimensions follows from the proof done in 2 dimensions. Both the Feller property (Lemma 6.11) and positivity of the top Lyapunov exponent (Lemma 6.10) are exactly the same as the proofs done in dimension 2. Irreducibility of the single point and two point processes also follows immediately from Lemma 6.6. For each pair of coordinates we can perform the same computations done in Lemma 6.6. This implies that the span of the vector fields that generate the two point process has dimension $2d$, and so the single point and two point processes are irreducible.

The added assumption in Theorem 1.3 for high dimensions that (1.14) holds on a small ball, $B(\hat{x},\hat{\varepsilon})$ allows us to generalize the proof of irreducibility of the derivative and projective processes. Irreducibility follows if we show irreducibility on the subset $B(\hat{x},\hat{\varepsilon})$. Indeed, since the single point process is irreducible we can first move the position component into $B(\hat{x},\hat{\varepsilon})$, control the derivative without moving

the single point out of $B(\hat{x}, \hat{\varepsilon})$, and then move the position and derivative pair to the end condition. Thus, irreducibility on the set $B(\hat{x}, \hat{\varepsilon})$ implies irreducibility of both processes.

Irreducibility of the processes on the set $B(\hat{x}, \hat{\varepsilon})$ follows from counting the dimension of the Lie algebra of the rescaled derivative process. Consider the d-dimension counterparts to (6.17) and (6.18). We will refer to the coordinate that corresponds to the (i, j) entry of A_t as coordinate (i, j). Since the potential is separable on $B(\hat{x}, \hat{\varepsilon})$, the Lie bracket of the $d^2 + d$-dimensional counterparts to (6.18) can only be non-zero in 6 positions. That is, for a pair of coordinates $i < j$

$$[\tilde{v}_{\alpha,(i,j),1}, \tilde{v}_{\alpha,(i,j),2}]$$

can be non-zero in the i, j, (i, i), (j, j), (i, j), and (j, i) coordinates. Thus, for each pair of coordinates we can perform the same computations done in the 2 dimensional cases.

First, consider the case of tent shears. Clearly, each coordinate gives rise to a linearly independent vector (vectors 1 and 2 in (6.19)). For each pair $i < j$ of coordinates we obtain 2 vectors with nonzero entries in the (i, j) and (j, i) coordinates (vectors 3 and 4 in (6.19)) that are not in the span of the first d vectors. This gives d^2 linearly independent vectors in total. The computations in Lemma 6.6 for the coordinate pair $i, i - 1$ add another $d - 1$ vectors to the collection of linearly independent vectors (5$^{\text{th}}$ vector found in Lemma 6.6). Thus the dimension of the Lie algebra of the rescaled derivative process is $d^2 + d - 1$ on $B(\hat{x}, \hat{\varepsilon}) \times Id$ and so both the derivative and projective processes are irreducible (see Lemma 6.6 for details). The conclusion of Theorem 1.3 now follows from the same exact argument done in the earlier parts of this section. Counting the dimension of the Lie algebra in the case of sine shears is similar, and omitted for brevity.   □

Appendix A. Lebesgue Exponential Mixing of Randomly Shifted Tent Flows

Studying mixing rates of incompressible flows is an area of active research [Bre06, CDL08, Thi12, Sei13, IKX14], and examples of exponentially mixing flows were only constructed recently [YZ17, ACM19, EZ19, BBPS22, BCZG22, Coo22]. We note that when F is the sawtooth shear (6.1) (shown in Figure 1), then Theorem 1.3 still holds when $U = 0$. In this case, μ is simply the Lebesgue measure, and hence the flow v (defined in (1.13), with F as in (6.1)) is almost surely (Lebesgue) exponentially mixing. The proof of Theorem 1.3, however, involves technical calculations to check irreducibility. If instead of using the sawtooth profiles (6.1), we use a localized tent function, then the proof of irreducibility becomes extremely simple. We devote this appendix to presenting this, and hence obtain a short, simple, proof showing a family of incompressible flows is almost surely (Lebesgue) exponentially mixing.

Let F_0 be the localized tent function shown in Figure 4. Explicitly, F_0 a piece-wise linear periodic function such that

$$(A.1) \qquad F_0\left(\frac{1}{2} \pm \frac{1}{8}\right) = 0, \quad F_0\left(\frac{1}{2}\right) = 1, \quad \text{and} \quad F_0(x) = 0 \text{ for } x \in \left[\frac{3}{8}, \frac{5}{8}\right]^c.$$

For $\alpha \in [0, 1]$ define

$$F_\alpha(x) = F_0(x - \alpha),$$
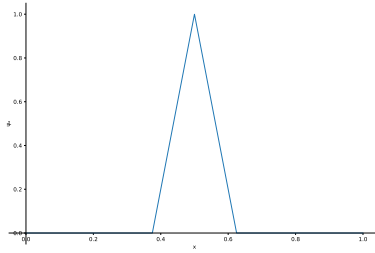
FIGURE 4. Profile of the localized tent function ψ_0.

so that F_α is a localized tent function with peak at $1/2 + \alpha$. Define the associated horizontal and vertical shear flows by

$$(A.2) \qquad v_{\alpha,1}(x) = \begin{pmatrix} F(x_2 - \alpha) \\ 0 \end{pmatrix}, \quad \text{and} \quad v_{\alpha,2}(x) = \begin{pmatrix} 0 \\ F(x_1 - \alpha) \end{pmatrix}.$$

respectively. We will now show that if we randomly shift (and modulate) the localized tent shears flows $v_{\alpha,i}$ then we are (Lebesgue) exponentially mixing almost surely.

Theorem A.1. *Define the velocity field v by*

$$v_t = \beta_n v_{\alpha_n, i_n} \qquad \text{when } t \in [n, n+1),$$

where (α_n, β_n, i_n) are i.i.d. random variables that are uniformly distributed on parameter space $[0,1] \times [0,1] \times \{1,2\}$. The flow of v is almost surely (Lebesgue) exponentially mixing with rate (1.6).

Remark A.2. If instead of localized tent shears (A.1), we use sawtooth shears, then Theorem 1.3 already guarantees v is almost surely (Lebesgue) exponentially mixing. The reason we choose localized tent shears here is because the proof of irreducibility is simpler, and does not rely on the Rachevsky–Chow Theorem (Theorem 6.5). The local property makes the argument direct and highlights the fact that for a collection of vector fields to be almost surely mixing you must be able to control 2 points independently.

Remark A.3. In dimensions $d > 2$, Lebesgue exponentially mixing flows can be constructed by setting

$$v_t = \nabla^\perp_{i_n, j_n} F_{\alpha_{i_n}}(x_{j_n}),$$

where $\nabla^\perp_{i,j}$ is the skew gradient in the x_i-x_j plane (see (1.12)).

Proof of Theorem A.1. As in Section 6, define V_n by (6.5), with $v_{\alpha,i}$ as in (A.2), and consider the Markov process X_n defined by (6.6). Using the same notation as Section 6, define the random flow ϕ_n by (6.7).

We will prove Theorem A.1 by showing the conditions of Proposition 6.2 hold. As in Section 6 we need to show that the single point process (6.6) has a positive Lyapunov exponent, and that the single point process (6.6), two point process (6.10), and projective process (6.8) are all irreducible and Feller.

Since each velocity field is a shear the single point process is clearly irreducible. The existence of a positive Lyapunov exponent is immediate due to the classical result of Furstenberg [Fur63], which we quote below. The version we state is Theorem 4.1 in [BL85].

Theorem A.4. *Let G_n be a sequence of IID random matrices in $SL_2(\mathbb{R})$ with probability measure μ. If the smallest closed subgroup containing the support of μ is not compact and G_n does not leave a set of one or two lines invariant then there exists a constant $\lambda^+ > 0$ such that for any $u \in \mathbb{R}^2$,*

$$\lim_{n \to +\infty} \frac{1}{n} \ln \|G_n G_{n-1} \cdots G_1 u\| = \lambda^+,$$

almost surely.

Since $D_x \phi_n$ is independent of x it is a random product of shear matrices and so Theorem A.4 implies that the Lyapunov exponent is positive. (We remark that this argument is not specific to the localized tent shears, and the randomly shifted sawtooth shears also have a positive Lyapunov exponent for the same reason.)

Both processes are Feller by the same argument done in Lemma 6.11. We show irreducibility of the two point process and of the projective process in the following two Lemmas.

Lemma A.5. *The two point process (6.10) is irreducible.*

Lemma A.6. *The projective process (6.8) is irreducible.*

This concludes the proof of Theorem A.1, modulo the above lemmas. $\square$

It remains to prove Lemmas A.5 and A.6. We recall that the proof of irreducibility for Theorem 1.3 (Lemma 6.6), involved the Rachevsky–Chow theorem, and technical calculations that were checked symbolically. For localized tent flows and the Lebesgue measure, the proofs are short and simple.

Proof of Lemma A.5. We show that given $(x, y), (w, z) \in \mathcal{D}^c$ we can find a sequence parameters

$$\alpha_n, \beta_n \in [0, 1], \ i_n \in \{1, 2\}$$

so that the composition of the time 1 flows of the vector fields $\beta_n v_{\alpha_n, i_n}$, defined in (A.2), map x to w and y to z. Continuity in α and β then implies irreducibility of the two point process. The key observation of the proof is that the vector fields (A.2) can move the coordinates of two points $(x, y) \in \mathcal{D}^c$ independently.

Fix position $(x, y), (w, z) \in \mathcal{D}^c$ and let $d(\cdot, \cdot)$ be the distance on S^1. Suppose without loss of generality that $x_2 \neq y_2$. Then we can fine an open interval $I = (a, b) \subset [0, 1]$ so that for all $\alpha \in I$

$$F_\alpha(x_2) = 8(x_2 - \alpha) \quad \text{and} \quad F_\alpha(y_2) = 0.$$

Furthermore $F_\alpha(x_2)$ at one of the endpoints of I is 0. That is, $F_a(x_2) = 0$ or $F_b(x_2) = 0$. Thus, we can find a vector field from (A.2) whose time one flow translates x_1 without changing x_2, y_1, and y_2. By repeatedly selecting $\alpha \in I$, we can map x_1 to w_1 without changing any other coordinate. By repeating this process with each coordinate we can map (x, y) to (w, z) as desired. $\square$

Proof of Lemma A.6. We show that for any two elements, $(x, u), (y, u') \in \mathbb{T}^2 \times \mathbb{P}^1$, there is a sequence of $\alpha_n, \beta_n \in [0, 1]$, $i_n \in \{1, 2\}$ such that the composition of the time 1 flows of the vector fields, $\beta_n v_{\alpha_n, i_n}$, map x to y. Furthermore, the derivative of the composition of time one flows map u to u'. Continuity in α and β implies irreducibility of the projective process. The lemma follows by observing that for any $x \in \mathbb{T}^2$ we can find v_n, defined in (A.2), such that the derivative matrix at x is given by

$$(A.3) \qquad A_\beta = \begin{pmatrix} 1 & 8\beta \\ 0 & 1 \end{pmatrix} \quad \text{or} \quad B_\beta = \begin{pmatrix} 1 & -8\beta \\ 0 & 1 \end{pmatrix},$$

which is notably independent α. Consider the projective process as elements of $\mathbb{R}^2 - \{0\}$ under the equivalence relation $u \sim u'$ if and only if $u = cu'$ for some constant c. For $u \in \mathbb{P}^1$, $u \neq (1, 0)$ we can rescale u so that the second coordinate is 1. Fix $(x, u), (y, u') \in \mathbb{T}^2 \times \mathbb{P}^1$. Without loss of generality assume that neither u nor u' point in the direction $\boldsymbol{e}_1 \overset{\text{def}}{=} (1, 0)$. Indeed, suppose that they they do. Then we can select a vertical shear so that (x, u) is mapped to say, $(x_0, u_0) \in \mathbb{T}^2 \times \mathbb{P}^1$ where $u_0 \neq \boldsymbol{e}_1$. Similarly, we can select a vertical shear that maps $(y_0, u_0') \in \mathbb{T}^2 \times \mathbb{P}^1$, $u_0' \neq \boldsymbol{e}_1$, to (y, u'). By mapping (x_0, u_0) to (y_0, u_0') we have a sequence of shears that maps (x, u) to (y, u'). We may further assume without loss of generality that $u_1 + 8 < u_1'$. This is possible by repeatedly selecting vector fields whose derivative at x is B_β, defined in (A.3).

We now give the proof. First map x_2 to y_2 without mapping u to $\boldsymbol{e}_1$. This is possible since for any $\beta \in [0, 1]$ there exists α so that the time 1 flow of $\beta v_{\alpha, 2}$ maps x_2 to y_2. Select a vector field $v = \beta v_{\alpha, 2}$ such that

$$D_x \varphi^v(x) u = A_\beta u = d = \begin{pmatrix} u_1 + 8\beta \\ 1 \end{pmatrix}$$

and $d_1 - u_1' \equiv 0 \mod 8$. By repeatedly selecting vector fields so that the derivative is always A_1 we can map (x, u) to

$$\left(z, \begin{pmatrix} u_1' - 8 \\ 1 \end{pmatrix} \right)$$

for some $z \in \mathbb{T}^2$ such that $z_2 = y_2$. Now select α such that $\varphi^{v_{\alpha, 1}}(z) = y$ and $D_x \varphi^{v_{\alpha, 1}} = A_1$. The sequence of parameters, α, β, i, provide a realization of the projective process which maps (x, u) to (y, u'). □

References

[ACM19]   G. Alberti, G. Crippa, and A. L. Mazzucato. Exponential self-similar mixing by incompressible flows. *J. Amer. Math. Soc.*, 32(2):445–490, 2019. doi:10.1090/jams/913.

[AdFDJ03] C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. An introduction to MCMC for machine learning. *Machine Learning*, 50:5 – 43, 2003. doi:10.1023/A:1020281327116.

[Ano67]   D. V. Anosov. Geodesic flows on closed Riemannian manifolds of negative curvature. *Trudy Mat. Inst. Steklov.*, 90:209, 1967.

[Arn98]   L. Arnold. *Random dynamical systems*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998. doi:10.1007/978-3-662-12878-7.

[ARNB20] E. Arnon, E. Rabani, D. Neuhauser, and R. Baer. Efficient Langevin dynamics for "noisy" forces. *The Journal of Chemical Physics*, 152(16):161103, 2020. doi:10.1063/5.0004954.

[BBPS22] J. Bedrossian, A. Blumenthal, and S. Punshon-Smith. Almost-sure exponential mixing of passive scalars by the stochastic Navier-Stokes equations. *Ann. Probab.*, 50(1):241–303, 2022. doi:10.1214/21-aop1533.

[BCP+21]   A. Blanca, P. Caputo, D. Parisi, A. Sinclair, and E. Vigoda. Entropy decay in the Swendsen-Wang dynamics on $\mathbb{Z}^d$. In *STOC '21—Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1551–1564. ACM, New York, [2021] ©2021. doi:10.1145/3406325.3451095.

[BCVD18]   A. Bouchard-Côté, S. J. Vollmer, and A. Doucet. The bouncy particle sampler: a nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.*, 113(522):855–867, 2018. doi:10.1080/01621459.2017.1294075.

[BCZG22]   A. Blumenthal, M. Coti Zelati, and R. S. Gvalani. Exponential mixing for random dynamical systems and an example of Pierrehumbert, 2022. doi:10.48550/ARXIV.2204.13651.

[BE85]   D. Bakry and M. Émery. Diffusions hypercontractives. In *Séminaire de probabilités, XIX, 1983/84*, volume 1123 of *Lecture Notes in Math.*, pages 177–206. Springer, Berlin, 1985. doi:10.1007/BFb0075847.

[Bet17]   M. Betancourt. A conceptual introduction to Hamiltonian Monte Carlo, 2017. doi:10.48550/ARXIV.1701.02434.

[BFR19]   J. Bierkens, P. Fearnhead, and G. Roberts. The zig-zag process and super-efficient sampling for Bayesian analysis of big data. *Ann. Statist.*, 47(3):1288–1320, 2019. doi:10.1214/18-AOS1715.

[BKK+85]   G. G. Batrouni, G. R. Katz, A. S. Kronfeld, G. P. Lepage, B. Svetitsky, and K. G. Wilson. Langevin simulations of lattice field theories. *Phys. Rev. D*, 32:2736–2747, Nov 1985. doi:10.1103/PhysRevD.32.2736.

[BL85]   P. Bougerol and J. Lacroix. *Products of random matrices with applications to Schrödinger operators*, volume 8 of *Progress in Probability and Statistics*. Birkhäuser Boston, Inc., Boston, MA, 1985. doi:10.1007/978-1-4684-9172-2.

[Bou88]   P. Bougerol. Comparaison des exposants de lyapounov des processus markoviens multiplicatifs. *Annales De L Institut Henri Poincare-probabilites Et Statistiques*, 24:439–489, 1988.

[Bre06]   A. Bressan. Prize offered for the solution of a problem on mixing flows. http://www.math.psu.edu/bressan/PSPDF/prize1.pdf, 2006.

[BS88]   P. H. Baxendale and D. W. Stroock. Large deviations and stochastic flows of diffeomorphisms. *Probab. Theory Related Fields*, 80(2):169–215, 1988. doi:10.1007/BF00356102.

[BW20]   O. Butterley and K. War. Open sets of exponentially mixing Anosov flows. *J. Eur. Math. Soc. (JEMS)*, 22(7):2253–2285, 2020. doi:10.4171/JEMS/964.

[BZ21]   S. Bianchini and M. Zizza. Properties of mixing bv vector fields, 2021, 2110.03581.

[CCAY+18]   X. Cheng, N. S. Chatterji, Y. Abbasi-Yadkori, P. L. Bartlett, and M. I. Jordan. Sharp convergence rates for Langevin dynamics in the nonconvex setting, 2018. doi:10.48550/ARXIV.1805.01648.

[CDL08]   G. Crippa and C. De Lellis. Regularity and compactness for the DiPerna-Lions flow. In *Hyperbolic problems: Theory, numerics, applications*, pages 423–430. Springer, Berlin, 2008. doi:10.1007/978-3-540-75712-2_39.

[CFIN23]   A. Christie, Y. Feng, G. Iyer, and A. Novikov. Symbolic computations verifying irreducibility conditions, 2023. URL https://gitlab.com/AlexanderCChristie/speeding-up-langevin-dynamics-by-mixing-auxiliary-code/-/tree/main.

[CGLL22]   S. Chewi, P. Gerber, H. Lee, and C. Lu. Fisher information lower bounds for sampling, 2022. doi:10.48550/ARXIV.2210.02482.

[Che23]   S. Chewi. *Log-Concave Sampling*. 2023. URL https://chewisinho.github.io/main.pdf.

[Cho40]   W.-L. Chow. Über systeme von liearren partiellen differentialgleichungen erster ordnung. *Mathematische Annalen*, 117(1):98–105, 1940. doi:10.1007/BF01450011.

[CKRZ08]   P. Constantin, A. Kiselev, L. Ryzhik, and A. Zlatoš. Diffusion and mixing in fluid flow. *Ann. of Math. (2)*, 168(2):643–674, 2008. doi:10.4007/annals.2008.168.643.

[Coo22]   W. Cooperman. Exponential mixing by shear flows, 2022, 2206.14239.

[Dal17a]   A. S. Dalalyan. Further and stronger analogy between sampling and optimization: Langevin Monte Carlo and gradient descent. In *Annual Conference Computational Learning Theory*, 2017. doi:10.48550/ARXIV.1704.04752.

[Dal17b]   A. S. Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *J. R. Stat. Soc. Ser. B. Stat. Methodol.*, 79(3):651–676, 2017. doi:10.1111/rssb.12183.

[DFY20]   M. Damak, B. Franke, and N. Yaakoubi. Accelerating planar Ornstein-Uhlenbeck diffusion with suitable drift. *Discrete Contin. Dyn. Syst.*, 40:4093, 2020. doi:10.3934/dcds.2020173.

[Dia09]   P. Diaconis. The Markov chain Monte Carlo revolution. *Bull. Amer. Math. Soc. (N.S.)*, 46(2):179–205, 2009. doi:10.1090/S0273-0979-08-01238-X.

[DKK04]   D. Dolgopyat, V. Kaloshin, and L. Koralov. Sample path properties of the stochastic flows. *Ann. Probab.*, 32(1A):1–27, 2004. doi:10.1214/aop/1078415827.

[DLP16]   A. B. Duncan, T. Lelièvre, and G. A. Pavliotis. Variance reduction using nonreversible Langevin samplers. *J. Stat. Phys.*, 163(3):457–491, 2016. doi:10.1007/s10955-016-1491-2.

[DM17]   A. Durmus and E. Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *Ann. Appl. Probab.*, 27(3):1551–1587, 2017. doi:10.1214/16-AAP1238.

[Dol98]   D. Dolgopyat. On decay of correlations in Anosov flows. *Ann. of Math. (2)*, 147(2):357–390, 1998. doi:10.2307/121012.

[DPZ17]   A. B. Duncan, G. A. Pavliotis, and K. C. Zygalakis. Nonreversible langevin samplers: Splitting schemes, analysis and implementation, 2017, 1701.04247.

[DT12]   A. S. Dalalyan and A. B. Tsybakov. Sparse regression learning by aggregation and Langevin Monte-Carlo. *J. Comput. System Sci.*, 78(5):1423–1443, 2012. doi:10.1016/j.jcss.2011.12.023.

[EZ19]   T. M. Elgindi and A. Zlatoš. Universal mixers in all dimensions. *Adv. Math.*, 356:106807, 33, 2019. doi:10.1016/j.aim.2019.106807.

[FBPR18] P. Fearnhead, J. Bierkens, M. Pollock, and G. O. Roberts. Piecewise deterministic Markov processes for continuous-time Monte Carlo. *Statist. Sci.*, 33(3):386–412, 2018. doi:10.1214/18-STS648.

[FI19]   Y. Feng and G. Iyer. Dissipation enhancement by mixing. *Nonlinearity*, 32(5):1810–1851, 2019. doi:10.1088/1361-6544/ab0e56.

[FKR06]   A. Fannjiang, A. Kiselev, and L. Ryzhik. Quenching of reaction by cellular flows. *Geom. Funct. Anal.*, 16(1):40–69, 2006. doi:10.1007/s00039-006-0554-y.

[Fur63]   H. Furstenberg. Noncommuting random products. *Transactions of the American Mathematical Society*, 108(3):377–428, 1963. URL http://www.jstor.org/stable/1993589.

[FW03]   A. Fannjiang and L. Wołowski. Noise induced dissipation in Lebesgue-measure preserving maps on d-dimensional torus. *J. Statist. Phys.*, 113(1-2):335–378, 2003. doi:10.1023/A:1025787124437.

[GCS+13] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 3 edition, 2013.

[GGZ22]   X. Gao, M. Gürbüzbalaban, and L. Zhu. Global convergence of stochastic gradient Hamiltonian Monte Carlo for nonconvex stochastic optimization: nonasymptotic performance bounds and momentum-based acceleration. *Oper. Res.*, 70(5):2931–2947, 2022. doi:10.1287/opre.2021.2162.

[GHKM21] N. E. Glatt-Holtz, J. A. Krometis, and C. F. Mondaini. On the accept-reject mechanism for metropolis-hastings algorithms, 2021, 2011.04493.

[GM16]   A. Guillin and P. Monmarché. Optimal linear drift for the speed of convergence of an hypoelliptic diffusion. *Electronic Communications in Probability*, 21(none):1 – 14, 2016. doi:10.1214/16-ECP25.

[Has70]   W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57(1):97–109, 1970. doi:10.1093/biomet/57.1.97.

[HHMS93] C.-R. Hwang, S.-Y. Hwang-Ma, and S.-J. Sheu. Accelerating gaussian diffusions. *The Annals of Applied Probability*, 3(3):897 – 913, 1993. doi:10.1214/aoap/1177005371.

[HS21]   X. Huang and C. D. Sogge. Weyl formulae for schrödinger operators with critically singular potentials. *Communications in Partial Differential Equations*, 46(11):2088–2133, 2021. doi:10.1080/03605302.2021.1925915.

[HWG+20] Y. Hu, X. Wang, X. Gao, M. Gurbuzbalaban, and L. Zhu. Non-convex optimization via non-reversible stochastic gradient langevin dynamics. *ArXiv e-prints*, 2020. doi:10.48550/ARXIV.2004.02823.

[IKX14]   G. Iyer, A. Kiselev, and X. Xu. Lower bounds on the mix norm of passive scalars advected by incompressible enstrophy-constrained flows. *Nonlinearity*, 27(5):973–985, 2014. doi:10.1088/0951-7715/27/5/973.

[ILN23]   G. Iyer, E. Lu, and J. Nolen. Using Bernoulli maps to accelerate mixing of a random walk on the torus, 2023, 2303.03528.

[IXZ21]   G. Iyer, X. Xu, and A. Zlatoš. Convection-induced singularity suppression in the Keller-Segel and other non-linear PDEs. *Trans. Amer. Math. Soc.*, 374(9):6039–6058, 2021. doi:10.1090/tran/8195.

[IZ22]   G. Iyer and H. Zhou. Quantifying the dissipation enhancement of cellular flows, 2022. doi:10.48550/ARXIV.2209.11645.

[JKO98]   R. Jordan, D. Kinderlehrer, and F. Otto. The variational formulation of the Fokker-Planck equation. *SIAM J. Math. Anal.*, 29(1):1–17, 1998. doi:10.1137/S0036141096303359.

[KH95]   A. Katok and B. Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1995. doi:10.1017/CBO9780511809187. With a supplementary chapter by Katok and Leonardo Mendoza.

[Led86]   F. Ledrappier. Positivity of the exponent for stationary sequences of matrices. In L. Arnold and V. Wihstutz, editors, *Lyapunov Exponents*, pages 56–73, Berlin, Heidelberg, 1986. Springer Berlin Heidelberg.

[Lie83]   E. H. Lieb. Sharp constants in the Hardy-Littlewood-Sobolev and related inequalities. *Ann. of Math. (2)*, 118(2):349–374, 1983. doi:10.2307/2007032.

[Lin82]   D. A. Lind. Dynamical properties of quasihyperbolic toral automorphisms. *Ergodic Theory and Dynamical Systems*, 2(1):49–68, 1982. doi:10.1017/S0143385700009573.

[Liv04]   C. Liverani. On contact Anosov flows. *Ann. of Math. (2)*, 159(3):1275–1312, 2004. doi:10.4007/annals.2004.159.1275.

[LLN19]   Y. Lu, J. Lu, and J. Nolen. Accelerating langevin sampling with birth-death, 2019, 1905.09863.

[LNP13]   T. Lelièvre, F. Nier, and G. A. Pavliotis. Optimal non-reversible linear drift for the convergence to equilibrium of a diffusion. *Journal of Statistical Physics*, 152(2):237–274, jun 2013. doi:10.1007/s10955-013-0769-x.

[LP17]   D. A. Levin and Y. Peres. *Markov chains and mixing times*. American Mathematical Society, Providence, RI, 2017. doi:10.1090/mbk/107. Second edition of [MR2466937], With contributions by Elizabeth L. Wilmer, With a chapter on "Coupling from the past" by James G. Propp and David B. Wilson.

[LPV15]   X. Li, E. Paquet, and H. L. Viktor. Molecular dynamics, Monte Carlo simulations, and Langevin dynamics: A computational review. *BioMed Research International*, 2015:183918, 2015. doi:10.1155/2015/183918.

[LS18]   J. Lu and K. Spiliopoulos. Analysis of multiscale integrators for multiple attractors and irreversible Langevin samplers. *Multiscale Model. Simul.*, 16(4):1859–1883, 2018. doi:10.1137/16M1083748.

[LSW22]   Y. Lu, D. Slepčev, and L. Wang. Birth-death dynamics for sampling: Global convergence, approximations and their asymptotics, 2022, 2211.00450.

[MCF15]   Y.-A. Ma, T. Chen, and E. Fox. A complete recipe for stochastic gradient MCMC. *Advances in neural information processing systems*, 28, 2015, 1506.04696.

[MRR$^+$53] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953. doi:10.1063/1.1699114.

[MS17]   G. Mazzola and S. Sorella. Accelerating ab initio molecular dynamics and probing the weak dispersive forces in dense liquid hydrogen. *Phys. Rev. Lett.*, 118:015703, Jan 2017. doi:10.1103/PhysRevLett.118.015703.

[MT06]   R. Montenegro and P. Tetali. Mathematical aspects of mixing times in Markov chains. *Found. Trends Theor. Comput. Sci.*, 1(3):x+121, 2006. doi:10.1561/0400000003.

[MT09]   S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009. doi:10.1017/CBO9780511626630. With a prologue by Peter W. Glynn.

[Nea96]   R. M. Neal. *Bayesian learning for neural networks*, volume 118. Springer Science & Business Media, 1996. doi:10.1007/978-1-4612-0745-0.

[Nov16] E. Novak. Some results on the complexity of numerical integration. In *Monte Carlo and quasi-Monte Carlo methods*, volume 163 of *Springer Proc. Math. Stat.*, pages 161–183. Springer, [Cham], 2016. doi:10.1007/978-3-319-33507-0_6.

[Pav14] G. A. Pavliotis. *Stochastic processes and applications*, volume 60 of *Texts in Applied Mathematics*. Springer, New York, 2014. doi:10.1007/978-1-4939-1323-7. Diffusion processes, the Fokker-Planck and Langevin equations.

[Pie94] R. T. Pierrehumbert. Tracer microstructure in the large-eddy dominated regime. *Chaos, Solitons & Fractals*, 4(6):1091–1110, 1994. doi:10.1016/0960-0779(94)90139-2.

[Pol85] M. Pollicott. On the rate of mixing of Axiom A flows. *Invent. Math.*, 81(3):413–426, 1985. doi:10.1007/BF01388579.

[QP04] D. Quigley and M. I. J. Probert. Langevin dynamics in constant pressure extended systems. *The Journal of Chemical Physics*, 120(24):11432–11441, jun 2004. doi:10.1063/1.1755657.

[Ras38] P. K. Rashevskii. E. cartan, geometry of riemannian spaces. *Uspekhi Matematicheskikh Nauk*, pages 266–268, 1938. URL https://www.mathnet.ru/rus/rm8969.

[Ray54] D. Ray. On spectra of second-order differential operators. *Trans. Amer. Math. Soc.*, 77:299–321, 1954. doi:10.2307/1990873.

[RBS15] L. Rey-Bellet and K. Spiliopoulos. Irreversible Langevin samplers and variance reduction: a large deviations approach. *Nonlinearity*, 28(7):2081, may 2015. doi:10.1088/0951-7715/28/7/2081.

[Ros61] H. H. Rosenbrock. An automatic method for finding the greatest or least value of a function. *Comput. J.*, 3:175–184, 1960/61. doi:10.1093/comjnl/3.3.175.

[RRT17] M. Raginsky, A. Rakhlin, and M. Telgarsky. Non-convex learning via stochastic gradient Langevin dynamics: a nonasymptotic analysis. In *Annual Conference Computational Learning Theory*, 2017. doi:10.48550/arXiv.1702.03849.

[RT96] G. O. Roberts and R. L. Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, 2(4):341–363, 1996. doi:10.2307/3318418.

[Sac22] Y. Sachkov. *Introduction to geometric control*, volume 192 of *Springer Optimization and Its Applications*. Springer, Cham, [2022] ©2022. doi:10.1007/978-3-031-02070-4. Translated from the Russian original.

[Sch80] Z. Schuss. Singular perturbation methods in stochastic differential equations of mathematical physics. *SIAM Rev.*, 22(2):119–155, 1980. doi:10.1137/1022024.

[Sei13] C. Seis. Maximal mixing by incompressible fluid flows. *Nonlinearity*, 26(12):3279–3289, 2013. doi:10.1088/0951-7715/26/12/3279.

[SM79] Z. Schuss and B. J. Matkowsky. The exit problem: a new approach to diffusion across potential barriers. *SIAM J. Appl. Math.*, 36(3):604–623, 1979. doi:10.1137/0136043.

[SOW06] R. Sturman, J. M. Ottino, and S. Wiggins. *The mathematical foundations of mixing*, volume 22 of *Cambridge Monographs on Applied and Computational Mathematics*. Cambridge University Press, Cambridge, 2006. doi:10.1017/CBO9780511618116. The linked twist map as a paradigm in applications: micro to macro, fluids to solids.

[Thi12] J.-L. Thiffeault. Using multiscale norms to quantify mixing and transport. *Nonlinearity*, 25(2):R1–R44, 2012. doi:10.1088/0951-7715/25/2/R1.

[TLF77] P. Turq, F. Lantelme, and H. L. Friedman. Brownian dynamics: Its application to ionic solutions. *The Journal of Chemical Physics*, 66(7):3039–3044, 1977, https://doi.org/10.1063/1.434317. doi:10.1063/1.434317.

[TZ23] M. Tsujii and Z. Zhang. Smooth mixing Anosov flows in dimension three are exponentially mixing. *Ann. of Math. (2)*, 197(1):65–158, 2023. doi:10.4007/annals.2023.197.1.2.

[Vil09] C. Villani. Hypocoercivity. *Mem. Amer. Math. Soc.*, 202(950):iv+141, 2009. doi:10.1090/S0065-9266-09-00567-5.

[VW19] S. Vempala and A. Wibisono. Rapid convergence of the unadjusted Langevin algorithm: Isoperimetry suffices. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. doi:10.48550/arXiv.1903.0856.

[Wib18] A. Wibisono. Sampling as optimization in the space of measures: The Langevin dynamics as a composite optimization problem. In S. Bubeck, V. Perchet, and P. Rigollet, editors, *Proceedings of the 31st Conference On Learning Theory*, volume 75 of *Proceedings of Machine Learning Research*, pages 2093–3027. PMLR, 06–09 Jul 2018. URL https://proceedings.mlr.press/v75/wibisono18a.html.

[YZ17]   Y. Yao and A. Zlatoš. Mixing and un-mixing by incompressible flows. *J. Eur. Math.
              Soc. (JEMS)*, 19(7):1911–1948, 2017. doi:10.4171/JEMS/709.

[Zla10]   A. Zlatoš. Diffusion in fluid flow: dissipation enhancement by flows in 2D. *Comm.
              Partial Differential Equations*, 35(3):496–534, 2010. doi:10.1080/03605300903362546.

DEPARTMENT OF MATHEMATICS, PENN STATE UNIVERSITY, STATE COLLEGE, PA 16803.

Email address: acc5843@psu.edu

SCHOOL OF MATHEMATICAL SCIENCES, SHANGHAI KEY LABORATORY OF PMMP, EAST CHINA
NORMAL UNIVERSITY, SHANGHAI, 200241, P.R. CHINA.
Email address: yyfeng@math.ecnu.edu.cn

DEPARTMENT OF MATHEMATICAL SCIENCES, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PA
15213.
Email address: gautam@math.cmu.edu

DEPARTMENT OF MATHEMATICS, PENN STATE UNIVERSITY, STATE COLLEGE, PA 16803.
Email address: novikov@psu.edu