Exponential Convergence for the Stochastically Forced Navier-Stokes Equations and Other Partially Dissipative Dynamics

Jonathan C. Mattingly*

March 1st, 2001 Revised 6/01, 12/01, 2/02

Abstract

We prove that the two dimensional Navier-Stokes equations possesses an exponentially attracting invariant measure. This result is in fact the consequence of a more general "Harris-like" ergodic theorem applicable to many dissipative stochastic PDEs and stochastic processes with memory. A simple iterated map example is also presented to help build intuition and showcase the central ideas in a less encumbered setting. To analyze the iterated map, a general "Doeblin-like" theorem is proven. One of the main features of this paper is the novel coupling construction used to examine the ergodic theory of the non-Markovian processes.

1 Introduction and Main Results

We are mainly concerned with the ergodic theory of stochastic processes with infinite dimensional phase spaces. Such systems are difficult to handle with the standard theory. To handle them we show that many estimates can be preformed in the context of a finite dimensional non-Markovian stochastic process. Hence, we also in this paper develop a fairly general ergodic theory for a certain class of non-Markovian stochastic processes.

The central structural assumption in this work is that the system has only a finite number of unstable directions. In other words, the unstable manifold is finite dimensional. We mainly have in mind stochastically forced PDEs where the symbol of the differential operator is dissipative. This holds for our primary expository example, the stochastically forced 2-D Navier Stokes equations (SNS).

$$\begin{cases} \dfrac{\partial u}{\partial t} + (u\cdot\nabla)u + \nabla p - \nu\Delta u = \sigma(u)\dfrac{\partial W(x,t)}{\partial t} \\ \nabla\cdot u = 0 \end{cases} \tag{1}$$

We will concentrate on this equation for definiteness; however, the ideas developed and many of our lemmas could be applied to a large number of equations such as: reaction diffusion

*Department of Mathematics, Stanford University, Stanford CA 94305

equations, the Cahn-Hilliard equations, or the Kuramoto-Sivashisky equations. In fact very little of the fine structure of the Navier-Stokes equations is used. Our methods could also be used to analysis certain classes of stochastic delay equations, as the central results are concerned with the ergodic theory stochastic systems with memory.

In developing our theory, we will also discuss a system generated by a simple iterated map. This is mainly for expository purposes. We use it to highlight the essence of our ideas while avoiding a number of technical complications which the SNS equations present. These complications are standard in the ergodic theory of Markov processes. They are essentially the difference between Doeblin's condition and Harris's condition (see [Dur96, Har56]). The first is extremely uniform while the second requires some care and extra estimates to deal with the lack of uniformity. The iterated map example will place us in the uniform "Doeblin-like" situation where our central ideas are less obscured by other technicalities.

This paper gives the proofs of the exponential convergence results mentioned in [EMS01] and referred to in [BKL01] as being similar to theirs. In fact, it is an expanded and clarified version of the coupling section from an earlier circulated draft of [EMS01]. At first glance, that article takes a very different point of view from this one. [EMS01] quickly reduces the equation to a finite dimensional process with memory and then rests in that world for the remainder of the discussion. In this paper, we prefer to stay in the Markov setting, as we are interested in the convergence to the invariant measure from an initial state. Answering this question seems more natural in the Markovian setting. However both points of view were developed in tandem and both have been useful in improving our understanding.

In equation (1), we will interpret $\frac{\partial W(x,t)}{\partial t}$ in the sense of Itô. For simplicity of presentation, we will take $\sigma(u)dW$ to be of the form

$$\sigma(u)dW(x,t) = \sum_{k\in\mathcal{W}} \sigma_k(u)e_k(x)dw_k(t,\omega) \tag{2}$$

where the w_k's are standard i.i.d complex valued Wiener process satisfying $w_{-k}(t) = w_k(t)$ and $\{e_k(x) = \binom{-ik_2}{ik_1}\frac{e^{ik\cdot x}}{|k|}, k \in \mathbb{Z}^2\}$ are the basis in the space of L^2 divergence-free, mean zero vector fields on $\mathbb{T}^2$, the two dimensional torus. $\mathcal{W}$ is a finite subset of $\mathbb{Z}^2$ specifying the modes which are forced. The σ_k are complex valued functions. Further assumptions on the σ_k are necessary and will be outlined below.

Generally, we restrict to the case where only a finite number of modes are forced. Parts of our analysis cover the case when all modes are forced and the case with finitely many forced modes is within reach of our techniques; however, it presents extra complications which we do not tackle in this article. We do give a general outline of this case in the map example.

To give the conditions on the σ_k, we define $\mathbb{L}^2_{\mathcal{W}} = \text{span}\{e_k : k \in \mathcal{W}\}$. Notice that $\mathbb{L}^2_{\mathcal{W}}$ is finite dimensional. We assume that each σ_k is a smooth function from $\mathbb{L}^2_{\mathcal{W}} \to \mathbb{C}$ such that there exist positive numbers σ_k^{max} and σ_k^{min} with $\sigma_k^{min} \le |\sigma_k| \le \sigma_k^{max}$. The requirement that the σ_k be bounded from below is critical to the current approach. The requirement that they be bounded from above provides simple estimates. It is reasonable to expect that many unbounded functions could be analyzed. The requirement that the lower σ_k only depend on the forced modes is an artifact of how we prove Lemma 4.5 and could certainly be removed (as long as only a finite number of modes are forced).

Define

$$|\mathcal{W}|_{int} = \sup \{ N : k \in \mathcal{W} \text{ for all } k \text{ with } 0 < |k| \leq N \} ,$$

and $\mathcal{E}_0^* = \sum (\sigma_k^{max})^2$. Of course $(0,0) \notin \mathcal{W}$ since we allow no mean flow. Letting $P_t(u_0, \cdot)$ denote the measure induced at time t by the dynamics starting from u_0, we will prove the following theorem:

Theorem 1. *In the above setting, there is an absolute positive constant $\mathcal{C}$ (depending only on the domain) so that if $|\mathcal{W}|_{int}^2 > \mathcal{C}\frac{\mathcal{E}_0^*}{\nu^3}$ then following holds. There exist positive constants K and γ so that for any initial conditions $u_0^{(1)}$ and $u_0^{(2)}$*

$$\|P_t(u_0^{(1)}, \cdot) - P_t(u_0^{(2)}, \cdot)\|_* \leq K\Big[1 + \Big|u_0^{(1)}\Big|_{\mathbb{L}^2}^2 + \Big|u_0^{(2)}\Big|_{\mathbb{L}^2}^2\Big] e^{-\gamma t} .$$

The exact setting and mode of convergence will be described in section 2.4. For now let us simply say that the $\| \cdot \|_*$ is a norm on measure which interpolates between the total variation norm and the Wasserstein norm. Hence, it is a complete metric on the space of probability measures. The $| \ |_{\mathbb{L}^2}$ norm will be defined in section 4, but it is essentially the L^2 norm in the space variables.

We do not claim that the map examples we consider are of any relevance other than as illustrative examples. A typical example of the type of iterated random map we will consider is $\phi(\omega) : [0,1] \times l_1^2 \to [0,1] \times l_1^2$. Here l_1^2 is the unit ball in the space of square summable real sequences and ϕ is the map

$$\phi_\omega : \begin{pmatrix} z \\ y \end{pmatrix} \longmapsto \begin{pmatrix} e^{z^2} \cdot |y * y * y|^2 + (1 + z|y|^2) \cdot W(\omega) \qquad (\mod 1 \ ) \\ \frac{z^2}{6} \cdot (y * y) + \frac{\sin(2\pi z)^2}{4} \cdot B(\omega) \end{pmatrix} , \tag{3}$$

where W is a Gaussian random variable, B is any l_1^2 valued random variable independent of W, and "$*$" denotes the convolution on l_1^2. The exact form of this map is unimportant. The point is that it is a map of an infinite dimensional phase space with the property that the map has a smooth density on some finite dimensional part of the space (namely $[0,1]$) and is contractive in a way to be made clear in the sequel on the remaining part of the phase space (namely l_1^2).

The first ergodic results for equation (1) were proved in [FM95]. However, the results there required forcing which was extremely rough in the space variable. In particular, the forcing could not be differentiable in the space variable and all of the eigen functions e_k had to be forced. In [Mat99], ergodicity and stronger results were proved without any assumptions on the fine structure of the forcing but only assumptions on the balance between energy dissipation and energy influx. In many ways, the results in this paper can be seen as applying the ideas from [Mat99] to one part of the phase space and ideas from classical probability theory to the remainder.

Recently there have been a number of papers on the subject of ergodicity of stochastically forced PDEs and the 2D Navier Stokes equation in particular. Others working independently of the author, have announced similar results to those in this note. In [KS00] ergodicity was proved for the two dimensional Navier Stokes equation forced by an impulsive forcing.

Recently they have also announced that they can demonstrate convergence of the transition kernels to this unique invariant measure [KS]. The abstract theorems contained in this paper can also be used to prove exponential convergence of the 2D Navier Stokes equation with impulsive forcing. In [BKL01], results similar to ours for the Navier Stokes equation were proven independently. The techniques are more analytical in nature and the control by the initial data less explicit. In this paper, we take a more probabilistic approach than [KS00, KS, BKL01]. In addition the convergence to the invariant measure is slightly stronger than those obtained the the other works. We utilize the ideas and imagery of coupling which are novel in this context. We believe that coupling provides a powerful, extremely adaptive tool for proving ergodicity in this setting. Although there are many differences between the three approaches, at the core there are a number of similarities. This is not surprising as they all reflect the same underlying physics. It will be interesting to see what common theory distills out of these different treatments, pulling the best features from each.

The theory presented in this paper could be described as "effectively elliptic" in that all of the important coordinates are forced directly. Some steps towards a hypo-elliptic theory have been made. In [EH], a reaction diffusion system with all but a finite number of modes was examined. This equation is much more uniform than the Navier Stokes equation, yet their results are suggestive. More importantly the analysis in [EH] addresses issues orthogonal to those in this paper; hence, they should be understood as complimentary. In [EM01], it was shown that arbitrary finite dimensional projections of the Navier Stokes equations were ergodic under quite general assumptions. The approach presented here is well suited to the hypo-elliptic setting; however, additional estimates are needed.

Lastly, we show in the map setting how the coupling construction can produce a factorization of the invariant measure when modes other than the "active variables" are forced. This makes clear the connection to [Mat99] mentioned previously. In that paper, each realization of forcing was in one to one correspondence to a stationary solution. This simple structure was because the dissipative linear term was taken strong enough to ensure that there were no active variables. If we allow there to be active variables, then for each realization of the forcing on the "enslaved" variables we obtain an attracting invariant measure on the active variables. (See the next section for a precise definition of "active" and "enslaved" variables.) Hence, we see that the invariant measures can be disintegrated over each fiber of the forcing on the enslaved variables. We only scratch the surface of this property. It is clear that an interesting theory can be developed exploiting the skew flow structure of the enslaved modes viewed as a random dynamical system driven by the active modes..

2 Abstract Setting

Let $\mathbb{X}$ be a Banach space and $\mathcal{B}$ the associated Börel σ-algebra. Let P be a Markov transition kernel on $\mathbb{X}$. Let $\phi(\omega) : \mathbb{X} \to \mathbb{X}$ be a random map, built on the probability space $(\Omega, \mathbb{P}, \mathcal{F})$, which realizes the Markov chain associated with P. That is to say, $\mathbb{P}\{\phi(\omega)x \in B\} = P(x, B)$ for $B \in \mathcal{B}$.

We wish to compare the measure obtained by evolving forward in time from two different initial conditions $x_0^{(1)}$ and $x_0^{(2)}$ namely $P_t(x_0^{(1)}, \cdot)$ and $P_t(x_0^{(2)}, \cdot)$. To accomplish this, we will build a measure on the product space $\mathbb{X} \times \mathbb{X}$ such that its marginals agree with $P_t(x_0^{(1)}, \cdot)$ and $P_t(x_0^{(2)}, \cdot)$ and which possesses properties useful for our analysis. Such a construction is usually referred to as a coupling. The idea is construct to a coupling which makes explicit the shared component of $P_t(x_0^{(1)}, \cdot)$ and $P_t(x_0^{(2)}, \cdot)$. Usually one constructs a coupling which ensures that at some random moment of time in the future the two Markov process are at the same position in phase space. This random time is referred to as the "coupling time." Since the two processes agree at the coupling time, the Markov property ensures that the two processes have identical distributions at any time after the coupling time.

Since the phase space is infinite dimensional, we will not be able to construct a coupling which ensures that the two processes agree in all coordinates at some random time. Instead, we will change the measure so that asymptotically as $t \to \infty$ the trajectories will converge to each other in a particular norm.

Though the theorems in this note apply to this general setting; all of the the examples we will present have a shared structure. We will construct a coupling which ensures that some finite subset of the coordinates agree for all times in the future after some moment of time. We emphasis that we will not construct a coupling which insures that all of the coordinates are ever equal. Instead, we only ensure that a finite number of coordinates agree at some moment after the coupling time; however, the construction will guarantee that they agree for all future times. This will in turn imply that the remaining degrees of freedom converge to each other as $t \to \infty$. Both facts are linked to the special structure of the systems we consider. These systems possess some subset of the coordinates which we will refer to as the "active variables" with the following properties. If the active variables agree from time 0 to t, then the distance between the remaining coordinates is reduced by a factor exponentially small in t over the interval of time. Hence if the "active variables" agree for all future times, the remaining coordinates converge to each other. This also gives an indication of why it is possible to make the active variables agree for all future time with some positive probability. The longer the active variables agree the closer the remaining degrees of freedom are to each other. Hence if the transition probabilities possess any smoothness relative to the initial states then it will become easier to continue coupling the active variables the longer they agree. This feedback is the critical element of the systems we consider. More general constructions will be useful to handle other situations such as the hypo-elliptic cases.

In general the "active variable" refers to some coordinates on which we will change the measure. The "enslaved variables" can then be determined from some initial condition and the trajectories of the active variables. The enslaved variables can be viewed as a random dynamical system over a probability space given by the active variables. However with the added complication that the measure on the active variables depends on the state of the whole system; not just the active variables as is the normal case.

Consider the following two examples. The first illustrates the difficulty in making two processes agree for all time. The second illustrate the type of feedback all of our systems have. Let B_t, $\tilde{W}_t$, and $\tilde{W}_t$ be standard Brownian Motion on $\mathbb{R}$ and set

$$\begin{aligned} dX_t &= -X_t dt + dW_t & \qquad dY_t &= -Y_t dt + e^{-t} dt + d\tilde{W}_t \\ X_0 &= x_0 & Y_0 &= x_0 \ . \end{aligned}$$

A straightforward application of Girsanov's theorem implies that for any finite T the measures induced by B_t, X_t, and Y_t on the path space $C\big([0,T];\mathbb{R}\big)$ are mutually absolutely continuous. Hence, it is possible to build a coupling construction with a positive chance that $B_t = X_t = Y_t$ for all $t \in [0,T]$. However as T increases the Radon-Nikodym derivatives between the densities induced by X_t and B_t or between the densities induced by Y_t and B_t be come increasingly large. This is because the density induced by B_t on $C\big([0,\infty);\mathbb{R}\big)$ is singular relative to the other two. But X_t and Y_t are absolutely continuous relative to each other on $C\big([0,\infty);\mathbb{R}\big)$. Hence, it would be possible to couple them for all times.

For a caricature of the feedback mechanism, consider the following two systems of Itô stochastic differential equations

$$\begin{aligned} dZ_t &= -Z_t dt + F_t dt + dW_t & \qquad d\tilde{Z}_t &= -\tilde{Z}_t dt + \tilde{F}_t dt + d\tilde{W}_t \\ \frac{dF}{dt} &= -F_t + Z_t & \qquad \frac{d\tilde{F}}{dt} &= -\tilde{F}_t + \tilde{Z}_t \\ Z_0 &= z_0 \qquad F_0 = f_0 & \qquad \tilde{Z}_0 &= z_0 \qquad \tilde{F}_0 = \tilde{f}_0 \end{aligned}$$

where we assume that $f_0 \neq \tilde{f}_0$. A straight forward calculation shows that if $Z_s = \tilde{Z}_s$ for $s \in [0,t]$ then $|F_t - \tilde{F}_t| = |f_0 - \tilde{f}_0|e^{-t}$. Hence, asymptotically in time, the F_t forgets the initial condition, becoming completely enslaved to the Z motion. Again Girsanov's theorem implies that the measures induced by Z_t and $\tilde{Z}_t$ on the path space $C\big([0,T];\mathbb{R}\big)$ are mutually absolutely continuous. Consider $Z_t = \tilde{Z}_t$ for $t \in [0,T]$. Then we are interested in the percentage of the time, under the most efficient coupling possible, that $Z_t = \tilde{Z}_t$ for $t \in [T, T+1]$. If Q_T and $\tilde{Q}_T$ represent the measure induced on $C\big([T,T+1],\mathbb{R}\big)$ by Z_t and $\tilde{Z}_t$ respectively, then Girsanov's theorem followed by the Cauchy-Schwartz inequality implies that

$$|Q_T - \tilde{Q}_T|_{TV} = \mathbb{E}\left|1 - \frac{dQ_T}{d\tilde{Q}_T}\right| \leq \mathbb{E}\left(\frac{dQ_T}{d\tilde{Q}_T}\right)^2 - 1 = \exp\left\{\int_T^{T+1} \frac{1}{2}|F_t - \tilde{F}_t|^2 dt\right\} - 1$$

where $|\cdot|_{TV}$ is the total variational norm. If we assume that $Z_t = \tilde{Z}_t$ for $t \in [0,T]$ then

$$|Q_T - \tilde{Q}_T|_{TV} \leq \exp\left(\frac{1}{2}e^{-T}\right) - 1 \ .$$

The total variational distance between the two probability measures can be interpreted as the minimum chance, over all couplings, that two random variables distributed as the measures are not equal. Hence from the above calculation, we see we should be able to construct a coupling so that the longer the Z agree the more likely they are to agree at the next moment of time.

In the next few sections, we develop these ideas in an abstract setting. In the above discussion, we have assumed that the coupling was constructed so the active modes eventually were equal for all future times. This will not be true in all of the settings to which we intend to apply the abstract theory. We wish to allow the possibility that the active coordinates simply become more and more similar, yet this is enough to entrain the enslaved variables. This will be useful in the case when all of the active modes are not forced. In this "hypo-elliptic" setting, it is impossible to make the active modes equal at all times.

2.1 An Enslaving Splitting

In light of the discussion in the last section, we assume that $\mathbb{X} = \mathbb{X}_+ \times \mathbb{X}_-$. We will refer to $\mathbb{X}_+$ as the active space and $\mathbb{X}_-$ as the enslaved space. Furthermore, we assume the existence of a function $\Phi : \mathbb{X}_+ \times \mathbb{X} \to \mathbb{X}_-$ such that if $x_1 = \phi(\omega)x_0$ with $x_i = (z_i, y_i) \in \mathbb{X}_+ \times \mathbb{X}_-$ then $y_1 = \Phi(z_1, x_0)$. For $z \in \mathbb{X}_+^k$, we will used $\Upsilon(z, x_0)$ to denote the $x = (z, y) \in \mathbb{X}_+^k \times \mathbb{X}_-^k = \mathbb{X}^k$ where y is obtained by $y = \Phi(z, x_0)$. In the context of the example at the end of the last section, Φ would reconstruct the F_t given the initial condition of F and the information $Z_s, s \in [0, T]$ and Υ returns the pair (F_t, Z_t).

Our terminology is reasonable because given a trajectory $z_1, \ldots, z_n$ and an initial y_0, we can reconstruct the associated $y_2, \ldots, y_n$. Φ can be viewed as a random dynamical system over the space $\mathbb{X}_+$ with the caveat that the distribution on $\mathbb{X}_+$ used to drive Φ and obtain a path in $\mathbb{X}_-$ depends on both the current value in $\mathbb{X}_+$ and $\mathbb{X}_-$. Our assumption that Φ is deterministic means that any knowledge of the randomness in a given step must be transmitted through z. In section 3.1, we will relax this condition. In the simple map setting Φ needs only be a function of the $x_0 \in \mathbb{X}$. The "look ahead" allowed by the inclusion of $z_1 \in \mathbb{X}_+$ is useful in the continuous in time setting.

We extend the definition of Φ and Υ to sequences in $\mathbb{X}_+$ as follows: If $z = (z_1, \ldots, z_k) \in \mathbb{X}_+^k$ then $y = (y_1, \ldots, y_k) = \Phi(z, x_0) \in \mathbb{X}_-^k$ means $y_j = \Phi(z_j, x_{j-1})$ where $x_i = (z_i, y_i)$. We will use " : " to denote concatenation. Namely, if $z = (z_1, \ldots, z_k) \in \mathbb{X}_+^k$ and $\tilde{z} \in (z_1, \ldots, z_m) \in \mathbb{X}_+^m$ then $z : \tilde{z} = (z_1, \ldots, z_k, \tilde{z}_1, \ldots \tilde{z}_m) \in \mathbb{X}_+^{k+m}$. We will denote by Π_+ and Π_- the projections onto $\mathbb{X}_+$ and $\mathbb{X}_-$ respectively. We also extend all of our maps to sets of vectors in the natural way. Hence if $Z \subset \mathbb{X}_+^k$ and $x \in \mathbb{X}$ then

$$x_0 : \Upsilon(Z, x_0) = \big\{(x_0, x_1, \ldots x_k) \in \mathbb{X}^{k+1} : \exists z \in Z \text{ so that } (x_1, \ldots, x_k) = \Upsilon(z, x_0)\big\} \ .$$

We will denote by $\pi_k x$ the projection onto the k-th component of a vector $x \in \mathbb{X}^m$ where $m \geq k$. Similarly we will denote by $\pi_{[n,k]} x$ as the vector of length $k-n+1$ obtained by taking the nth through kth element of x. At times we will not want to mention explicitly the length of a vector but wish to obtain remainder after the first n elements are removed. If $x \in \mathbb{X}^m$ and $n \leq k \leq m$, we will write $\pi_{[n,k]}^{\perp} x$ for the remaining $m-(k-n+1)$ elements of x after the elements n through k have been removed. For example $\pi_{[2,k]}^{\perp}(x_1, x_2, \ldots, x_k, x_{k+1}, \ldots, x_m) = (x_1, x_{k+1}, \ldots, x_m)$.

Let $P^n(x, \cdot)$ be the measure induced on $\mathbb{X}$ by the chain after n steps starting from the initial condition x. Let $Q^n(x, \cdot)$ be the measure induced on the path space $\mathbb{X}^n$ by taking n steps of the chain starting from x. We emphasize that $Q^n(x, \cdot)$ is a measure on the entire path space from 1 to n. Let $Q_+^n(x, \cdot)$ be the projection of $Q^n(x, \cdot)$ onto $\mathbb{X}_+^n$ and $\mathcal{B}_+$ the sigma algebra generated by the projection of $\mathcal{B}$ on to $\mathbb{X}_+$. In general, if $U(x, \cdot)$ is a measure on the paths in $\mathbb{X}_+^n$ and $\hat{U}(x, \cdot)$ is a measure on the path space $\mathbb{X}_+^m$ for $x \in \mathbb{X}$, then $(\hat{U}U)(x, \cdot)$ will denote the measure on the path space $\mathbb{X}_+^{n+m}$ constructed as follows. For any $A \subset \mathbb{X}_+^{n+m}$

$$(\hat{U}U)(x, A) = \int_{\pi_{[1,n]} A} U\big(x, dz\big) \hat{U}\big(\pi_n \Upsilon(z, x), \pi_{[1,n]}^{-1}(z) \cap A\big)$$

where $\pi_{[1,n]}^{-1}(z)$ is viewed as a subset of $\mathbb{X}_+^{n+m}$. If the U and $\hat{U}$ are normalized to be probability measures then $\hat{U}U$ is the distribution of the $\mathbb{X}_+^{n+m}$-valued random variable $z' = z : \hat{z}$

obtained by first drawing $z = (z_1, \cdots, z_n)$ from $\mathbb{X}_+^n$ according to $U(x, \cdot)$. Then drawing $\hat{z} = (\hat{z}_1, \cdots, \hat{z}_m)$ from $\mathbb{X}_+^m$ according to $\hat{U}(x_n, \cdot)$ where x_n is the random variable $\pi_n \Upsilon(z, x)$. We extend all of our definitions to measures on the product spaces$\mathbb{X}_+^k \times \mathbb{X}_+^k$ in the natural way.

Lastly, given two measure $\hat{\mathbf{U}}$ and $\mathbf{U}$ on some product space, we will write $\mathbf{U} \overset{\text{M}}{\sim} \hat{\mathbf{U}}$ if the two measures have the same marginals. For instance if $\hat{\mathbf{U}}$ and $\mathbf{U}$ are measures on $\mathbb{X}^k \times \mathbb{X}^k$ then $\mathbf{U} \overset{\text{M}}{\sim} \hat{\mathbf{U}}$ means that $\mathbf{U}(A \times \mathbb{X}^k) = \hat{\mathbf{U}}(A \times \mathbb{X}^k)$ and $\mathbf{U}(\mathbb{X}^k \times A) = \hat{\mathbf{U}}(\mathbb{X}^k \times A)$ for all measurable A.

2.2 Basic Assumptions on the Coupling and Convergence

We begin with an overview of the central construction of this paper. We will build a coupling which makes certain features of the dynamics transparent. The construction requires has three basic components. First, a default coupling $\mathbf{Q}_+$ which will be used when nothing better can be done. It is usually taken to be $Q_+ \times Q_+$; that is uncorrelated motion. Secondly, for a distinguished subset of initial conditions, called $\mathbf{C}$, a factorization of the dynamics into a family of measures $\mathbf{s}_k$ and $\mathbf{r}_k$. This factorization has the property that if segment are drawn from the support of $\mathbf{s}_k$ then the trajectories of the whole system will converge to each other when viewed under the norm described momentarily. The third ingredient is this norm which is adapted to the factorization of the dynamics into $\mathbf{s}$ and $\mathbf{r}$.

To help build intuition, it is often the case that the $\mathbf{s}$ measure contains the parts of the dynamics which leads to similar behavior of the two trajectories and $\mathbf{r}$ is the remainder left over when the similar part has been removed. However, strictly speaking one only needs a factorization which satisfies the assumptions below.

More precisely we assume that we are provided with a probability transition kernel $\mathbf{Q}_+ : (\mathbb{X} \times \mathbb{X}) \times \boldsymbol{\mathcal{B}_+} \to [0, 1]$ where $\boldsymbol{\mathcal{B}_+}$ is the product σ-algebra on $\mathbb{X}_+ \times \mathbb{X}_+$ generated by the elements of $\mathcal{B}_+ \times \mathcal{B}_+$. We make the following basic assumption on $\mathbf{Q}_+$:

A 1: $\mathbf{Q}_+$ has the correct marginals: *For all $x, x' \in \mathbb{X}$, $\mathbf{Q}_+(x', x, \cdot) \overset{M}{\sim} Q_+(x', \cdot) \times Q_+(x, \cdot)$.*

This is the absolute minimal assumption on $\mathbf{Q}_+$. In all cases in this paper $\mathbf{Q}_+$ will be taken to be $Q_+ \times Q_+$ but one can imagine cases where another choice might be advantages.

Convergence will be measured relative to a collection of test functions $\mathcal{G}$. We will always assume that $\mathcal{G}$ only contains measurable functions $\mathsf{g} : \mathbb{X} \to \mathbb{R}$ with $\|\mathsf{g}\|_\infty = \sup |\mathsf{g}(x)| \leq 1$ for all $\mathsf{g} \in \mathcal{G}$. Given any signed measure μ on $\mathbb{X}$, we define

$$\|\mu\|_{\mathcal{G}} = \sup_{\mathsf{g} \in \mathcal{G}} \int_{\mathbb{X}} \mathsf{g}(x) d\mu(x) \ . \tag{4}$$

Once we describe the factorization, we will give further conditions on the set $\mathcal{G}$.

As mentioned, we will not always have a factorization for all initial conditions. Rather, we assume that we are provided with a measurable subset $\mathbf{C} \subset \mathbb{X} \times \mathbb{X}$ and that we will only be asked to provide an admissible factoring for initial conditions in $\mathbf{C}$.

A2: Uniformly over $\mathbf{C}$, there is some "similar" motion induced by active transition densities: *For any $k = 1, 2, \cdots$ or ∞ there exist sub-probability kernels $\mathbf{r}_k$ and $\mathbf{s}_k$ so that: (A sub-probability kernel simply means that they are normalized to a number less than or equal to one.)*

i) The two measures factor $\mathbf{Q}_+$ for initial conditions in $\mathbf{C}$. *If* $(x^{(1)}, x^{(2)}) \in \mathbb{X}^k \times \mathbb{X}^k$, *with* $x^{(i)} = (x_0^{(i)}, \ldots, x_{k-1}^{(i)})$ *then*

$$\mathbf{s}_k(x^{(1)}, x^{(2)}, \cdot\,) + \mathbf{r}_k(x^{(1)}, x^{(2)}, \cdot\,) \overset{M}{\sim} \mathbf{Q}_+(\pi_{k-1}x^{(1)}, \pi_{k-1}x^{(2)}, \cdot\,) \ .$$

ii) The factorization is symmetric in total mass. $\mathbf{s}_k(x^{(1)}, x^{(2)}, \mathbb{X}_+ \times \mathbb{X}_+) = \mathbf{s}_k(x^{(2)}, x^{(1)}, \mathbb{X}_+ \times \mathbb{X}_+)$

iii) The ratio of mass in the factors is uniformly bounded over initial conditions in $\mathbf{C}$. *There is a positive constant* ρ_k^* *so when* $(x^{(1)}, x^{(2)}) \in \mathbf{C}$, $\mathbf{s}^k(x^{(1)}, x^{(2)}, \mathbb{X}_+^k \times \mathbb{X}_+^k) \geq \rho_*^k > 0$.

Here $\mathbf{s}^k(x^{(1)}, x^{(2)}, \cdot\,)$ *is the measure on* $\mathbb{X}_+^k \times \mathbb{X}_+^k$ *obtained by weighting the first step according to* $\mathbf{s}_1$*, the second step according to* $\mathbf{s}_2$ *taking into account the first step, and so on. Symbolically,* $\mathbf{s}^k = \mathbf{s}_k \cdots \mathbf{s}_1$*.*

In the above, $\mathbf{s}^k = \mathbf{s}_k \cdots \mathbf{s}_1$ is a slight generalization of the definition given previously since $\mathbf{s}_k$ takes argument from $\mathbb{X}^k$. However, the idea is the same. Given initial conditions $(x_0^{(1)}, x_0^{(2)}) \in \mathbb{X} \times \mathbb{X}$, we first draw a $(z_1^{(1)}, z_1^{(2)})$ from $\mathbf{s}_1(x_0^{(1)}, x_0^{(2)}, \cdot\,)$. Then we use Υ to obtain $x_1^{(i)}$. Next we feed the $(x_0^{(i)}, x_1^{(i)})$ into $\mathbf{s}_2$ and draw $(z_2^{(1)}, z_2^{(1)})$ from which we reconstruct $(x_2^{(1)}, x_2^{(2)})$. Continuing in this manor, we see that $\mathbf{s}^k = \mathbf{s}_k \cdots \mathbf{s}_1$ is a well defined measure on $\mathbb{X}_+^k \times \mathbb{X}_+^k$ taking arguments $(x_0^{(1)}, x_0^{(2)}) \in \mathbb{X} \times \mathbb{X}$.

The $\mathbf{s}$ measures are our replacement for the active modes being exactly equal in a classical coupling argument. Our construction is only useful if $\mathbf{s}^\infty$ is supported on pairs of trajectories which, if followed, determine the asymptotic behavior of entire system. This is the content of the next assumption which describes the relation between the factorization and the mode of convergence given by $\| \ \|_{\mathcal{G}}$. We now state the remaining assumptions on $\mathcal{G}$.

A3: If the active modes are drawn from $\mathbf{s}^\infty$ the trajectories approach each other over time as seen through test function which induce the $\mathcal{G}$-distance. *There exists a fixed function* $\mathsf{G} : \mathbb{Z} \to [0, \infty)$ *and, for every* $(x_0^{(1)}, x_0^{(2)}) \in \mathbf{C}$*, a set* $Z(x_0^{(1)}, x_0^{(2)}) \subset \mathbb{X}_+^\infty \times \mathbb{X}_+^\infty$ *with the following properties:*

i) $Z(x_0^{(1)}, x_0^{(2)})$ has full measure: $\mathbf{s}^\infty(x_0^{(1)}, x_0^{(2)}, Z) = \mathbf{s}^\infty(x_0^{(1)}, x_0^{(2)}, \mathbb{X}_+^\infty \times \mathbb{X}_+^\infty)$.

ii) Decreases to zero: $\mathsf{G}(n)$ *is monotone decreasing with* $\mathsf{G}(n) \to 0$ *as* $n \to \infty$.

iii) Drawing from $\mathbf{s}^\infty$ causes the trajectories to converge in the $\mathcal{G}$-norm as $t \to \infty$: *If* $(z^{(1)}, z^{(2)}) \in Z$ *with* $z^{(i)} = (z_1^{(i)}, z_2^{(i)}, \cdots)$ *and* $\chi^{(i)} = \Upsilon(z^{(i)}, x_0^{(i)})$ *then for any* $\mathsf{g} \in \mathcal{G}$ *we have* $|\mathsf{g}(\chi_n^{(1)}) - \mathsf{g}(\chi_n^{(2)})| \leq \mathsf{G}(n)$.

It is now clear why we built our coupling as we did. It was constructed with the hope that eventually the coupled process would draw its infinite z futures from $\mathbf{s}^\infty$. Assumption A3 then guarantees that the x's obtained from these z will converge to zero.

For this to happen the system has to regularly find itself in the set $\mathbf{C}$. This is the content of the next assumption which is the last of our basic assumptions.

A 4: The set $\mathbf{C}$ is visited infinitely often: *For any* $(x^{(1)}, x^{(2)}) \in \mathbb{X} \times \mathbb{X}$*, there exists a measurable set* $A \subset \mathbb{X}_+^\infty \times \mathbb{X}_+^\infty$ *of full* $\mathbf{Q}_+^\infty$*-measure so that for each* $(f^{(1)}, f^{(2)}) \in A$*, where* $f^{(i)} = (f_1^{(i)}, \ldots, f_n^{(i)}, \ldots)$*, there exists a strictly increasing sequence* $\{t_k\}_{k=1}^\infty$ *with* $(f_{t_k}^{(1)}, f_{t_k}^{(2)}) \in \mathbf{C}$.

2.3 The Coupling Construction

We now use the assumptions in the previous section to construct a specific representation of two copies of the Markov chain built on a common probability space. This coupling will then be used in the following sections to prove ergodic theorems.

2.3.1 Setup and Basic Features

In preparation, define $\rho^k(x^{(1)},x^{(2)}) = \mathbf{s}^k(x^{(1)},x^{(2)},\mathbb{X}_+^k\times\mathbb{X}_+^k)$ and the probability measures

$$\mathbf{S}^k(x^{(1)},x^{(2)},\,\cdot\,) = \frac{\mathbf{s}^k(x^{(1)},x^{(2)},\,\cdot\,)}{\rho^k(x^{(1)},x^{(2)})} \qquad \mathbf{R}^k(x^{(1)},x^{(2)},\,\cdot\,) = \frac{\mathbf{r}^k(x^{(1)},x^{(2)},\,\cdot\,)}{1-\rho^k(x^{(1)},x^{(2)})}$$
$$\mathbf{S}_k(x^{(1)},x^{(2)},\,\cdot\,) = \frac{\mathbf{s}_k(x^{(1)},x^{(2)},\,\cdot\,)}{\mathbf{s}_k(x^{(1)},x^{(2)},\mathbb{X}_+\times\mathbb{X}_+)} \qquad \mathbf{R}_k(x^{(1)},x^{(2)},\,\cdot\,) = \frac{\mathbf{r}_k(x^{(1)},x^{(2)},\,\cdot\,)}{\mathbf{r}_k(x^{(1)},x^{(2)},\mathbb{X}_+\times\mathbb{X}_+)}\,.$$

If $\rho^k=0$ then we set the corresponding $\mathbf{S}^k$ to the null measure. Clearly by the construction of the $\mathbf{s}^k$, $\rho^k(x^{(1)},x^{(2)}) > \rho^{k+1}(x^{(1)},x^{(2)})$. Since $\rho^k\in[0,1]$, notice that they form a partition of the unit interval. Lastly we set $\mathbf{U}^n(x^{(1)},x^{(2)},\,\cdot\,) = (\mathbf{R}_n\mathbf{S}^{n-1})(x^{(1)},x^{(2)},\,\cdot\,)$ for $n>1$, $\mathbf{U}^1=\mathbf{R}^1$, and $\rho^0=1$. Recall that, as defined at the start of section 2.1, $\mathbf{R}_n\mathbf{S}^{n-1}$ represents the measure on $\mathbb{X}_+^n\times\mathbb{X}_+^n$ obtained by first taking $n-1$ steps according to $\mathbf{S}^{n-1}$, updating the initial condition and then taking one step according to $\mathbf{R}_n$. In this setting, we have the following simple but central lemma.

The following lemma is a fundamental in our investigation. It provides a factorization of the future trajectories into a “similar” part and a “remainder” part. We find it useful to think of $\mathbf{S}^k(x^{(1)},x^{(2)},\,\cdot\,)$ as concentrated along the diagonal in $\mathbb{X}_+\times\mathbb{X}_+$. Strictly speaking, our assumptions do not guarantee this. However, in all of the simple cases in this paper, this will be true and in the more complicated cases, such as the SNS, it is essentially true in a way that will be made clear later. However, we will only use the properties assumed in the assumptions.

The following lemma is fundamental to this paper. It gives the factorization of the infinite future.

Lemma 2.1 (Future Factoring Lemma). *Consider the setting of Assumption 2. For* $(x^{(1)},x^{(2)})\in\mathbf{C}$,

$$\mathbf{Q}_+^\infty(x^{(1)},x^{(2)},\,\cdot\,) \overset{M}{\sim} \rho^\infty(x^{(1)},x^{(2)})\mathbf{S}^\infty(x^{(1)},x^{(2)},\,\cdot\,) + \sum_{k=1}^{\infty}\left[\rho^{k-1}(x^{(1)},x^{(2)})-\rho^k(x^{(1)},x^{(2)})\right](\mathbf{Q}_+^\infty\mathbf{U}^k)(x^{(1)},x^{(2)},\,\cdot\,)$$

This factorization has the following intuitive meaning. The measure $\mathbf{Q}_+^\infty$ on pairs of complete futures staring from $(x_0^{(1)},x_0^{(2)})$ can be decomposed into a mixture of the measure $\mathbf{S}^\infty$, which lead to convergence for all time, and the measures $\mathbf{Q}_+^\infty\mathbf{U}^k$, which lead to convergence for the first $k-1$ steps and then not on the kth step. The ρ factors can be understood as the probability of drawing from a given term in the factorization. Hence there is a ρ^∞ chance of drawing from the “similar for all time” part of the distribution for all time and a $\rho^{k-1}-\rho^k$

chance of drawing from the part "similar" for exactly $k-1$ steps. It is important to observe that all but the totally "similar" part of the decomposition $\mathbf{S}^\infty$ have a finite length part which is not the standard distribution $\mathbf{Q}_+^\infty$.

Proof of lemma 2.1.

$$\begin{aligned}\mathbf{Q}_+(x^{(1)}, x^{(2)}, \cdot\,) &\overset{\text{M}}{\sim} \mathbf{s}_1(x^{(1)}, x^{(2)}, \cdot\,) + \mathbf{r}_1(x^{(1)}, x^{(2)}, \cdot\,)\\ &\overset{\text{M}}{\sim} \rho^1(x^{(1)}, x^{(2)})\mathbf{S}^1(x^{(1)}, x^{(2)}, \cdot\,) + (1-\rho^1(x^{(1)}, x^{(2)}))\mathbf{R}^1(x^{(1)}, x^{(2)}, \cdot\,)\end{aligned}$$

In the following, we suppress the dependence on the initial condition in the interest of brevity. Next

$$\mathbf{Q}_+^2 \overset{\text{M}}{\sim} \mathbf{Q}_+\mathbf{s}_1 + \mathbf{Q}_+\mathbf{r}_1 \overset{\text{M}}{\sim} \mathbf{s}_2\mathbf{s}_1 + \mathbf{r}_2\mathbf{s}_1 + \mathbf{Q}_+\mathbf{r}_1\ .$$

Evaluating these measures on the set $\mathbb{X}_+^2 \times \mathbb{X}_+^2$ produces

$$1 = \rho^2 + \mathbf{r}_2\mathbf{s}_1(\mathbb{X}_+^2 \times \mathbb{X}_+^2) + (1-\rho^1)\ .$$

Simplifying produces $\mathbf{r}_2\mathbf{s}_1(\mathbb{X}_+^2 \times \mathbb{X}_+^2) = \rho^1 - \rho^2$ and hence

$$\begin{aligned}\mathbf{Q}_+^2 &\overset{\text{M}}{\sim} \rho^2\mathbf{S}^2 + (\rho^1-\rho^2)\mathbf{R}_2\mathbf{S}_1 + (1-\rho^1)\mathbf{Q}_+\mathbf{R}_1\\ &\overset{\text{M}}{\sim} \rho^2\mathbf{S}^2 + (\rho^1-\rho^2)\mathbf{U}^2 + (1-\rho^1)\mathbf{Q}_+\mathbf{U}^1\end{aligned}$$

Continuing in this fashion produces the quoted result. □

From the proof of the above lemma, one easily extracts the following corollary.

Corollary 2.2.

$$\rho^{k-1}(x^{(1)}, x^{(2)}) - \rho^k(x^{(1)}, x^{(2)}) = \mathbf{r}_k\mathbf{s}^{k-1}(x^{(1)}, x^{(2)}, \mathbb{X}^k \times \mathbb{X}^k)$$

This corollary also has an intuitive meaning. $\rho^{k-1}(x^{(1)}, x^{(2)}) - \rho^k(x^{(1)}, x^{(2)})$ is the difference in the probability of drawing from $\mathbf{s}^{k-1}$ and $\mathbf{s}^k$. The only way this can happen is to draw from $\mathbf{s}^{k-1}$ for the first $k-1$ steps and but not from $\mathbf{s}_k$ on the kth step. This is exactly what $\mathbf{r}_k\mathbf{s}^{k-1}$ represents.

2.3.2 The Reconstructed Process

We now build a Markov process $(x_n^{(1)}, x_n^{(2)}, M_n^{(1)}, M_n^{(2)})$ which will be central to all the following analysis. By construction, the $x_n^{(i)} \in \mathbb{X}$ will be a realization of the Markov chain with transition kernel P and initial condition $x_0^{(i)}$. We emphasis that $x_n^{(1)}$ and $x_n^{(2)}$ will not be independent realizations. The entire point of coupling is to build useful correlations between the two processes. At times we will want to view $x_n^{(i)}$ as an element of $\mathbb{X}_+ \times \mathbb{X}_-$ and we write $x_n^{(i)} = (z_n^{(i)}, y_n^{(i)})$. The $M_n^{(i)}$ can be seen as some internal state which will help our analysis. $M_n^{(i)}$ will be used as a stack to store future states of $z_n^{(i)}$. More precisely for each n, $M_n^{(i)}$ is an element of $\mathbb{X}_+^k$ for some k, $k \in \{1, 2, 3, \ldots, \infty\}$. Recall that π_1 is the projection of an element of $\mathbb{X}_+^k$ onto its first element and $\pi_1^\perp$ was the projection orthogonal π_1. Hence π_1 returns the first element in the stack and $\pi_1^\perp$ returns everything else. By combining these two maps, we implement what is usually referred to as a "pop" from a stack. π_1 returns the top element in the stack and $\pi_1^\perp$ returns an updated stack which has the previous top element removed.

The dynamics of $(x_n^{(1)}, x_n^{(2)}, M_n^{(1)}, M_n^{(2)})$ is given by the following rules where $z_n = \Pi_+ x_n$ and $y_n = \Pi_- x_n$:

i) There is something on the stack; pop it off and use it: More precisely if $M_n^{(1)}$ is not the empty set, then for $i = 1, 2$ we set $z_{n+1}^{(i)} = \pi_1 M_n^{(i)}$, $M_{n+1}^{(i)} = \pi_1^\perp M_n^{(i)}$, and $y_{n+1}^{(i)} = \Phi(z_{n+1}^{(i)}, y_n^{(i)})$.

ii) The stack is empty but we are not in the set $\mathbf{C}$ where we have control over the factorization. Take a step from $\mathbf{Q}_+$: More precisely if $M_n^{(1)}$ is the empty set and $(x_n^{(1)}, x_n^{(2)}) \notin \mathbf{C}$ then we chose $(z_{n+1}^{(1)}, z_{n+1}^{(2)})$ according to the distribution $\mathbf{Q}_+(x_n^{(1)}, x_n^{(2)}, \cdot)$ and set $y_{n+1}^{(i)} = \Phi(z_{n+1}^{(i)}, y_n^{(i)})$ and $M_{n+1}^{(i)} =$ empty set.

iii) The stack is empty and we are in the set $\mathbf{C}$ where we have control over the factorization. Try to couple for all future times: More precisely if $M_n^{(1)}$ is the empty set and $(x_n^{(1)}, x_n^{(2)}) \in \mathbf{C}$ then, we pick an α from $\{1, 2, 3, \ldots, \infty\}$ according to

$$\begin{aligned} \alpha =& k \qquad \text{with probability } \rho^{k-1}(x_n^{(1)}, x_n^{(2)}) - \rho^k(x_n^{(1)}, x_n^{(2)}) \\ \alpha =& \infty \qquad \text{with probability } \rho^{\infty}(x_n^{(1)}, x_n^{(2)}) \ . \end{aligned}$$

Recall by convention $\rho^0 = 1$. Since the ρ^k partition the unit interval, this construction is well defined. We now choose an element $(\zeta^{(1)}, \zeta^{(2)})$ of $\mathbb{X}_+^\alpha \times \mathbb{X}_+^\alpha$ by the following prescription: If $\alpha = \infty$, we pick $(\zeta^{(1)}, \zeta^{(2)})$ according to the distribution $\mathbf{S}^\infty(x_n^{(1)}, x_n^{(2)}, \cdot)$ otherwise according to $\mathbf{U}^\alpha(x_n^{(1)}, x_n^{(2)}, \cdot)$. In all cases, we set $z_{n+1}^{(i)} = \pi_1 \zeta^{(i)}$ and $M_{n+1}^{(i)} = \pi_1^\perp \zeta^{(i)}$. Lastly, we set $y_{n+1}^{(i)} = \Phi(z_{n+1}^{(i)}, y_n^{(i)})$.

This completes the coupling construction.

Let us pause for a moment and highlight the main features of this construction. When the chain is not in the set $\mathbf{C}$ at time n, each coordinate draws its next step from $\mathbf{Q}_+$. It is useful to recall that usually $\mathbf{Q}_+ = Q_+ \times Q_+$, so the two coordinates move independently in this case. The first time the chain finds itself in $\mathbf{C}$, it draws a segment of future of a random length which is specified by the random variable α. If $\alpha = \infty$ then the choice is made from $\mathbf{S}^\infty$ and the chains are said to have "coupled." This infinite future is placed on the stack M where it is "popped off" one by one for the rest of time. By the properties of $\mathbf{S}^\infty$, the enslaved degrees of freedom will converge to each other asymptotically in time. If $\alpha \neq \infty$, then a future of finite length is drawn from $\mathbf{U}^\alpha$. This future is placed on the stack, where it is again "popped off" one by one until nothing is left. Once the stack the stack is again empty, we are free to draw a new from $\mathbf{Q}_+$ until the chain is again in $\mathbf{C}$ and we can again try to couple for all times. We will refer to the times what the stack is empty ($|M| = 0$) as "unbiased" because the future depends only on the current state and we are free to pick the next step basses on the transition kernel. When there are elements on the stack we are required to use them as the decision about next step had already been made.

2.4 A Basic Ergodic Result

In this section, we give basic convergence results, postponing discussion of the convergence rate until section 2.5. We have the following basic ergodic result.

Theorem 2. *Under Assumptions A1- A4, for any* $(x_0^{(1)}, x_0^{(2)}) \in \mathbb{X} \times \mathbb{X}$

$$\|P^n(x_0^{(1)}, \cdot\,) - P^n(x_0^{(2)}, \cdot\,)\|_{\mathcal{G}} \to 0 \text{ as } n \to \infty \;.$$

Before giving the proof, we establish some notation. For $(x_0^{(1)}, x_0^{(2)}) \in \mathbb{X} \times \mathbb{X}$ define the stopping time τ by

$$\tau(x_0^{(1)}, x_0^{(2)}) = \inf\{n : |M_n| = \infty\} \;. \tag{5}$$

Since

$$\|P^n(x_0^{(1)}, \cdot\,) - P^n(x_0^{(2)}, \cdot\,)\|_{\mathcal{G}} = \sup_{\mathsf{g} \in \mathcal{G}} \left|\mathbb{E}\mathsf{g}(x_n^{(1)}) - \mathbb{E}\mathsf{g}(x_n^{(2)})\right|$$

the importance of the stopping time τ shown by the following central lemma.

Lemma 2.3 (Basic convergence estimate). *For any* $\mathsf{g} \in \mathcal{G}$ *and* $(x_0^{(1)}, x_0^{(2)}) \in \mathbb{X} \times \mathbb{X}$,

$$\left|\mathbb{E}\mathsf{g}(x_n^{(1)}) - \mathbb{E}\mathsf{g}(x_n^{(2)})\right| \le \mathbb{P}\left\{\tau > \frac{n}{2}\right\} + \mathsf{G}\left(\frac{n}{2}\right)$$

where G *is described in Assumption A3 and* τ *defined by* (5).

Since by construction of the chain, τ must occur after a visit to $\mathbf{C}$ when M_n is the empty set. ($\tilde{M}_n$ will also be empty as M_n and $\tilde{M}_n$ always have the same length by construction.) Hence, it is important to track these visits. To this end, we introduce the following stopping times:

$$\begin{aligned} &\upsilon_0 = 0 \\ &t_k = \inf\{n \ge \upsilon_{k-1} : (x_n^{(1)}, x_n^{(2)}) \in \mathbf{C}\} \quad k = 1, 2 \cdots \\ &\upsilon_k = \inf\{n > t_k : |M_n| = 0 \text{ or } |M_n| = \infty\} \quad k = 1, 2, \cdots \end{aligned} \tag{6}$$

The t_1 is the first visit to $\mathbf{C}$ by the chain. Assuming the chain has not yet coupled, and hence $|M| \neq \infty$, υ_1 will be the first time the stack is empty. As such, υ_1 is the first time when we are again eligible to couple. Since t_2 is the first visit to $\mathbf{C}$ on or after υ_1, t_2 is precisely the next time we try to couple. Assuming that coupling attempt failed, υ_2 will again be the first time when we can try to couple if we find the chain in $\mathbf{C}$. The time between t_1 and υ_1 can be thought of as a blackout region where the future has already been chosen and we can not manipulate it. Once the chain has coupled, we turn off this "blackout" feature and the t_k simply track the visits to $\mathbf{C}$. This is done so that all of the t_k and υ_k are finite which simplifies some notation and arguments. However, this feature of the construction has no mathematical content.

As the visits to various sets (namely $\mathbf{C}$) will be important, we define the following hitting times. For any $\mathbf{x}_0 = (x_0^{(1)}, x_0^{(2)}) \in \mathbb{X} \times \mathbb{X}$ and measurable set $\mathbf{A}$, we define

$$\tau_{\mathbf{A}}(\mathbf{x}_0) = \inf\{n \geq 0 : \mathbf{x}_n \in \mathbf{A}\} \text{ and } \dot{\tau}_{\mathbf{A}}(\mathbf{x}_0) = \inf\{n > 0 : \mathbf{x}_n \in \mathbf{A}\} \tag{7}$$

The two definition differ only in there treatment of the initial condition.

We now return to the proof of Theorem 2 using Lemma 2.3.

Proof of Theorem 2. Since $\|P^n(x_0^{(1)}, \cdot) - P^n(x_0^{(2)}, \cdot)\|_{\mathcal{G}} = \sup_{\mathsf{g}\in\mathcal{G}} \left|\mathbb{E}\mathsf{g}(x_n^{(1)}) - \mathbb{E}\mathsf{g}(x_n^{(2)})\right|$ and $\lim \mathsf{G}(n) = 0$, Lemma 2.3 reduces the proof of the theorem to showing that $\mathbb{P}\left\{\tau > \frac{n}{2}\right\} \to 0$ as $n \to 0$. We now analyze this using the stopping times defined in (6). Since with probability one, the trajectory enters $\mathbf{C}$ in finite time regardless of the initial condition, we know that t_1 is almost surely finite. Since $|M_n|$ decreases by one on each step until it reaches zero, $\upsilon_1 - t_1$ is finite with probability one. (When $|M_{1+t_1}| = \infty$, $\upsilon_1 - t_1 = 1$ by construction.) In all cases, υ_1 is finite with probability one. Again since with probability one the trajectory will again enter $\mathbf{C}$ regardless of the state at time υ_1, we see that t_2 is finite with probability one. Continuing in this manner, we see that each t_k is almost surely finite and $\lim t_k = \infty$.

We now return to showing that $\mathbb{P}\left\{\tau > \frac{n}{2}\right\} \to 0$ as $n \to 0$. First observe that $\mathbb{P}\left\{\tau > 1 + t_k\right\} \leq (1 - \rho_*^\infty)^k$ because on each visit to $\mathbf{C}$ when $|M_n| = 0$ there is at most $(1 - \rho_*^\infty)$ chance that the chain does not couple. Now pick an $\epsilon > 0$ and fix a k so that $(1 - \rho_*^\infty)^k < \frac{\epsilon}{2}$. Since $t_k < \infty$ almost surely, there exists a finite m so that $\mathbb{P}\{t_k \geq m\} < \frac{\epsilon}{2}$. To complete the proof observe that if $n/2 > m$ then

$$\begin{aligned}
\mathbb{P}\left\{\tau > \frac{n}{2}\right\} &= \mathbb{P}\left\{t_k < \frac{n}{2} < \tau\right\} + \mathbb{P}\left\{\frac{n}{2} < \tau \;\&\; t_k \geq \frac{n}{2}\right\} \\
&\leq \mathbb{P}\left\{\tau > 1 + t_k\right\} + \mathbb{P}\left\{t_k \geq \frac{n}{2}\right\} \\
&\leq (1 - \rho_*^\infty)^k + \mathbb{P}\{t_k > m\} \leq \epsilon
\end{aligned}$$

Since ϵ was arbitrary, $\lim_{n\to\infty} \mathbb{P}\{\tau > \frac{n}{2}\} = 0$ which completes the proof. ☐

Proof of Lemma 2.3. For any n,

$$\begin{aligned}
\mathbb{E}\{\mathsf{g}(x_n^{(1)})\} =& \mathbb{E}\left\{\mathsf{g}(x_n^{(1)}) \;\&\; \tau > \frac{n}{2}\right\} + \mathbb{E}\left\{\mathsf{g}(x_n^{(1)}) \;\&\; \tau \leq \frac{n}{2}\right\} \\
=& \mathbb{E}\left\{\mathsf{g}(x_n^{(1)}) \;\&\; \tau > \frac{n}{2}\right\} + \mathbb{E}\left\{\mathsf{g}(x_n^{(2)}) \;\&\; \tau \leq \frac{n}{2}\right\} + \mathbb{E}\left\{[\mathsf{g}(x_n^{(1)}) - \mathsf{g}(x_n^{(2)})] \;\&\; \tau \leq \frac{n}{2}\right\} \\
\leq& \mathbb{P}\left\{\tau > \frac{n}{2}\right\} + \mathbb{E}\left\{\mathsf{g}(x_n^{(2)})\right\} + \mathbb{E}\left\{|\mathsf{g}(x_n^{(1)}) - \mathsf{g}(x_n^{(2)})| \;\Big|\; \tau \leq \frac{n}{2}\right\} .
\end{aligned}$$

Hence

$$\mathbb{E}\{\mathsf{g}(x_n^{(1)})\} - \mathbb{E}\left\{\mathsf{g}(x_n^{(2)})\right\} \leq \mathbb{P}\left\{\tau > \frac{n}{2}\right\} + \mathbb{E}\left\{|\mathsf{g}(x_n^{(1)}) - \mathsf{g}(x_n^{(2)})| \;\Big|\; \tau \leq \frac{n}{2}\right\} .$$

Reversing the role of $x^{(1)}$ and $x^{(2)}$ using the fact that $\mathbb{E}\left\{|\mathsf{g}(x_n^{(1)}) - \mathsf{g}(x_n^{(2)})| \;\Big|\; \tau \leq \frac{n}{2}\right\} \leq \mathsf{G}(\frac{n}{2})$, because the trajectory was draw from $\mathbf{S}^\infty$, produces the quoted estimate. ☐

2.5 Exponential Convergence

We now explore the rate of convergence. From Lemma 2.3, we see that the convergence rate is dictated by the speed at which G and $\mathbb{P}\{\tau > n\}$ converge to zero. Though an entire zoology of convergence rates are possible, we concentrate on exponential convergence.

As the rate at which $\mathbb{P}\{\tau > n\} \to 0$ depends on the frequency of visits to $\mathbf{C}$, again consider the stopping times $\{t_k\}$ defined in (6). With the following assumptions, we can prove exponential convergence to the invariant measure.

A5: The convergence is exponential when the appropriate active variable paths are used: *There exists constants K_0 and $\lambda_0 > 0$, so that the function G from assumption A3 satisfies $\mathsf{G}(n) < K_0 e^{-\lambda_0 n}$.*

A6: The stopping times measuring the recurrence of needed events have exponential moments.*: For some fixed $\lambda_1 > 0$ the following hold.*

i) The time between unbiased visits to $\mathbf{C}$ of the Markov chain constructed in Section 2.3 has exponential tails (unbiased visits are those when $|M_n| = 0$): *There exists a finite constant K_1 so that $\mathbb{E}\{\exp[\lambda_1(t_{k+1} - t_k)] | \tau > t_{k+1}\} \leq K_1$ for all $k \geq 1$.*

ii) Furthermore, for any initial conditions the time to first enter $\mathbf{C}$ has exponential tails (which may depend on the initial condition) : *For any initial condition $(x_0^{(1)}, x_0^{(2)}) \in \mathbb{X} \times \mathbb{X}$, $\mathbb{E}\exp(\lambda_1 \tau_{\mathbf{C}}) < \infty$. (The hitting time $\tau_{\mathbf{C}}(x_0^{(1)}, x_0^{(2)})$ is defined in (7).)*

Simply put the assumption A5 says that once the chain couples the enslaved variables converge to each other exponentially quickly. This is need if we want exponential convergence. The two parts of assumption A6, give up exponential control over the coupling time. The second part of the assumption ensures that the time to enters the set $\mathbf{C}$ in which we have control of the needed estimates has exponential tails. The first part of the assumption, ensures that failed coupling attempts do not introduce waiting times with fat tails. Hence it is not completely surprising that we have the following theorem.

Theorem 3. *Under assumptions A1– A3 and A5 –A6, there exists constants positive K and γ so that for any $x_0^{(i)} \in \mathbb{X}$*

$$\|P^n(x_0^{(1)}, \cdot\,) - P^n(x_0^{(2)}, \cdot\,)\|_{\mathcal{G}} \leq K\big[1 + \mathbb{E}\exp(\lambda_1 \tau_{\mathbf{C}})\big]e^{-\gamma n} \ .$$

where $\tau_{\mathbf{C}} = \tau_{\mathbf{C}}(x_0^{(1)}, x_0^{(2)})$.

Proof of Theorem 3. By our assumption on $\mathsf{G}(n)$ and Lemma 2.3, we have

$$\|P^n(x_0^{(1)}, \cdot\,) - P^n(x_0^{(2)}, \cdot\,)\|_{\mathcal{G}} \leq \mathbb{P}\left\{\tau > \frac{n}{2}\right\} + K_0 e^{-\lambda_0 \frac{n}{2}} \ .$$

Hence we need to show that $\mathbb{P}\left\{\tau > \frac{n}{2}\right\}$ decays exponentially in n. Begin by picking a $\lambda' > 0$ so that $a = \mathbb{E}\{\exp(\lambda'(t_{k+1} - t_k)) | \tau > t_{k+1}\}(1 - \rho_*^\infty) < 1$. This is always possible by the first part of assumption A6. Next let $\mathcal{F}_i$ be the sigma algebra generated by the process through step i. By Chebyshev's inequality and the definition of τ, we know that for any $k > 1$

$$\mathbb{P}\left\{\frac{n}{2} \le t_k \middle| t_k < \tau\right\} \le \exp(-\lambda'\frac{n}{2})\mathbb{E}\{\exp(\lambda' t_k)|\tau > t_k\}(1-\rho_*^\infty)^k$$
$$\le \exp\left[-\lambda_1\frac{n}{2}\right]\mathbb{E}\exp\left[\lambda_1 t_1\right]\prod_{i=2}^{k}\mathbb{E}\left\{\exp\left[\lambda_1(t_i - t_{i-1})\right]\middle|\mathcal{F}_i \ \&\ \tau > t_i\right\} \ .$$

By the definition of τ, we know that $\mathbb{P}\{t_k < \tau\} \le (1-\rho_*^\infty)^k$. Combining these two estimates and the definition of a defined above with the fact that τ must occur immediately after some t_k produces the estimate

$$\mathbb{P}\left\{\frac{n}{2} \le t_k \middle| t_k < \tau\right\}\mathbb{P}\{t_k < \tau\} \le \exp(-\lambda'\frac{n}{2})\mathbb{E}\exp\left[\lambda_1 t_1\right]a^{k-1} \ .$$

Since t_1 is just the first $n \ge 0$ so $(x_n^{(1)}, x_n^{(2)}) \in \mathbf{C}$, we know that $\tau > t_1$. Hence combining all of the estimates produces

$$\mathbb{P}\left\{\frac{n}{2} < \tau\right\} \le \sum_{k=2}^{\infty}\mathbb{P}\left\{\frac{n}{2} \le t_k < \tau\right\} = \sum_{k=2}^{\infty}\mathbb{P}\left\{\frac{n}{2} \le t_k \middle| t_k < \tau\right\}\mathbb{P}\{t_k < \tau\}$$
$$\le \exp(-\lambda'\frac{n}{2})\mathbb{E}\exp\left[\lambda_1 t_1\right]\frac{a}{1-a} = \exp(-\lambda'\frac{n}{2})\mathbb{E}\exp\left[\lambda_1 \tau_{\mathbf{C}}\right]\frac{a}{1-a}$$

completes the proof. □

Assumption 6 made the proof of Theorem 3 straight forward; however, it is not immediately obvious how to connect it with a particular system. In the remainder of this section and the subsequent two, we introduce a number of assumptions in this direction.

The stopping time $t_{k+1} - t_k$, when conditioned on $\tau > t_{k+1}$, is composed of two elements. The time to drain the stack after a failed coupling attempt and then the time to return to the set $\mathbf{C}$ after the stack has drained. The next assumption addresses the first of these issues.

A 7: It is exponentially unlikely to couple for a long time without coupling for all time: *There exist positive constants* λ_2 *and* K_2 *such that for all* $(x^{(1)}, x^{(2)}) \in \mathbf{C}$ *and* k, $\rho^{k-1}(x^{(1)}, x^{(2)}) - \rho^k(x^{(1)}, x^{(2)}) \le K_2 e^{-\lambda_2 k}$.

In terms of the coupling construction, this ensures that the random variable α_k defined by

$$\alpha_k = \upsilon_k - t_k \tag{8}$$

has exponential moments. When the system has not coupled, α_k is the time needed to drain the stack. In all cases, it is the time one needs to what until a visit to $\mathbf{C}$ can be used for t_{k+1}. Hence we have that $t_{k+1} - t_k = \alpha_k + \tau_{\mathbf{C}}(\mathbf{x}_{\upsilon_k})$. The exponential moments of α_k are encapsulated in the following lemma.

Lemma 2.4. *Under assumption A7 there exist positive constants* K_3 *and* γ_3 *so that for any* $n \ge 0$, $\mathbb{P}\{\alpha_k > n|\tau > \upsilon_k\} \le K_3 e^{-\lambda_3 n}$.

Proof of Lemma 2.4. From the construction of the coupled chain, we see that this is a question about the existence of exponential moments of the random variable α used in the construction in section 2.3. We see that

$$\mathbb{P}\{\alpha_k = j | \tau > \upsilon_k\} = \begin{cases} 1 - \rho^1(x^{(1)}_{t_k}, x^{(2)}_{t_k}) & \text{if } j = 1 \\ \rho^{j-1}(x^{(1)}_{t_k}, x^{(2)}_{t_k}) - \rho^j(x^{(1)}_{t_k}, x^{(2)}_{t_k}) & \text{for } 2 \leq j < \infty \ . \end{cases} \tag{9}$$

Hence the lemma is just a restatement of assumption A7. □

This lemma gives control over the time until the stack drains after each failed coupling attempt. The next two sections, make various assumptions which allow us to control the return time to the set $\mathbf{C}$. First we assume that the set $\mathbf{C}$ is the entire space. This makes all of the estimates uniform and has the flavor of a Doeblin Condition. In section 2.5.2, a different tact is used to control the times $t_{k+1} - t_k$. The section posits the existence of a Lyapunov function to control the excursions out of the set $\mathbf{C}$. The resulting theorem has the flavor of Harris's Condition [Har56].

2.5.1 Doeblin-like Condition

We now explore exponential convergence in a Doeblin-like setting where all of our estimates are uniform over the entire phase space.

Theorem 4. *If assumptions A1–A5 and assumption A7 all hold with a $\mathbf{C}$ equal the entire space $\mathbb{X} \times \mathbb{X}$ then there exist constants positive K and γ so that for any $(x^{(1)}_0, x^{(2)}_0) \in \mathbb{X} \times \mathbb{X}$*

$$\|P^n(x^{(1)}_0, \cdot) - P^n(x^{(2)}_0, \cdot)\|_{\mathcal{G}} \leq K e^{-\gamma n} \ .$$

Proof of Theorem 4. We will prove this theorem by connecting with Theorem 3. We begin by observing that since $\mathbf{C} = \mathbb{X} \times \mathbb{X}$, we have that $\tau_{\mathbf{C}} = 0$ for any starting point. Hence $t_{i+1} - \upsilon_i = 0$. Hence assumption A4 and the second part of assumption A6 hold trivially as we are always in the set $\mathbf{C}$. All that remains it to show that the first part of assumption A6 holds, after which Theorem 3 will apply and the proof will be complete. Since $\mathbf{C} = \mathbb{X} \times \mathbb{X}$ and $t_{k+1} = \upsilon_k$, we know that $t_{k+1} - t_k = \upsilon_k - t_k$. This is precisely the random variable defined in (8). Hence the fact that $\mathbb{E}\{\exp(\lambda'(t_{k+1} - t_k) | \tau > t_{k+1}\} < \infty$ uniformly in k for some $\lambda' > 0$ follows directly from Lemma 2.4 which complete the proof. □

2.5.2 Harris-like Condition

We now turn to the "Harris-like" setting where we will no longer assume that all of our estimates are uniform over the entire phase space. Rather, we will assume that we have the needed estimates over some central region of the phase space. To these weaker assumptions, we need to add the existence of some structure which ensures that the dynamics visit the center region regularly. With this concern in mind, we introduce the following assumption.

A8: There is a Lyapunov structure which pushes the dynamics back in to the center of the phase space. *There exists a $V : \mathbb{X} \to \mathbb{R}$ such that $V(x) \geq 0$ and if x_1 is distributed as $P(x_0, \cdot)$ then*

$$\mathbb{E}V(x_1) \leq aV(x_0) + b$$

for some $b > 0$ and $a \in (0, 1)$.

Let $\mathbf{V}(x^{(1)}, x^{(2)}) = V(x^{(1)}) + V(x^{(2)})$. Fixing an $\hat{a} \in (a, 1)$, we define $\hat{\mathbf{C}}(\hat{a}) = \{(x^{(1)}, x^{(2)}) \in \mathbb{X} \times \mathbb{X} : \mathbf{V}(x^{(1)}, x^{(2)}) \leq \frac{2b}{\hat{a}-a}\}$.

Theorem 5. *If assumptions A1–A3, A5 and A7–A8 hold with a* $\mathbf{C}$ *such that* $\mathbf{C} = \hat{\mathbf{C}}(\hat{a})$ *for some* $\hat{a}$ *(where* $\hat{\mathbf{C}}$ *is defined above) then there exist constants positive* K *and* γ *so that for any* $(x_0^{(1)}, x_0^{(2)}) \in \mathbb{X} \times \mathbb{X}$

$$\|P^n(x_0^{(1)}, \cdot) - P^n(x_0^{(2)}, \cdot)\|_{\mathcal{G}} \leq K\big[1 + \mathbb{E}V(x_0^{(1)}) + \mathbb{E}V(x_0^{(2)})\big]e^{-\gamma n} \ .$$

Proof of Theorem 5. All we have to do is to prove that together Assumption A7 and Assumption A8 together imply Assumption A6 and that there is a constant K' so that $\mathbb{E}\exp(\lambda_1 \tau_{\mathbf{C}}) \leq K'\big[1 + \mathbb{E}\mathbf{V}(x_0^{(1)}, x_0^{(2)})\big]$. Once this is done, the theorem is implied by Theorem 3. For notational brevity, we will write $\mathbf{x}_n$ for $(x_n^{(1)}, x_n^{(2)})$. Recall the hitting times defined in (7). We have the following estimates whose proofs are discussed at the end of this section:

Lemma 2.5. *Under Assumption A8, if* $x_n^{(i)}$ *is the nth step of a chain starting from* $x_0^{(i)}$ *and* $\mathbf{x}_n = (x_n^{(1)}, x_n^{(2)})$ *then*

$$\mathbb{E}\mathbf{V}(\mathbf{x}_n) \leq a^n \mathbb{E}\mathbf{V}(\mathbf{x}_0) + \frac{2b}{1-a}(1 - a^n) \ .$$

Lemma 2.6. *Under Assumption A8, for* $n > 0$

$$\mathbb{P}\{\dot{\tau}_{\hat{\mathbf{C}}}(\mathbf{x}_0) > n\} \leq \hat{a}^n \left(\frac{\hat{a}-a}{2b}\mathbb{E}\mathbf{V}(\mathbf{x}_0) + \frac{\hat{a}-a}{\hat{a}}\mathbb{E}\mathbf{1}_{\hat{\mathbf{C}}}(\mathbf{x}_0)\right)$$
$$\mathbb{P}\{\tau_{\hat{\mathbf{C}}}(\mathbf{x}_0) > n\} \leq \hat{a}^n \left(\frac{\hat{a}-a}{2b}\right)\mathbb{E}\mathbf{V}(\mathbf{x}_0) \ .$$

Since $\hat{a} \in (a, 1)$, it is clear from the second lemma that for some $\lambda_{1,1} > 0$ and $K' > 0$, $\mathbb{E}\exp\big(\lambda_{1,1}\tau_{\hat{\mathbf{C}}}(\mathbf{x}_0)\big) \leq K'[1 + \mathbf{V}(\mathbf{x}_0)]$. Assuming that we pick $\lambda_1 \leq \lambda_{1,1}$, this gives the needed estimate on $\tau_{\hat{\mathbf{C}}}(\mathbf{x}_0)$ to satisfy the second part of Assumption A6. We now turn to the estimates on $t_{k+1} - t_k$ for $k \geq 1$. For definiteness, fix some $k \geq 1$. Set $d_k = t_{k+1} - t_k$ and again $\alpha_k = \upsilon_k - t_k$. As observe before, $d_k = \alpha_k + \tau_{\hat{\mathbf{C}}}(\mathbf{x}_{\upsilon_k})$.

Fixing a $\delta \in (0, 1)$ and k, observe that

$$\mathbb{P}\{d_k > n\} = \mathbb{P}\{d_k > n \ \& \ \alpha_k \geq \delta n\} + \mathbb{P}\{d_k > n \ \& \ \alpha_k < \delta n\} \ . \tag{10}$$

Lemma 2.4 implies that

$$\mathbb{P}\{d_k > n \ \& \ \alpha_k \geq \delta n\} \leq \mathbb{P}\{\alpha_k \geq \delta n\} \leq K_3 \exp(-\lambda_3 \delta n) \ . \tag{11}$$

The second term requires a bit more work. First notice that if we condition on $\alpha_k = j$ then $\mathbf{x}_{\upsilon_k}$ is distributed as $(\chi_j^{(1)}, \chi_j^{(2)})$ where $\chi^{(i)} = \Upsilon(\zeta^{(i)}, x_{t_k}^{(i)})$ and $(\zeta^{(1)}, \zeta^{(2)})$ is distributed according to $\mathbf{U}^j(\mathbf{x}_{\upsilon_k})$. We need to estimate $\mathbb{P}\{\tau_{\hat{\mathbf{C}}}(\mathbf{x}_{\upsilon_k}) > l | \alpha_k = j\}$. From Lemma 2.6, we see that there exists a $K > 0$ and $\gamma > 0$ so that for any $l > 0$

$$\mathbb{P}\{\tau_{\hat{\mathbf{C}}}(\mathbf{x}_{\upsilon_k}) > l | \alpha_k = j\} \leq K e^{-\gamma l} \mathbb{E}\mathbf{V}(\chi_j^{(1)}, \chi_j^{(2)}) \ .$$

We need to estimate $\mathbb{E}\mathbf{V}(\chi_j^{(1)}, \chi_j^{(2)})$. By construction (see the proof of Lemma 2.1)

$$\mathbf{Q}_+^j(\mathbf{x}_{t_k},\,\cdot\,) \geq \big[\rho^{j-1}(\mathbf{x}_{t_k}) - \rho^j(\mathbf{x}_{t_k})\big]\mathbf{U}^j(\mathbf{x}_{t_k},\,\cdot\,) = \mathbb{P}\{\alpha_k = j\}\mathbf{U}^j(\mathbf{x}_{t_k},\,\cdot\,)\ .$$

Since $\mathbf{V} \geq 0$, this implies that if $(X^{(1)}, X^{(2)}) \in \mathbb{X}^j \times \mathbb{X}^j$ is distributed as $(Q^j \times Q^j)(\mathbf{x}_{t_k},\,\cdot\,)$ then

$$\mathbb{E}\mathbf{V}(\chi_j^{(1)}, \chi_j^{(2)}) \leq \frac{\mathbb{E}\mathbf{V}(X_j^{(1)}, X_j^{(2)})}{\mathbb{P}\{\alpha_k = j\}}\ .$$

Because $\mathbf{V}(\mathbf{x}_{t_k}) \leq K_0$, Lemma 2.5 implies that there exists a j independent constant K_0' so that $\mathbb{E}\mathbf{V}(X_j^{(1)}, X_j^{(2)}) \leq K_0'$. Hence we have that

$$\mathbb{P}\{\tau_{\hat{\mathbf{C}}}(\mathbf{x}_{v_k}) > l \ \&\ \alpha_k = j\} \leq K e^{-\gamma l}\mathbb{E}\mathbf{V}(\chi_j^{(1)}, \chi_j^{(2)})\mathbb{P}\{\alpha_k = j\} \leq K K_0' e^{-\gamma l}$$

Returning to (11), we conclude by observing that

$$\begin{aligned}\mathbb{P}\{d_k > n \ \&\ \alpha_k < \delta n\} &= \sum_{j=1}^{\lfloor \delta n \rfloor} \mathbb{P}\{\tau_{\hat{\mathbf{C}}}(\mathbf{x}_{v_k - 1}) > n - j \ \&\ \alpha_k = j\} \\ &\leq K K_0' \sum_{j=1}^{\lfloor \delta n \rfloor} e^{-\gamma(n-l)} \leq \hat{K} e^{-\gamma(1-\delta)n}\ .\end{aligned}$$

Combining all of our estimates we obtain that d_k has exponentially decaying tails independent with bounds independent of k which concludes the proof. □

Proof of Lemma 2.6 and Lemma 2.5. Both lemmas are simple consequences of the Lyapunov structure. The proof of the continuous analog of Lemma 2.6 can be found in Lemma 3.2 of [EMS01]. In the discrete setting, similar statements are proved in Lemma 9.3 of [MSH01] or Lemma 11.3.9 of [MT93]. Lemma 2.5 is just integrating up the differential inequality given by the Lyapunov structure. This can be found in many references. For instance see Lemma 9.3 of [MSH01] or [Has80] for the continuous setting. □

3 A First Application

Consider the probability space $(\Omega, \mathbb{P}, \mathcal{F}, \theta)$ where each $\omega \in \Omega$ is an infinite sequence of the form $(\omega_1, \omega_2, \dots)$ and θ is the shift which maps $(\omega_1, \omega_2, \dots)$ to $(\omega_2, \omega_3, \dots)$. We assume that θ is ergodic with respect to $\mathbb{P}$. Let θ^n denote the n-fold composition of θ.

Let l^2 denote the space of square summable sequences that is the $y = (y(1), y(2), y(3), \dots) \in \mathbb{R}^\infty$ with $\|y\|_2 = \big[\sum_k y(k)^2\big]^{\frac{1}{2}} < \infty$. We denote by l_1^2 the set $\{y \in l^2 : \|y\|_2 \leq 1\}$. Consider the following iterated random map from $\phi : [0,1]^n \times l_1^2 \to [0,1]^n \times l_1^2$

$$\phi_\omega : \begin{pmatrix} z \\ y \end{pmatrix} \longmapsto \begin{pmatrix} f_1(z,y) + f_2(z,y)\cdot W(\omega) \quad (\mod 1\ ) \\ h(z)\cdot(y * y) + g_1(z) + g_2(z)\cdot B(\omega) \end{pmatrix}\ . \tag{12}$$

Here "$*$" denotes the convolution and "$\cdot$" denotes coordinate by coordinate multiplication. $W(\omega)$ is a non-degenerate Gaussian random variable on $\mathbb{R}^n$ and $B(\omega)$ is any random variable such that $\|B\|_2 \leq 1$ almost surely. We assume that the collection of random variables

$$\{W(\omega), W(\theta\omega), W(\theta^2\omega), \dots, B(\omega), B(\theta\omega), B(\theta^2\omega), \dots\}$$

are jointly independent. Furthermore we take $f_i : [0,1]^n \times l_1^2 \to l_1^2$ such that $f_i(z,y)$ is continuous in y. We also assume that there exist fixed, positive numbers f^* and f_* so that $0 < f_* \leq \big(f_2(z,y)\big)_j \leq f^* < \infty$ for all z and y and $j = 1, \dots, n$. Here $(\quad)_j$ is the jth coordinate of the vector. $h : [0,1]^n \to l^2$ such that there exists a positive h^* so $\sup_z \|h(z)\|_2 \leq \frac{h^*}{2} < \frac{1}{2}$. The $g_i : [0,1]^n \to l^2$ are such that $\sup_z \|g_1(z)\|_2 + \|g_2(z)\|_2 \leq \frac{1}{2}$.

We define $x_n^{(i)} = (z_n^{(i)}, y_n^{(i)}) = \phi_{\theta^n\omega}(z_{n-1}^{(i)}, y_{n-1}^{(i)})$ for initial conditions $x_0^{(1)}, x_0^{(2)} \in [0,1]^n \times l_1^2$. In this setting, we can use Theorem 4 to prove

Theorem 6. *There exist positive constants K a γ so that for almost every realization $\mathbf{B}(\omega) = (B(\omega), B(\theta\omega), B(\theta^2\omega) \dots)$*

$$\sup_{\mathsf{g}\in\mathcal{G}} \left|\mathbb{E}\{\mathsf{g}(x_n^{(1)})|\mathbf{B}\} - \mathbb{E}\{\mathsf{g}(x_n^{(2)})|\mathbf{B}\}\right| \leq K e^{-\gamma n}$$

where

$$\mathcal{G} = \Big\{\mathsf{g} : [0,1]^n \times l_1^2 \to \mathbb{R} : \mathsf{g}(z,y) \textit{ measurable}, \ \sup_{z,y} |\mathsf{g}(z,y)| \leq 1, \\ |\mathsf{g}(z,y) - \mathsf{g}(z,y')| < \|y - y'\|_2 \textit{ for all } z, y, y'\Big\}$$

Of course the above theorem implies that

$$\sup_{\mathsf{g}\in\mathcal{G}} \left|\mathbb{E}\{\mathsf{g}(x_n^{(1)})\} - \mathbb{E}\{\mathsf{g}(x_n^{(2)})\}\right| \leq K e^{-\gamma n} .$$

If we define

$$\mathcal{G}_{TV} = \Big\{\mathsf{g} : [0,1]^n \times l_1^2 \to \mathbb{R} : \mathsf{g}(z,y) \text{ measurable}, \ \sup_{z,y} |\mathsf{g}(z,y)| \leq 1, \Big\}$$

and

$$\mathcal{G}_W = \Big\{\mathsf{g} : [0,1]^n \times l_1^2 \to \mathbb{R} : \mathsf{g}(z,y) \text{ measurable}, \ \sup_{z,y} |\mathsf{g}(z,y)| \leq 1, \\ |\mathsf{g}(z,y) - \mathsf{g}(z',y')| < |z - z'| + \|y - y'\|_2 \text{ for all } z, , z', y, y'\Big\}$$

then $|\cdot|_{\mathcal{G}_{TV}}$ and $|\cdot|_{\mathcal{G}_W}$ correspond respectively to the standard total variation and Wasserstein distances on probability measures. Both of these metrics are complete. (see [Dud76].) Since $\mathcal{G}_W \subset \mathcal{G}$, convergence in $|\cdot|_{\mathcal{G}}$ implies convergence in $|\cdot|_{\mathcal{G}_W}$. Yet, if the test function depends only on z then convergence in $|\cdot|_{\mathcal{G}}$ is equivalent to $|\cdot|_{\mathcal{G}_{TV}}$.

Our conditions on ϕ_ω were designed to ensure that the map had a number of properties:

i) ϕ is a map from $[0,1]^n \times l_1^2 \to [0,1]^n \times l_1^2$: Since the new z is calculated module 1, it is clear that it is in $[0,1]^n$. The fact that the y component stays in l_1^2 follows from a combination of Young's and Hölder's inequality. Fix $(z,y) \in [0,1]^n \times l_1^2$ then

$$\begin{aligned}\|h(z)\cdot(y*y)+g_1(z)+g_2(z)\cdot B\|_2 &\le \|h(z)\|_2\|y\|_2^2+\|g_1(z)\|_2+\|g_2(z)\|_2\|B\|_2\\ &< \frac{1}{2}+\|g_1(z)\|_2+\|g_2(z)\|_2 \le 1\end{aligned}$$

and hence $\phi_\omega(z,y) \in [0,1]^n \times l_1^2$ almost surely.

ii) For fixed z and B the map is a contraction: Fix $(z,y),(z,\tilde y) \in [0,1]^n \times l_1^2$ and ω and $\tilde\omega$ such that $B(\omega)=B(\tilde\omega)$. Letting Π_y denote the projection onto the y coordinate,

$$\begin{aligned}\|\Pi_y\phi_\omega(z,y)-\Pi_y\phi_{\tilde\omega}(z,\tilde y)\|_2 &= \|h(z)\cdot(y*y)-h(z)\cdot(\tilde y*\tilde y)\|_2 = \|h(z)\cdot[(y-\tilde y)*(y+\tilde y)]\|_2\\ &\le \|h(z)\|_2\|y-\tilde y\|_2\|y+\tilde y\|_2 \le \frac{h^*}{2}\|y-\tilde y\|_2 2 \le h^*\|y-\tilde y\|_2\end{aligned}$$

Since $h^* < 1$, the claim holds.

iii) The z motion can be described by a density which is smooth in the initial y, everywhere positive, and uniformly bounded from above and below: Because W is a non-degenerate Gaussian there exists a function $p(z,y,z_1)$ such that for measurable $A \subset \mathbb{R}^n$

$$\mathbb{P}\left\{\phi_\omega(z,y)\in A\right\} = P_z(z,y,A) = \int_A p(z,y,z_1)dz_1$$

Because f_2 is uniformly bounded from above and below, we know that p is everywhere positive and uniformly bounded from above and below. Because f_1 and f_2 are continuous in y, p is also continuous in y. Since $[0,1]^n$ is compact, we know that p is continuous in y uniformly in (z,y). Putting these facts together implies that there exist positive constants ρ^* and C so

$$\sup_{z,y,\tilde z,\tilde y}|P_z(z,y,\,\cdot\,)-P_z(\tilde z,\tilde y,\,\cdot\,)|_{TV} \le \rho^* < 1$$

and

$$\sup_z|P_z(z,y,\,\cdot\,)-P_z(z,\tilde y,\,\cdot\,)|_{TV} \le C\|y-\tilde y\|_2\ .$$

Having made these observations, Theorem 6 is a consequence of the following two theorems when we set $\mathbb{X}_+ = [0,1]^n$ and $\mathbb{X}_- = l_1^2$. The first theorem addresses the simplified case of Theorem 6 when we take $B(\omega)$ equal to zero. After discussing this case, we will turn to the slightly more complicated setting of $\mathcal{B}(\omega)$ not equal to zero.

Theorem 7. *Let $\mathbb{X}_+$ and $\mathbb{X}_-$ be two Banach spaces. Consider ϕ_ω a random map from $\mathbb{X}_+\times\mathbb{X}_- \to \mathbb{X}_+\times\mathbb{X}_-$ on a probability space $(\Omega,\mathbb{P},\mathcal{F},\theta)$. We assume ϕ has the form*

$$\phi_\omega(z,y) = \big(F_\omega(z,y),G(z,y)\big),$$

for $(z, y) \in \mathbb{X}_+ \times \mathbb{X}_-$. *Given two initial conditions* $x_0^{(i)} = (z^{(i)}, y^{(i)}) \in \mathbb{X}_+ \times \mathbb{X}_-$, $i = 1, 2$, *two noise realizations* $\omega^{(i)}$, *we define* $x_{k+1}^{(i)} = \phi_{\theta^k \omega^{(i)}}(x_k^{(i)})$. *There exist positive constants* K *and* γ,

$$\sup_{\mathsf{g} \in \mathcal{G}} \left| \mathbb{E}\{\mathsf{g}(x_n^{(1)})\} - \mathbb{E}\{\mathsf{g}(x_n^{(2)})\} \right| \leq K e^{-\gamma n}$$

where

$$\mathcal{G} = \Big\{ \mathsf{g} : \mathbb{X}_+ \times \mathbb{X}_- \to \mathbb{R} : \mathsf{g}(z, y) \text{ measurable }, \ \sup_{z,y} |\mathsf{g}(z, y)| \leq 1, \\ |\mathsf{g}(z, y) - \mathsf{g}(z, y')| < \|y - y'\| \text{ for all } z, y \Big\}$$

if the following conditions hold:

i) For $A \subset \mathbb{X}_+$, *define* $P_+(z, y, A) = \mathbb{P}\{\phi_\omega(z, y) \in A\}$. *There is a fixed positive constant* $\hat{\rho} < 1$, *with*

$$\sup_{z,y,\bar{z},\bar{y}} |P_+(z, y, A) - P_+(\bar{z}, \bar{y}, A)|_{TV} \leq \hat{\rho}$$

and a positive constant $\hat{C}_1$ *so*

$$\sup_{z,y,\bar{z},\bar{y}} |P_+(z, y, A) - P_+(\bar{z}, \bar{y}, A)|_{TV} \leq \hat{C}_1 \|y - \bar{y}\|$$

ii) There exist positive constants $\hat{\gamma}$ *and* $\hat{C}_2$, *such that for any sequence* $z_1, \ldots, z_n$

$$\sup_{y,\hat{y}} \|G^n(\{z_k\}_1^n, y) - G^n(\{z_k\}_1^n, \bar{y})\| \leq \hat{C}_2 e^{-\hat{\gamma} n}$$

where $G^n(\{z_k\}_1^n, y) \overset{def}{=} G(z_n, \cdot\,) \circ \cdots \circ G(z_2, \cdot\,) \circ G(z_1, y)$.

Proof of Theorem 7. This theorem will be implied by Theorem 4 after we validate Assumptions A1-A2, A3-A5, and Assumption A7.

We begin by defining $\mathbf{Q}_+(x^{(1)}, x^{(2)}, \cdot\,)$ as the product measure $P_+(x^{(1)}, \cdot\,) \times P_+(x^{(2)}, \cdot\,)$. Clearly this choice satisfies Assumption A1. We now define $s(x^{(1)}, x^{(2)}, \cdot\,) = P_+(x^{(1)}, \cdot\,) \wedge P_+(x^{(2)}, \cdot\,)$. This notation is explained in section C of the appendix. As defined $s(x^{(1)}, x^{(2)}, \cdot\,)$ is a measure on $\mathbb{X}_+$. We define $\mathbf{s}(x^{(1)}, x^{(2)}, \cdot\,)$ as the measure on $\mathbb{X}_+ \times \mathbb{X}_+$ concentrated on the diagonal elements (z, z) where z is distributed according to $s(x^{(1)}, x^{(2)}, \cdot\,)$. Next define $r(x^{(1)}, x^{(2)}, \cdot\,) = [P_+(x^{(1)}, \cdot\,) - P_+(x^{(2)}, \cdot\,)]^+$ and

$$\mathbf{r}(x^{(1)}, x^{(2)}, \cdot\,) = \frac{r(x^{(1)}, x^{(2)}, \cdot\,) \times r(x^{(2)}, x^{(1)}, \cdot\,)}{r(x^{(1)}, x^{(2)}, \mathbb{X})}$$

with the convention that $\frac{0}{0} = 0$. Observe that

$$\mathbf{Q}_+(x^{(1)}, x^{(2)}, \cdot\,) \overset{\mathrm{M}}{\sim} \mathbf{s}(x^{(1)}, x^{(2)}, \cdot\,) + \mathbf{r}(x^{(1)}, x^{(2)}, \cdot\,)$$

The added flexibility of defining different measures for different steps is not needed in this example. For $x^{(i)} = (x_0^{(i)}, \dots, x_{k-1}^{(i)}) \in \mathbb{X}^k \times \mathbb{X}^k$, for all k we define $\mathbf{s}_k(x^{(1)}, x^{(2)}, \cdot\,) = \mathbf{s}(\pi_{k-1}x^{(1)}, \pi_{k-1}x^{(2)}, \cdot\,)$ and $\mathbf{r}_k(x^{(1)}, x^{(2)}, \cdot\,) = \mathbf{r}(\pi_{k-1}x^{(1)}, \pi_{k-1}x^{(2)}, \cdot\,)$.

We set $\Phi(z, x) = G(x)$. Notice in the simple map setting when we define Φ, we do not need the "look ahead" of knowing the z at the moment of time we are reconstructing the y. This is needed in the continuous in time setting when each step of our chain will be a entire segment of trajectory.

We begin by showing that our choice of measures $\mathbf{s}^k$ and functions $\mathcal{G}$ satisfy assumptions A3–A5. Since $\mathbf{s}^\infty$ is concentrated on the diagonal, we can pick a set of full measure $Z \subset \mathbb{X}_+^\infty \times \mathbb{X}_+^\infty$ such that if $(\zeta^{(1)}, \zeta^{(2)}) \in Z$ then $\zeta^{(1)} = \zeta^{(2)}$. Next, for any two $x_0^{(i)} \in \mathbb{X}$, we define $\chi^{(i)} = \Upsilon(\zeta(i), x_0^{(i)}) \in \mathbb{X}^\infty$ and $y^{(i)} = \Pi_- \chi^{(i)}$. Then for any $\mathsf{g} \in \mathcal{G}$, the continuity of $\mathsf{g}(z, y)$ in y and the assumption on G implies $|\mathsf{g}(\pi_n \chi^{(1)}) - \mathsf{g}(\pi_n \chi^{(2)})| \le \mathsf{G}(n) = \hat{C}_2 e^{-\gamma n}$. This establishes assumptions A3–A5.

Then clearly our construction of $\mathbf{s}_k$ and $\mathbf{r}_k$ satisfies the first part of Assumption A2. The uniform lower bound on $\rho^k(x_0^{(1)}, x_0^{(2)}) = \mathbf{s}_k(x^{(1)}, x^{(2)}, \mathbb{X}_+^k \times \mathbb{X}_+^k)$ for finite k follows from the uniform upper bound $\hat{\rho}$ on the total variational for one step. To see this recall that

$$\begin{aligned}\mathbf{s}(x^{(1)}, x^{(2)}, \mathbb{X}_+ \times \mathbb{X}_+) &= s(x^{(1)}, x^{(2)}, \mathbb{X}_+)\\ &= 1 - |P_+(x^{(1)}, \cdot\,) - P_+(x^{(2)}, \cdot\,)|_{TV} \ge 1 - \hat{\rho}\end{aligned}$$

and thus

$$\begin{aligned}\mathbf{s}^k(x^{(1)}, x^{(2)}, \mathbb{X}_+^k \times \mathbb{X}_+^k) =& \mathbf{s}_k \cdots \mathbf{s}_1(x^{(1)}, x^{(2)}, \mathbb{X}_+^k \times \mathbb{X}_+^k)\\ &\ge \left[\inf_{x^{(i)}} \mathbf{s}(x^{(1)}, x^{(2)}, \mathbb{X}_+ \times \mathbb{X}_+)\right]^k \ge [1 - \hat{\rho}]^k > 0\end{aligned}$$

Clearly a better bound is needed for $k = \infty$. We will turn to this last. It will require an understanding of the contraction in the enslaved variables. From corollary 2.2

$$\rho^k(x_0^{(1)}, x_0^{(2)}) - \rho^{k+1}(x_0^{(1)}, x_0^{(2)}) = \int_{\mathbb{X}_+^k \times \mathbb{X}_+^k} \mathbf{s}^k(x_0^{(1)}, x_0^{(2)}, d\zeta^{(1)} \times d\zeta^{(2)}) \mathbf{r}_{k+1}(\chi^{(1)}, \chi^{(2)}, \mathbb{X}_+ \times \mathbb{X}_+)]$$

where $\chi^{(i)} = \pi_k^\perp \Upsilon(\zeta^{(i)}, \Pi_- x_0^{(i)})$. Continuing

$$\le \mathbf{s}^k(x_0^{(1)}, x_0^{(2)}, \mathbb{X}_+^k \times \mathbb{X}_+^k) \sup_{(\chi^{(1)}, \chi^{(2)}) \in D} \mathbf{r}_{k+1}(\chi^{(1)}, \chi^{(2)}, \mathbb{X}_+ \times \mathbb{X}_+)$$

where $D = \{(\chi^{(1)}, \chi^{(2)}) : \exists \zeta = (\zeta_1, \dots, \zeta_k) \in \mathbb{X}_+^k \text{ with } \chi^{(i)} = \pi_k \Upsilon(\zeta, x_0^{(i)})\}$. Because the same ζ is used to obtain both $\chi^{(1)}$ and $\chi^{(2)}$, our assumption on G implies that for all $(\chi^{(1)}, \chi^{(2)}) \in D$, $\|\Pi_- \chi^{(1)} - \Pi_- \chi^{(2)}\| \le \hat{C}_2 e^{-\hat{\gamma} k}$. Combining this with the second assumption P_+ produces

$$\sup_{(\chi^{(1)}, \chi^{(2)}) \in D} \mathbf{r}_{k+1}(\chi^{(1)}, \chi^{(2)}, \mathbb{X}_+ \times \mathbb{X}_+) = \sup_{(\chi^{(1)}, \chi^{(2)}) \in D} |P_+(\chi^{(1)}, \cdot\,) - P_+(\chi^{(2)}, \cdot\,)|_{TV} \le \hat{C}_1 \hat{C}_2 e^{-\hat{\gamma} k}\ .$$

Returning to our estimate of $\rho^k - \rho^{k+1}$ we see that

$$1 - \frac{\rho^{k+1}(x_0^{(1)}, x_0^{(2)})}{\rho^k(x_0^{(1)}, x_0^{(2)})} \le \hat{C}_1 \hat{C}_2 e^{-\hat{\gamma} k} \tag{13}$$

Hence assumption A7 holds in this example. In addition with a little work this estimate shows that ρ^∞ is uniformly bounded away from zero which is the missing element of assumption A2.

For any fixed n

$$\rho^\infty(x^{(1)}, x^{(2)}) = \rho^n(x^{(1)}, x^{(2)}) \prod_{k=n}^{\infty} \frac{\rho^{k+1}(x^{(1)}, x^{(2)})}{\rho^k(x^{(1)}, x^{(2)})}.$$

Since $\rho^n(x^{(1)}, x^{(2)}) \geq (1-\hat\rho)^n > 0$ for any n, we need only show that the product term is uniformly bounded away from zero for some n. Using estimate (13), we see that

$$\log \prod_{k=n}^{\infty} \frac{\rho^{k+1}(x^{(1)}, x^{(2)})}{\rho^k(x^{(1)}, x^{(2)})} = \sum_{k=n}^{\infty} \log \frac{\rho^{k+1}(x^{(1)}, x^{(2)})}{\rho^k(x^{(1)}, x^{(2)})} \geq \sum_{k=n}^{\infty} \log\left(1 - \hat C_1 \hat C_2 e^{-\hat\gamma k}\right) \ .$$

Since there exists a C' so that for large enough k, $\log\left(1 - \hat C_1 \hat C_2 e^{-\hat\gamma k}\right) > K' e^{-\hat\gamma k}$ it is clear that the sum is bounded away from $-\infty$ uniformly in $(x^{(1)}, x^{(2)})$. (Since $\rho^k \geq \rho^{k+1}$, the sum is bounded from above by zero.) Hence there exists a positive ρ^∞_* so that $\rho^\infty(x^{(1)}, x^{(2)}) \geq \rho^\infty_*$. This completes Assumption A2 and hence the proof. □

3.1 A Small Generalization

To handle the case of Theorem 6 when the B is not identically zero, we need a small generalization of our abstract theory. We could have addressed this case from the start, but the notation seemed complicated enough.

Since we are interested in general theorem applicable to the setting of Theorem 6, we pause fro a moment to explore that setting. Given our independence assumptions on B and W, we can split Ω into $\Omega_\beta \times \Omega_\eta$ such that $W(\omega) = W(\beta)$ and $B(\omega) = B(\eta)$. We will prove the following:

Theorem 8. *Let $\mathbb{X}_+$ and $\mathbb{X}_-$ be two Banach spaces. Consider ϕ_ω a random map from $\mathbb{X}_+ \times \mathbb{X}_- \to \mathbb{X}_+ \times \mathbb{X}_-$ on a probability space $(\Omega = \Omega_\beta \times \Omega_\eta, \mathbb{P} = \mathbb{P}_\beta \times \mathbb{P}_\eta, \mathcal{F}, \theta)$. We assume that ϕ has the form*

$$\phi_{(\beta,\eta)}(z, y) = \big(F_\beta(z, y), G_\eta(z, y)\big).$$

for $(z, y) \in \mathbb{X}_+ \times \mathbb{X}_-$. For any two initial conditions $x_0^{(i)} = (z^{(i)}, y^{(i)}) \in \mathbb{X}_+ \times \mathbb{X}_-$, $i = 1, 2$, noise realization $\omega^{(i)}$ we define $x_{k+1}^{(i)} = \phi_{\theta^k \omega^{(i)}}(x_k^{(i)})$. Then there exist positive constants K and γ such that for almost every sequence η

$$\sup_{\mathsf{g} \in \mathcal{G}} \left| \mathbb{E}\{\mathsf{g}(x_n^{(1)}) | \eta\} - \mathbb{E}\{\mathsf{g}(x_n^{(2)}) | \eta\} \right| \leq K e^{-\gamma n}$$

where

$$\mathcal{G} = \Big\{ \mathsf{g} : \mathbb{X}_+ \times \mathbb{X}_- \to \mathbb{R} : \mathsf{g}(z, y) \ \textit{measurable} \ , \ \sup_{z,y} |\mathsf{g}(z, y)| \leq 1, \\ |\mathsf{g}(z, y) - \mathsf{g}(z, y')| < \|y - y'\| \ \textit{for all} \ z, y \Big\}$$

as long as the following conditions holds for almost every realization η with constants independent of η:

i) For $A \subset \mathbb{X}_+$ and $\eta \in \Omega_\eta$, define $P_+(z,y,A|\eta) = \mathbb{P}_\beta\{F_{(\beta,\eta)}(z,y) \in A\}$. There exists a fixed positive constant $\hat{\rho} < 1$, with

$$\sup_{z,y,\bar{z},\bar{y}} |P_+(z,y,A|\eta) - P_+(\bar{z},\bar{y},A|\eta)|_{TV} \le \hat{\rho}$$

and a positive constant $\hat{C}_1$ so

$$\sup_{z,y,\bar{z},\bar{y}} |P_+(z,y,A|\eta) - P_+(\bar{z},\bar{y},A|\eta)|_{TV} \le \hat{C}_1 \|y - \bar{y}\| \ .$$

ii) There exist positive constants $\hat{\gamma}$ and $\hat{C}_2$, such that for any sequence $z_1, \dots, z_n$

$$\sup_{y,\hat{y}} \|G^n_\eta(\{z_k\}^n_1, y) - G^n_\eta(\{z_k\}^n_1, \bar{y})\| \le \hat{C}_2 e^{-\hat{\gamma} n}$$

where $G^n_\eta(\{z_k\}^n_1, y) \stackrel{def}{=} G_{\theta^{n-1}\eta}(z_n, \cdot\,) \circ \cdots \circ G_{\theta\eta}(z_2, \cdot\,) \circ G_\eta(z_1, y)$.

Proof of Theorem 8. The proof of this theorem is essentially identical to the proof of Theorem 7. It requires a small generalization of Theorem 4. The assumption that the map Φ, used to reconstruct the enslaved modes, is deterministic forces the structure of Theorem 7 from the previous section. In our present setting, we also need to know the realization of η to reconstruct the enslaved modes. In this case, we set $\Phi^\eta(z_1, z_0, y_0) = G_\eta(z_0, y_0)$. Now all of the quantities in section 2 depend on the realization of η. After taking each step, we now have to shift the η with θ. Though this seems more complicated, since we have assumed that all of our bounds are uniform in η, it amounts only to more complicated notation. The analysis of section 2 carries over with out modification. With this in hand, the proof of this theorem is the same as the previous theorem.

To proceed in cases were we do not have estimates uniform in η, we will need control of the constants ρ^*_k as we shift along η. As the shift along η is ergodic, it is reasonable to expect that such an estimate would exist in many settings. □

4 Ergodicity of the Navier Stokes Equation

We will work on a probability space $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P}, \theta_t)$. We associate Ω with the canonical space generated by all $d\omega_k(t)$. $\mathcal{F}$ and $\mathcal{F}_t$ are respectively the associated global σ-algebra and filtration generated by the $\{\omega_k(s)\}_{k\in\mathcal{W}}$ for $s \le t$. Expectations with respect to $\mathbb{P}$ will be denoted by $\mathbb{E}$.

Define $B(u,v) = -P_{div}(u\cdot\nabla)v$ and $\Lambda^2 u = -P_{div}\Delta u$ where P_{div} is the L^2 projection operator onto the space of divergence-free vector fields. Recall that $\sigma^{max}_k = \sup|\sigma_k(x)|$ and let $\sigma_{\max} = \max \sigma^{max}_k$ and $\mathcal{E}^*_j = \sum_k |k|^{2j}|\sigma^{max}_k|^2$. Writing $u(x,t) = \sum_k u_k(t)e_k(x)$, we will define

$$\mathbb{H}^\alpha = \Big\{ u = (u_k)_{k\in\mathbb{Z}^2},\ u_0 = 0,\ \sum_k |k|^{2\alpha}|u_k|^2 < \infty \Big\}$$

and $\mathbb{L}^2 = \mathbb{H}^0$.

Projecting (1) onto $\mathbb{L}^2$, we obtain the following system of Itô stochastic equation

$$du(x,t) + \nu\Lambda^2 u(x,t)dt = B(u,u)dt + \sigma(u)dW(x,t) \ . \tag{14}$$

Recall $\sigma(u)dW$ was defined in the introduction. Since only a finite number of modes are forced and the $\sigma_k(u)$ are uniformly bounded from above and below, equation (14) possess a global solution almost surely. It can be shown that the solution is almost surely continuous in $\mathbb{L}^2$ and in $\mathbb{H}^2$ at any moment after the initial moment. (In fact since we force only a finite number of modes, it can be shown that only the $|u_k|$ decay exponentially in $|k|$. See [Mat98].) We will take the state space of (14) to be $\mathbb{L}^2$ equipped with the Borel σ-algebra and write $u(t,\omega;u_0)$ for the solution at time t, noise realization ω and initial condition u_0. $P_t(u_0,\,\cdot\,)$ will denote the measure induced on $\mathbb{L}^2$ by the dynamics starting from u_0.

4.1 The Splitting into High and Low Modes

For this section and section 4.2, we momentarily allow forcing with an infinite number of modes. That is, we briefly allow the set of forced modes, $\mathcal{W}$, to be an infinite subset of $\mathbb{Z}^2$. We do this to make clear the generality under which our construction is valid. Of course, this raises delicate questions of uniqueness. Since the case of infinite $\mathcal{W}$ is an aside, we simply assume that σ has been chosen to insure the existence of a unique solution which is almost surely in $C\big([0,\infty),\mathbb{H}^2\big)$. (See [Fla94, DPZ96, Fer97, Mat98] for discussions of these issues.)

Fixing a finite subset $\mathcal{Z}$ of $\mathbb{Z}^2$, we define the splitting

$$\mathbb{L}^2_\ell = \text{span}\{e_k, k \in \mathcal{Z}\}, \quad \mathbb{L}^2_h = \text{span}\{e_k, k \notin \mathcal{Z}\} \ .$$

We will say that a splitting is elliptic if $k \in \mathcal{Z}$ implies that $k \in \mathcal{W}$.

Similarly, we also split the probability space into $\Omega = \Omega_\ell \times \Omega_h$. Ω_ℓ is generated by the increments of the Brownian motions w_k with $k \in \mathcal{Z}$ and Ω_h by those outside of $\mathcal{Z}$. We will write $\omega = (\xi,\zeta) \in \Omega_\ell \times \Omega_h$. As before we define the inner diameter of $\mathcal{Z}$, denoted $|\mathcal{Z}|_{int}$, by

$$|\mathcal{Z}|_{int} = \sup\{N : k \in \mathcal{Z} \text{ for all } k \text{ with } 0 < |k| \le N\ \} \ .$$

With "high/low" splitting, we can decompose (14) into

$$d\ell(t) = \big[-\nu\Lambda^2\ell + P_\ell B(\ell,\ell)\big]\,dt + P_\ell G(\ell,h)dt + \sigma_\ell(\ell,h)d\beta(t,\xi) \tag{15}$$

$$dh(t) = \big[-\nu\Lambda^2 h + P_h B(h,h)\big]\,dt + P_h G(h,\ell)dt + \sigma_h(\ell,h)d\eta(t,\zeta) \tag{16}$$

where $G(f,g) \overset{\text{def}}{=} B(g,f) + B(f,g) + B(g,g)$, P_ℓ and P_h are respectively the projection onto $\mathbb{L}^2_\ell$ and $\mathbb{L}^2_h$, $\sigma_\ell d\beta = P_\ell \sigma dW$, and $\sigma_h d\eta = P_h \sigma dW$. In defining the noise, we have used the fact that $W(t)$ is independent on the high and low mode space. In principle, this is not essential. In a more general setting, we would set $\eta = P_h W$ and $\beta = P_\ell \mathbb{E}\{W|\eta\}$.

We now consider (16) from a slightly different point of view. We can consider the $\ell(t)$ as some specified exogenous forcing. Hence if we are given some fixed trajectory $\ell(t)$ then we can solve (16) for any initial condition η-almost surely. We will write $\Phi^\eta_{t,s}(\ell,h_0)$ for solution to (16) at time t with initial condition h_0 at time s, noise realization η, and low mode forcing ℓ. Of course $\Phi^\eta_{t,s}(\ell,h_0)$ depends on the trajectory of ℓ η on $[s,t]$. When the starting time $s = 0$, we will write $\Phi^\eta_t(\ell,h_0)$ for $\Phi^\eta_{t,0}(\ell,h_0)$.

The decomposition presented here is discussed more detail in [Mat99] and [EMS01].

4.2 The Enslaving for Navier Stokes

For any $\epsilon \in (0,1)$ and $\delta > 0$, define the following set of nice future trajectories $\mathcal{U}^{\epsilon,\delta}$.

$$\mathcal{U}^{\epsilon,\delta} = \left\{ f \in C\big([0,\infty);\mathbb{L}^2\big) : \exists K' \text{ with } |f(t)|^2_{\mathbb{L}^2} + 2\nu\epsilon \int_0^t |\Lambda f(s)|^2_{\mathbb{L}^2}\, ds \le K' + (1+\delta)\mathcal{E}_0^* t,\ \forall t \ge 0 \right\}$$

The paths are "nice" because we have good control over their growth. In particular, observe that for all $f \in \mathcal{U}^{\epsilon,\delta}$,

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t |\Lambda f(s)|^2_{\mathbb{L}^2} \le \frac{1+\delta}{\epsilon} \frac{\mathcal{E}_0^*}{2\nu}\ . \tag{17}$$

These paths are typical in the following sense. Let $\mathbb{P}\{\,\cdot\,|\,\eta\}$ the probability measure obtained by conditioning on particular realization of high mode noise η.

Lemma 4.1. *For any initial condition $u_0 \in \mathbb{L}^2$, $\epsilon \in (0,1)$ and $\delta > 0$, $\mathbb{P}\{u(\cdot,\omega;u_0) \in \mathcal{U}^{\epsilon,\delta}\} = 1$. Furthermore, for almost every η, $\mathbb{P}\{u(\cdot,\omega;u_0) \in \mathcal{U}^{\epsilon,\delta}\,|\,\eta\} = 1$.*

Proof of Lemma 4.1. From Lemma A.2 of the appendix it is clear that $\mathbb{P}\{\varphi_t^\omega u_0 \in \mathcal{U}^{\epsilon,\delta}\} = 1$. By Fubini the conditional expectations are well defined for almost every η. Since the set has full measure for the full expectation it must have full measure for almost every conditional expectation. □

We now make a critical assumption on the structure of σ. We assume that σ_h depends only on ℓ. It is the key to the analysis proceeding in its present form.

Lemma 4.2. *Let $\mathcal{C}$ be the same constant as in Theorem 1. Assume that σ_h is independent of h and $\epsilon|\mathcal{Z}|^2_{int} > \mathcal{C}\frac{\mathcal{E}_0^*}{\nu^3}(1+\delta)$ for some $\epsilon \in (0,1]$ and $\delta > 0$. Then there exists a positive constant K and γ, depending on ϵ and δ, so the following holds:*

Given a $\ell \in C\big([0,t];\mathbb{L}^2_\ell\big)$, a $h_0 \in \mathbb{L}^2_h$, and a high mode noise trajectory η such that $\ell(s) + \Phi^\eta_s(\ell,h_0) \in \mathcal{U}^{\epsilon,\delta}$ then for any $\tilde{h}_0 \in \mathbb{L}^2_h$

$$\left|\Phi^\eta_s(\ell,h_0) - \Phi^\eta_s(\ell,\tilde{h}_0)\right|^2_{\mathbb{L}^2} \le K \left|h_0 - \tilde{h}_0\right|^2_{\mathbb{L}^2} e^{-\gamma s}$$

for any $s \in [0,t]$.

Proof of Lemma 4.2 . For $\epsilon = 1$ this is Lemma 2.2 from [EMS01]. Though the exponential convergence is not explicitly in the statement, it can be found in the proof. The fact that we allow the high modes to be forced is not of concern as the proof deals with the differences of two solutions; and hence, the noise cancels out because σ_h depends only on the low modes. The proof for smaller ϵ is identical. It simply uses the fact that $|\mathcal{Z}|^2_{int}$ is strictly greater than $\frac{\mathcal{E}_0^*}{\nu^3}$. □

4.3 The Main Result: Theorem 1

We now return to the setting where only a finite number of modes are forced and begin examining the principle result of this paper: Theorem 1. To make the statement of Theorem 1 complete we need to define the norm in which we are measuring the convergence. The norm on signed measures $\|\cdot\|_*$ is simply the norm defined in (4) for the class of measurable test functions $\mathcal{G}_* = \big\{g : \mathbb{L}^2_\ell \times \mathbb{L}^2_h \to \mathbb{R} : \sup|\mathsf{g}(\ell,h)| \le 1 \text{ and } |\mathsf{g}(\ell,h_1) - \mathsf{g}(\ell,h_2)| \le |h_1 - h_2|_{\mathbb{L}^2}\big\}$. This norm is analogous to the $|\cdot|_{\mathcal{G}}$-norm defined in section 3. In particular, convergence under $|\cdot|_{\mathcal{G}}$ implies convergence under the Wasserstein distance. And if the test function depends only on the low modes then it implies convergence in total variation.

After some work, Theorem 1 will be a consequence of Theorem 5. The proof is contained in the next section and is composed of a number of steps. The first step is to factor the transition kernel induced by the dynamics of the Navier-Stokes equation and then reconstitute it to build the $\mathbf{s}^k$ and $\mathbf{r}^k$ kernels from Assumption A2. The second step is to obtain the analytic estimates on the $\mathbf{s}^k$ needed to verify assumptions A3-A5. Lastly we prove the existence of the Lyapunov structure required by assumption A8.

As a preliminary step, we fix the splitting by taking the splitting set to be $\mathcal{W}$. In the notation of section 4.1, we take $\mathcal{Z} = \mathcal{W}$. We use the notation of section 4.1 decomposing $\mathbb{L}^2 = \mathbb{L}^2_\ell \times \mathbb{L}^2_h$. For the remainder or discussion we fix a $\epsilon_0 \in (0,1)$ and $\delta > 0$ so that $\epsilon_0|\mathcal{Z}|_{int} > (1+\delta)\mathcal{C}\frac{\mathcal{E}_0^*}{\nu^3}$. This is always possible since we have assumed that $|\mathcal{Z}|_{int} > \mathcal{C}\frac{\mathcal{E}_0^*}{\nu^3}$. We set $\mathcal{U} = \mathcal{U}^{\epsilon_0,\delta}$, where $\mathcal{U}^{\epsilon,\delta}$ was defined in the last section.

Since we now are only forcing the "low modes" the split system (15) and (16) can be written as

$$d\ell(t) = \left[-\nu\Lambda^2\ell + P_\ell B(\ell,\ell)\right]dt + P_\ell G(\ell,h)dt + \sigma(\ell)dW(t,\omega) \tag{18}$$

$$\frac{\partial h(t)}{\partial t} = -\nu\Lambda^2 h + P_h B(h,h) + P_h G(h,\ell)\ . \tag{19}$$

Since the high modes are now deterministic given a low mode trajectory, we simplify the notation of section 4.1. We will write $\Phi_t(\ell,h_0)$ for the solution to (19) with specified low modes ℓ and high mode initial condition h_0.

Step 1: Setup and Preliminaries

Though our process is continuous, we will think of solutions of (18)-(19) as a discrete chain where each step will be a segment of trajectory. For the remainder of this discussion, we fix a $T > 0$ and set $\mathbb{X} = C([0,T],\mathbb{L}^2)$, $\mathbb{X}_+ = C([0,T],\mathbb{L}^2_\ell)$, and $\mathbb{X}_- = C([0,T],\mathbb{L}^2_h)$. We define $P(x,\cdot)$ as the measure induced on $\mathbb{X}$ by starting from an initial condition x. We will allow x to be in either $\mathbb{X}$ or $\mathbb{L}^2$, this is reasonable as one only needs to evolve forward in time last position in $\mathbb{L}^2$. As before we define Q^n as the measure induced on $\mathbb{X}^n$ by taking n steps of the chain, Q^n_+ as its projection onto $\mathbb{X}^n_+$, and P_+ as the projection of P onto $\mathbb{X}_+$. Of course in our current setting, we can also view Q^n and Q^n_+ as measures on $C([0,nT],\mathbb{L}^2)$ and $C([0,nT],\mathbb{L}^2_\ell)$ respectively. We define the function $\Phi : \mathbb{X}_+ \times \mathbb{X} \to \mathbb{X}_-$ by $y = \Phi(z,x)$ with $y(s) = \Phi_s(z,x(T))$ with $s \in [0,T]$. Next define $\mathbf{Q}_+(x^{(1)},x^{(2)},\cdot)$ to be the product measure $Q_+(x^{(1)},\cdot) \times Q_+(x^{(2)},\cdot)$. Clearly this choice satisfies assumption A1.

We begin by establishing a number of constants, whose values will be fixed later, and using them to define a number of subsets in the pathspace which will be needed in our constructions. For any positive K_0, K_1, and C_1 we define the following sets. (Recall that ϵ_0 and δ were positive constants fixed once and for all in the last section.) In general "+" will adorn subsets of the active variable and boldface will be used for sets which contain pairs of trajectories.

First set $A_1^+ = \{z \in \mathbb{X}_+ : \sup_{t\in[0,T]} |z(t)|_{\mathbb{L}^2}^2 \le C_1\}$. For $k = 2, 3, \dots$ and $x = (x_0, \dots, x_{k-1}) \in \mathbb{X}^k$, we define

$$A_k^+(x) = \Bigg\{ z_k \in \mathbb{X}_+ : |x_k(t)|_{\mathbb{L}^2}^2 + 2\nu\epsilon_0 \sum_{j=2}^{k} \int_0^{T_j(k)} |\Lambda x_j(s)|_{\mathbb{L}^2}^2 \, rs \le K_1 + (1+\delta)\mathcal{E}_0^*[(k-2)T + t] \text{ for all } t \in [0,T] \Bigg\}$$

where $x_k = \Upsilon(z_k, \Pi_- x_{k-1})$ and $T_j(k) = t\mathbf{1}_{\{k\}}(j) + T\mathbf{1}_{\{j<k\}}(j)$. Notice that the definition of A_1^+ does not depend on a base point x while the higher k definitions do. These sets are subsets of $\mathbb{X}_+$. Next we string them together to define subsets of the sequence space. For $x = (x_0, \dots, x_{m-1}) \in \mathbb{X}^m$ we define $A_+^{m,k}(x) \subset \mathbb{X}_+^{k-m-1}$ as

$$A_+^{m,k}(x) = \Big\{ (z_m, \cdots, z_k) \in \mathbb{X}_+^{k-m-1} : z_l \in A_l^+(x_0, \cdots, x_{l-1}) \text{ for } l = m \dots k \text{ where } x_j = (z_j, y_j) \text{ and } y_j = \Phi(z_j, x_{j-1}) \text{ for } j = m, \dots, k \Big\} . \quad (20)$$

When $m = 1$, we will simply write $A_+^k(x)$. Observe that we continue the convention from the previous sections of subscripts denoting objects having to do with the k-element in a sequence and superscripts denoting objects associated with a path of length k. Notice that by construction if one is sitting at $x \in \mathbb{X}$ then $A_+^k(x)$ is precisely the choices of active variable which keeps one in the set of "nice" futures $\mathcal{U}$ from the previous section for k steps.

Finally we build a particular set of pairs of paths in the product path space. These will be the paths on which our "**s**" measures in the factorization will be concentrated. For $x^{(i)} = (x_0^{(i)}, \dots, x_{k-1}^{(i)}) \in \mathbb{X}^k$ with $i = 1, 2$, we define $A_k^+(x^{(1)}, x^{(2)}) = A_k^+(x^{(1)}) \cap A_k^+(x^{(2)})$. Then for $k \ge 1$ and $x_0^{(i)} \in A_0\chi$, we define

$$\mathbf{A}_+^k(x_0^{(1)}, x_0^{(2)}) = \Big\{ (z^{(1)}, z^{(2)}) \in \mathbb{X}_+^k \times \mathbb{X}_+^k : \text{ where } z^{(i)} \in A_+^{1,k}(x_0^{(1)}, x_0^{(2)}), z_1^{(1)}(T) = z_1^{(2)}(T) \text{ and if } k \ge 2, z_m^{(1)} = z_m^{(2)} \text{ for } m \in \{2, \dots, k\} \Big\} .$$

These will be our "similar" paths in the product of the active path spaces. For these sets to be useful we have to show that we can construct a coupling (a measure on the product space) so that these sets have positive measure and in addition if these paths are followed the system has desirable properties. The latter point will be the topic of the next section while the first will be covered the following sections.

Before continuing, we define two last set of paths. Namely the paths in the whole phase space determined by following our disguised active mode paths and a group of nice initial conditions. In particular, define

$$\mathbf{A}^k(x_0^{(1)}, x_0^{(2)}) = \Big\{(\hat{x}^{(1)}, \hat{x}^{(2)}) \in \mathbb{X}^k \times \mathbb{X}^k \text{ with } \hat{x}^{(i)} = (x_0^{(i)}, \dots, x_{k-1}^{(i)}) : \exists\, (z^{(1)}, z^{(2)}) \in \mathbf{A}_+^k(\hat{x}_0^{(1)}, \hat{x}_0^{(2)}) \\ \text{with } \Pi_+ \hat{x}_j = z_j \text{ and } \Pi_- \hat{x}_j = y_j \text{ where } y_j^{(i)} = \Phi(z_j^{(i)}, x_{j-1}^{(i)})\Big\}$$

Next we define a set of nice initial conditions by $A_0 = \{x \in \mathbb{X} : |x(T)|_{\mathbb{L}^2}^2 \le K_0\}$ and

$$\mathbf{A}^{0,k} = \Big\{(x_0^{(1)}, \dots, x_k^{(1)}) \times (x_0^{(2)}, \dots, x_k^{(2)}) : x_0^{(i)} \in A_0 \text{ and } \\ (x_1^{(1)}, \dots, x_k^{(1)}) \times (x_1^{(2)}, \dots, x_k^{(2)}) \in \mathbf{A}^k(x_0^{(1)}, x_0^{(2)})\Big\}$$

Step 2: The sets A_+^k have probability close to one

This will be obtained by choosing the various constants. Because we know that the moments of the norm of the trajectory are controlled by the norm of the initial condition, for any $\epsilon > 0$ we can pick a C_1 so that $\mathbb{P}\{\sup_{t\in[0,T]} |z_1(t)|^2 \le C_1\} \ge 1 - \frac{\epsilon}{2}$ for any $x_0 \in A_0$ and $z_1(t) = P_\ell \varphi_t x_0(T)$. Then by Lemma D.1 the fact that the low modes are less that C_1 on $[0,T]$ implies that there is some bound K_1' so that the norm of the total solution is bounded by K_1' over the interval $[0,T]$. Lastly we pick $K_1 \ge K_1'$ so that for any $x_1 \in \mathbb{X}$ with $|x_1(T)|_{\mathbb{L}^2}^2 \le K_1'$, we have

$$\mathbb{P}\Bigg\{ |x_k(t)|_{\mathbb{L}^2}^2 + 2\nu\epsilon_0 \sum_{j=2}^{k} \int_0^{T_j(k)} |\Lambda x_j(s)|_{\mathbb{L}^2}^2\, ds \le K_1 + (1+\delta)\mathcal{E}_0^*[(k-2)T + t] + \\ \text{for all } t \in [0,T] \text{ and } k \in \{2, 3, \cdots\} \Bigg\} > 1 - \frac{\epsilon}{2}$$

Combining the two estimates give that for any $x_0 \in A_0$, $\mathbb{P}\{z(\cdot) \in A^k(x_0)\} > 1 - \epsilon$ for all k. After these considerations the only free parameters in our construction are ϵ and the K_0 used to define A_0.

Step 3: Contractive Properties

For paths in $\mathbf{A}^{0,k}$, Lemma 4.2 describing the high mode contraction has the following simple form.

Corollary 4.3. *There exist positive constants K' and γ' such that for $k \ge 2$ and $n \in \{2, \dots, k\}$*

$$\sup_{(x^{(1)}, x^{(2)}) \in \mathbf{A}^{0,k}} \left(\sup_{t \in [0,T]} \left| y_n^{(1)}(t) - y_n^{(2)}(t) \right|_{\mathbb{L}^2}^2 \right) \le K' e^{-\gamma' n T} \tag{21}$$

where $y_n^{(i)} = \Pi_- \pi_n x^{(i)}$.

Proof of Corollary 4.3. Fix any $(x^{(1)}, x^{(2)}) \in \mathbf{A}^{0,k}$. Because $(\pi_{[1,k]}x^{(1)}, \pi_{[1,k]}x^{(2)}) \in \mathbf{A}^k(x_0^{(1)}, x_0^{(2)})$ with $\pi_0 x^{(i)} \in A_0$, we know that $\pi_1 x^{(i)} \in A_1$ and $\Pi_+\pi_n x^{(i)} = \Pi_+\pi_n x^{(i)}$ for $n \in \{2, \dots, k\}$. Then Lemma 4.2 implies that

$$\sup_{t\in[0,T]} \left|y_n^{(1)}(t) - y_n^{(2)}(t)\right|_{\mathbb{L}^2}^2 \le K' \left|y_1^{(1)}(T) - y_1^{(2)}(T)\right|_{\mathbb{L}^2}^2 e^{-\gamma'(n-2)T}$$

for $n = 2, \dots, k$ and some fixed positive K' and γ', independent of the initial conditions or n. We know that $\left|y_1^{(1)}(T) - y_1^{(2)}(T)\right|_{\mathbb{L}^2}^2 \le 2K_1$ from the definition of A_1^+ and the choices of constants made in the last section. Hence for new positive constants K' and γ', we have stated result. □

Step 4: Deconstruction and Reconstruction of the Measures

We now build the factorizations $\mathbf{s}_k$ and $\mathbf{r}_k$ as needed to satisfy assumption A2. For $k \ge 2$, we proceed in a fashion similar to section 3. However, we do not have the luxury of using the same factorization of all k as we did in section 3. For $k = 1$, we will do something slightly more complicated but in the same spirit.

Given $k \ge 2$, we define

$$\begin{aligned} s_k(x^{(1)}, x^{(2)}, \,\cdot\,) =& Q_+\big(\pi_{k-1}x^{(1)}, \,\cdot\, \cap A_k^+(x^{(1)}, x^{(2)})\big) \wedge Q_+\big(\pi_{k-1}x^{(2)}, \,\cdot\, \cap A_k^+(x^{(1)}, x^{(2)})\big) \\ r_k(x^{(1)}, x^{(2)}, \,\cdot\,) =& \Big[Q_+\big(\pi_{k-1}x^{(1)}, \,\cdot\, \cap A_k^+(x^{(1)}, x^{(2)})\big) - Q_+\big(\pi_{k-1}x^{(2)}, \,\cdot\, \cap A_k^+(x^{(1)}, x^{(2)})\big)\Big]^+ \\ &+ Q_+\big(\pi_{k-1}x^{(1)}, \,\cdot\, \cap A_k^+(x^{(1)}, x^{(2)})^c\big) \end{aligned}$$

Observe that $Q_+(\pi_{k-1}x^{(1)}, \,\cdot\,) = s_k(x^{(1)}, x^{(2)}, \,\cdot\,) + r_k(x^{(1)}, x^{(2)}, \,\cdot\,)$ and $s_k(x^{(1)}, x^{(2)}, \,\cdot\,) = s_k(x^{(2)}, x^{(1)}, \,\cdot\,)$. Next we define, $\mathbf{s}_k(x^{(1)}, x^{(2)}, \,\cdot\,)$ to be the measure on $\mathbb{X}_+ \times \mathbb{X}_+$ concentrated on the diagonal elements (ζ, ζ) where ζ is distributed according to $s_k(x^{(1)}, x^{(2)}, \,\cdot\,)$. Lastly as in the simple map example of section 3, we define

$$\mathbf{r}_k(x^{(1)}, x^{(2)}, \,\cdot\,) = \frac{r_k(x^{(1)}, x^{(2)}, \,\cdot\,) \times r_k(x^{(2)}, x^{(1)}, \,\cdot\,)}{r_k(x^{(1)}, x^{(2)}, \mathbb{X})}$$

again with the convention that $\frac{0}{0} = 0$.

As mentioned at the onset, we only use the above construction for $k \ge 2$. This is because $Q_+\big(x^{(1)}, \,\cdot\,\big)$ is singular with respect to $Q_+\big(x^{(2)}, \,\cdot\,\big)$ whenever $x^{(1)}(T) \ne x^{(2)}(T)$. Hence for $k = 1$, we can not compare measures on the entire path space, instead we look at measures induce on $\mathbb{L}_\ell^2$ at time T. More precisely for any $B \subset \mathbb{L}_\ell^2$, define

$$[B] = \Big\{z \in \mathbb{X}_+ = C\big([0,T], \mathbb{L}_\ell^2\big) : z(T) \in B\Big\} \; .$$

We set

$$\begin{aligned} v_{A_1}(x, B) &= Q_+(x, [B] \cap A_1^+) \; , \qquad (22) \\ v^s(x^{(1)}, x^{(2)}, \,\cdot\,) &= v_{A_1}(x^{(1)}, \,\cdot\,) \wedge v_{A_1}(x^{(2)}, \,\cdot\,) \text{ and,} \\ v^r(x^{(1)}, x^{(2)}, \,\cdot\,) &= [v_{A_1}(x^{(1)}, \,\cdot\,) - v_{A_1}(x^{(2)}, \,\cdot\,)]^+ \; . \end{aligned}$$

Next we let $Q_+^{A_1}(x,\,\cdot\,|\ell)$ be the probability measure induced on $\mathbb{X}_+ = C\big([0,T],\mathbb{L}^2_\ell\big)$ by $Q_+(x,\,\cdot \cap A_1^+)$ conditioned on the trajectory being ℓ at time T. Also define $\mathbf{Q}_+^{A_1}(x^{(1)},x^{(2)},\,\cdot\,|\ell) = Q_+^{A_1}(x^{(1)},\,\cdot\,|\ell) \times Q_+^{A_1}(x^{(2)},\,\cdot\,|\ell)$. Lastly, we define

$$\begin{aligned}
\mathbf{s}_1(x^{(1)},x^{(2)},\,\cdot\,) &= \int \mathbf{Q}_+^{A_1}(x^{(1)},x^{(2)},\,\cdot\,|\ell) v^s(x^{(1)},x^{(2)},d\ell)\\
\mathbf{r}_1(x^{(1)},x^{(2)},\,\cdot\,) &= \frac{r_1(x^{(1)},x^{(2)},\,\cdot\,) \times r_1(x^{(2)},x^{(1)},\,\cdot\,)}{r_1(x^{(1)},x^{(2)},\mathbb{X}_+)}\\
r_1(x^{(1)},x^{(2)},\,\cdot\,) &= Q_+(x^{(1)},\,\cdot\, \cap (A_1^+)^c) + \int Q_+^{A_1}(x^{(1)},\,\cdot\,|\ell) v^r(x^{(1)},x^{(2)},d\ell)
\end{aligned}$$

The critical feature of this construction is that if $(\zeta^{(1)},\zeta^{(2)})$ is in the support of $\mathbf{s}_1$ then $\zeta^{(1)}(T) = \zeta^{(2)}(T)$ and $|\zeta^{(i)}(t)|^2 \le C_1$ for $t \in [0,T]$. Recalling that by definition $\mathbf{s}^k = \mathbf{s}_k \cdots \mathbf{s}_1$, if $(\zeta^{(1)},\zeta^{(2)})$ is in the support of $\mathbf{s}^k$ with $k \ge 2$ then $\zeta^{(i)}$ can be viewed as a trajectory in $C\big([0,kT],\mathbb{L}^2_\ell\big)$. Viewing them as such $\zeta^{(1)}(t) = \zeta^{(2)}(t)$ for $t \in [T,kT]$ and $|\zeta^{(i)}|^2 \le C_1$ for $t \in [0,T]$.

Step 5: The lower bounds on ρ^k

Recall that we defined $\rho^k(x^{(1)},x^{(2)}) = \mathbf{s}^k(x^{(1)},x^{(2)},\mathbb{X}_+^k \times \mathbb{X}_+^k)$. To connect with Theorem 5, we need to establish that the ρ^k are each uniformly bounded away from zero of initial conditions with $x_0^{(i)} \in A_0$. This is the content of Assumption A2.

We begin by considering ρ^1. All we desire to show is that there exist some positive constant ρ_*^1 so that $\rho^1(x^{(1)},x^{(2)}) \ge \rho_*^1$ for all $x^{(i)} \in A_0$. (The constant ρ_*^1 will depend on our choice of K_0 which was used to define A_0. From the construction we see that $\rho^1(x^{(1)},x^{(2)}) = v^s(x^{(1)},x^{(2)},\mathbb{L}^2_\ell) = \|\upsilon_{A_1}(x^{(1)},\,\cdot\,) \wedge \upsilon_{A_1}(x^{(2)},\,\cdot\,)\|_{TV}$. We need to show that there exists a ρ_*^1 so $\rho^1(x^{(1)},x^{(2)}) \ge \rho_*^1$ for all $x^{(i)} \in A_0$. But this follows from Lemma C.1 because $Q_+(x^{(i)},A_1^+)$ is uniformly bounded from below by construction and Lemma 4.5 shows that the needed density moment is uniformly bounded from above.

We now turn to ρ^k for $k \ge 2$. Observe that for fixed $(x_0^{(1)},x_0^{(2)}) \in A_0$,

$$\rho^k(x_0^{(1)},x_0^{(2)}) = \int_{\mathbb{X}\times\mathbb{X}} \mathbf{s}^1(x_0^{(1)},x_0^{(2)},d\zeta^{(1)} \times d\zeta^{(2)}) \mathbf{s}^{2,k}(\chi^{(1)},\chi^{(2)},\mathbb{X}_+^{k-1} \times \mathbb{X}_+^{k-1})\ .$$

where $\chi^{(i)} = (x_0^{(i)},x_1^{(i)})$, $i=1,2$, with $x_1^{(i)} = (\zeta^{(i)},y^{(i)})$ and $y^{(i)} = \Phi(\zeta^{(i)},x_0^{(i)})$. Next note that $\mathbf{s}^1(x_0^{(1)},x_0^{(2)},\mathbb{X}\times\mathbb{X}) = v^s(x_0^{(1)},x_0^{(2)},\mathbb{L}^2_\ell)$ and that by the construction of v^s, $\zeta^{(1)}(T) = \zeta^{(2)}(T)$ and $\zeta^{(1)},\zeta^{(2)} \in A_1^+$. Recalling the definition of $\mathbf{A}^{0,1}$ from the last section we observe that

$$\rho^k(x_0^{(1)},x_0^{(2)}) \ge \big[v^s(x_0^{(1)},x_0^{(2)},\mathbb{L}^2_\ell)\big] \inf_{(\chi^{(1)},\chi^{(2)})\in\mathbf{A}^{0,1}} \mathbf{s}^{2,k}(\chi^{(1)},\chi^{(2)},\mathbb{X}_+^{k-1} \times \mathbb{X}_+^{k-1})$$

From the calculation for $k=1$, we know that the first factor is bounded from below by ρ_*^1. Hence we only need a lower bound on the second factor. Recalling the definition of $\mathbf{s}^k$ for $k \ge 2$, Lemma C.1 implies that to obtain the uniform lower bound on ρ^k it is sufficient to have an upper bound on

$$\sup_{(x^{(1)},x^{(2)})\in\mathbf{A}^{0,1}} \mathbb{E}^{(2)} \left(\frac{dQ_+^{k-1}\big(x_1^{(1)},\cdot \cap A_+^k(x_1^{(1)},x_1^{(2)})\big)}{dQ_+^{k-1}\big(x_1^{(2)},\cdot \cap A_+^k(x^{(1)},x^{(2)})\big)} \right)^4 \tag{23}$$

and a lower bound on

$$\inf_{(x^{(1)},x^{(2)})\in\mathbf{A}^{0,1}} Q_+^{k-1}(x_1^{(1)}, \cdot \cap A^k(x_1^{(1)}, x_1^{(2)})) \ .$$

The infimum is bounded from below by $1-\epsilon$ by construction. A bound on the supremum follows from (26) of Lemma 4.4.

In summary, have shown that for all k there exists a positive ρ_*^k so that $\rho^k(x_0^{(1)}, x_0^{(2)}) \geq \rho_*^k$ for all $x_0^{(i)} \in A_0$.

Step 6: $\rho^{k-1} - \rho^k$ decays exponentially

The second condition needed to connect with Theorem 5 is that the $\rho^{k-1}(x^{(1)}, x^{(2)}) - \rho^k(x^{(1)}, x^{(2)})$ decay, uniformly in x, at an rate exponential in k. This is the content of Assumption A7. We will show that $\rho^k(x_0^{(1)}, x_0^{(2)}) - \rho^{k+1}(x_0^{(1)}, x_0^{(2)})$ decays exponentially in k with constants uniform over initial datum in A_0.

Begin by fixing $x^{(i)} \in A_0$. From Corollary 2.2, we have that $\rho^k(x^{(1)}, x^{(2)}) - \rho^{k+1}(x^{(1)}, x^{(2)}) = \mathbf{r}_{k+1}\mathbf{s}^k(x^{(1)}, x^{(2)}, \mathbb{X}^{k+1} \times \mathbb{X}^{k+1})$. To simplify notation we will denote

$$r_{A_k}(x^{(1)}, x^{(2)}, \cdot) = [Q_+(\pi_k x^{(1)}, , \cdot \cap A_k^+(x^{(1)}, x^{(2)})) - Q_+(\pi_k x^{(2)}, \cdot \cap A_k^+(x^{(1)}, x^{(2)})]^+$$
$$r_{!A_k^+}(x^{(1)}, x^{(2)}, \cdot) = Q_+(x^{(1)}, \cdot \cap A_k^+(x^{(1)}, x^{(2)})^c) \ .$$

With this notation $r_k = r_{A_k} + r_{!A_k}$. Next observe that

$$\begin{aligned}\rho^k(x^{(1)}, x^{(2)}) - \rho^{k+1}(x^{(1)}, x^{(2)}) &= \mathbf{r}_{k+1}\mathbf{s}^k(x^{(1)}, x^{(2)}, \mathbb{X}^{k+1} \times \mathbb{X}^{k+1}) \\ &\leq \sup_{x^{(1)},x^{(2)}\in A_0} r_{k+1}\mathbf{s}^k(x^{(1)}, x^{(2)})\end{aligned}$$

Where

$$r_{k+1}\mathbf{s}^k(x^{(1)}, x^{(2)}) = \int_{\mathbb{X}_+^k\times\mathbb{X}_+^k} r_{k+1}(\chi^{(1)}, \chi^{(2)}, \mathbb{X}_+)\mathbf{s}^k(x^{(1)}, x^{(2)}, dz^{(1)} \times dz^{(2)})$$

with $\chi^{(i)} = \Upsilon(z^{(i)}, x^{(i)})$. (Notice that the r_{k+1} is not bold !) Hence, we have

$$\begin{aligned}\rho^k(x^{(1)}, x^{(2)}) - \rho^{k+1}(x^{(1)}, x^{(2)}) \leq \sup_{x^{(1)},x^{(2)}\in A_0} & r_{!A_{k+1}}\mathbf{s}^k(x^{(1)}, x^{(2)}) \\ &+ \sup_{x^{(1)},x^{(2)}\in A_0} r_{A_{k+1}}\mathbf{s}^k(x^{(1)}, x^{(2)})\end{aligned}$$

where $r_{!A_{k+1}}\mathbf{s}^k$ and $r_{A_{k+1}}\mathbf{s}^k$ are defined as above.

Intuitively, the first term is small because it is unlikely that the trajectory is outside of A_{k+1}^+ on the k-th step given that it started in A_0. Intuitively, the second term is small because there is little probability mass left in $r_{A_{k+1}}$ which is essentially difference of the two transition densities on the $k+1$-th step after agreeing on the first k steps. The contractive properties of the system make the density on the $k+1$th step close.

We begin with the first term.

$$\sup_{x^{(1)},x^{(2)}\in A_0} r_{!A_{k+1}}\mathbf{s}^k(x^{(1)}, x^{(2)}) \leq \sup_{x_0\in A_0} \mathbb{P}_{x_0}\left\{\sup_{0\leq t\leq T} |x_k(t)|_{\mathbb{L}^2}^2 > K_0 + (1+\delta)\mathcal{E}_0^*(k+1)T\right\}$$

By Lemma A.2 the last term decays exponentially, hence there exist positive constants K and γ so that

$$\sup_{x^{(1)},x^{(2)}\in A_0} r_{!A_{k+1}}\mathbf{s}^k(x^{(1)},x^{(2)}) \le Ke^{-\gamma k}$$

We now turn to the other term. Since $\mathbf{s}^k(x^{(1)},x^{(2)},\mathbb{X}_+^k\times\mathbb{X}_+^k)=\mathbf{s}^k(x^{(1)},x^{(2)},\mathbf{A}_+^k(x^{(1)},x^{(2)}))$,

$$\begin{aligned}\sup_{x^{(1)},x^{(2)}\in A_0} r_{A_{k+1}}\mathbf{s}^k(x^{(1)},x^{(2)}) &\le \sup_{(x^{(1)},x^{(2)})\in\mathbf{A}^{0,k}} r_{A_{k+1}}(x^{(1)},x^{(2)})\\ &\le \sup_{(x^{(1)},x^{(2)})\in\mathbf{A}^{0,k}} \mathbb{E}^{(2)}\left[\frac{dQ_+(\pi_k x^{(1)},\cdot\cap A_{k+1}^+(x^{(1)},x^{(2)})}{dQ_+(\pi_k x^{(2)},\cdot\cap A_{k+1}^+(x^{(1)},x^{(2)})}-1\right]^+\end{aligned}$$

by Cauchy-Schwartz,

$$\le \sup_{(x^{(1)},x^{(2)})\in\mathbf{A}^{0,k}}\left[\mathbb{E}^{(2)}\left[\frac{dQ_+(\pi_k x^{(1)},\cdot\cap A_{k+1}^+(x^{(1)},x^{(2)}))}{dQ_+(\pi_k x^{(2)},\cdot\cap A_{k+1}^+(x^{(1)},x^{(2)}))}-1\right]^2\right]^{\frac{1}{2}}$$

and by expression (27)

$$\le \left[\exp\left(Ke^{-\gamma k}\right)-1\right]^{\frac{1}{2}}$$

for possible larger K and γ. Putting all of the estimate together produces

$$\sup_{x^{(i)}\in A_0}\rho^k(x_0^{(1)},x_0^{(2)})-\rho^{k+1}(x_0^{(1)},x_0^{(2)}) \le Ke^{-\gamma k}+\left[\exp\left(Ke^{-\gamma k}\right)-1\right]^{\frac{1}{2}} \le K'e^{-\gamma' k} \tag{24}$$

for some other positive K' and γ'.

Step 7: Lyapunov Structure

The last requirement to apply Theorem 5 is the existence of a Lyapunov function. For $x\in\mathbb{X}$, we take $V(x)=|x(T)|_{\mathbb{L}^2}^2$. Let $x_n\in\mathbb{X}=C([0,T];\mathbb{L}^2)$ the solution to the Navier Stokes equation starting from initial data x_0. Because the paths are continuous almost surely we know that $x_{n-1}(T)=x_n(0)$ with probability one. The standard energy estimate applied to this setting (see [Mat99, EMS01]) produces:

$$\mathbb{E}\Big\{\,|x_n(T)|_{\mathbb{L}^2}^2\,|\mathcal{F}_{n-1}\Big\}=e^{-2\nu T}\,|x_{n-1}(T)|_{\mathbb{L}^2}^2+\frac{\mathcal{E}_0^*}{2\nu}T$$

Hence assumption A8 holds with $a=e^{-2\nu T}$ and $b=\frac{\mathcal{E}_0^*}{2\nu}T$ and

$$\mathbf{C}=\left\{(x^{(1)},x^{(2)})\in\mathbb{X}\times\mathbb{X}:\left|x^{(1)}(T)\right|_{\mathbb{L}^2}^2+\left|x^{(2)}(T)\right|_{\mathbb{L}^2}^2\le K_0\right\}. \tag{25}$$

Clearly if $(x^{(1)},x^{(2)})\in\mathbf{C}$ then $x^{(i)}\in A_0$ for both $i=1$ and $i=2$.

Step 8: Conclusion of the proof of Theorem 1

All of the assumptions needed to apply Theorem 5 have been satisfied. We summarize. In step 1, we constructed the needed $\mathbf{s}+\mathbf{r}$ decomposition. In step 2, we obtained the needed uniform bounds on the ρ's. Together step 1 and 2 establish assumptions A1, A2, and A7. Given the contraction shown in Lemma 4.2 and the fact that the test functions are continuous in the high modes, assumption A3 and A5 are satisfied. The Lyapunov structure shown in step 3 is compatible with assumption A8.

4.3.1 Estimates on the Radon-Nikodym Derivatives

We now use this estimate to control various Radon-Nikodym derivatives which we arise in the next section.

The next lemma shows how the measures on the future paths become exponentially similar if the past is taken from the support of $\mathbf{S}^k$.

Lemma 4.4. *From any integer $p \geq 1$, there exist fixed positive constants $\hat{K}(p)$ and $\hat{\gamma}(p)$ so that if $2 \leq n < k$ then*

$$\sup_{(x^{(1)},x^{(2)})\in\mathbf{A}^{0,n}} \mathbb{E}^{(2)} \left(\frac{dQ_+^{k-n}\big(\pi_n x^{(1)}, \cdot \cap A_+^{n+1,k}(x^{(1)},x^{(2)})\big)}{dQ_+^{k-n}\big(\pi_n x^{(2)}, \cdot \cap A_+^{n+1,k}(x^{(1)},x^{(2)})\big)} \right)^{2p} \leq \exp\left(\hat{K}(p)e^{-\hat{\gamma}(p)n}\right) . \qquad (26)$$

In particular, we have the estimate

$$\sup_{(x^{(1)},x^{(2)})\in\mathbf{A}^{0,n}} \mathbb{E}^{(2)} \left[1 - \frac{dQ_+^{k-n}\big(\pi_n x^{(1)}, \cdot \cap A_+^{n+1,k}(x^{(1)},x^{(2)})\big)}{dQ_+^{k-n}\big(\pi_n x^{(2)}, \cdot \cap A_+^{n+1,k}(x^{(1)},x^{(2)})\big)} \right]^{2} \leq \exp\left(\hat{K}(1)e^{-\hat{\gamma}(1)n}\right) - 1 . \qquad (27)$$

This next lemma is the key estimate to prove that density at time T have a minimal common component uniformly over all initial conditions in A_0

Lemma 4.5. *Recall the definition of $v_{A_1}(x,B)$ from (22) and assume that $C_1 > K_0$ (the constants used in defining A_1 and A_0). Then for any $p \geq 1$, there exists a constant C^* so that*

$$\sup_{x^{(1)},x^{(2)}\in A_0} \mathbb{E}^{(1)} \left(\frac{dv_{A_1}(x^{(1)}, \cdot)}{dv_{A_1}(x^{(2)}, \cdot)} \right)^p < C^*$$

Proof of Lemma 4.4. Recall that $Q_+^{k-n}(\pi_n x^{(i)}, \cdot)$ is just the measure on the low modes induced by the dynamics of (18) up to time $(k-n)T$ starting with low mode initial data $\ell_0^{(i)} = (\Pi_+ \pi_n x^{(i)})(T)$ and high mode initial condition $h_0^{(i)} = (\Pi_- \pi_n x^{(i)})(T)$. Since $(x^{(1)}, x^{(2)}) \in \mathbf{A}^{0,n}$ with $n \geq 2$, we know that $\ell_0^{(1)} = \ell_0^{(2)}$ hence we will simply write ℓ_0.

We need to compare the measures induced by

$$\begin{aligned} &d\ell^{(i)}(t) = \left[-\nu\Lambda^2 \ell^{(i)} + P_\ell B(\ell^{(i)}, \ell^{(i)})\right] dt + P_\ell G\big(\ell^{(i)}, \Phi_t(\ell^{(i)}, h_0^{(i)})\big)dt + \sigma(\ell^{(i)})d\beta(t,\xi) \\ &\ell^{(i)}(0) = \ell_0 \end{aligned}$$

for $i = 1,2$. If we define $D(t,\ell,h_0^{(1)},h_0^{(2)}) \overset{\text{def}}{=} G(\ell,\Phi_t(\ell,h_0^{(1)})) - G(\ell,\Phi_t(\ell,h_0^{(2)}))$ then $\ell^{(1)}$ is related to $\ell^{(2)}$ are require to apply Lemma B.1. Recalling that $\sigma(\ell)$ is uniformly bounded away from zero, all we need is control over $|D(t,\ell,h_0^{(1)},h_0^{(2)})|$ for initial conditions in $\mathbf{A}^{0,n}$ and ℓ trajectories in $A_+^{n+1,k}(x^{(1)},x^{(2)})$.

By Lemma E.1 and the bilinearity of $B(u,v)$,

$$|D_t(\ell,h_0^{(1)},h_0^{(2)})|^2 \le Cd(t)\left[|\ell(t)|^2 + \left|\Phi_t(\ell,h_0^{(1)})\right|_{\mathbb{L}^2}^2 + \left|\Phi_t(\ell,h_0^{(2)})\right|_{\mathbb{L}^2}^2\right] \tag{28}$$

where $d(t) = \left|\Phi_t(\ell,h_0^{(1)}) - \Phi_t(\ell,h_0^{(2)})\right|_{\mathbb{L}^2}^2$. Because our paths are restricted to $\ell \in A_+^{n+1,k}(x^{(1)},x^{(2)})$, we have

$$|\ell(t)|_{\mathbb{L}^2}^2 + \left|\Phi_t(\ell,h_0^{(1)})\right|_{\mathbb{L}^2}^2 + \left|\Phi_t(\ell,h_0^{(2)})\right|_{\mathbb{L}^2}^2 \le 2K_1 + 2(1+\delta)\mathcal{E}_0^*(nT+t)\ .$$

Since in the definition of D, $\Phi_t(\ell,h_0^{(1)})$ and $\Phi_t(\ell,h_0^{(2)})$ are evaluated with the same ℓ, Lemma 4.2 implies that

$$d(t) \le K\left|h_0 - \tilde{h}_0\right|_{\mathbb{L}^2}^2 e^{-\gamma t}$$

for some constants K and γ and all $t \ge 0$. Because $(x^{(1)},x^{(2)}) \in \mathbf{A}^{0,n}$, Corollary 4.3 tells us that

$$\left|h_0^{(1)} - h_0^{(2)}\right|_{\mathbb{L}^2}^2 \le K' e^{-\gamma' nT}\ . \tag{29}$$

for some $\gamma' \in (0,\gamma]$. Combining the two estimates give $d(t) \le K'' e^{-\gamma'(nT+t)}$.

Using these facts in (28), shows that for some K'' and γ'',

$$\sup_{(x^{(1)},x^{(2)})\in\mathbf{A}^{0,n}}\left[\sup_{\ell(\cdot)\in A_+^{n+1,k}(x^{(1)},x^{(2)})}\int_0^{(k-n)T}|D_t(\ell,h_0^{(1)},h_0^{(2)})|^2 dt\right] \le K'' e^{-\gamma n}$$

for any $k > n$. Using this estimate and Lemma B.1, we complete the proof. □

Proof of Lemma 4.5. Though the following proof is not the most elegant, it will be sufficient for our needs. We will prove the lemma by comparing the stochastic process $\ell(t)$ with

$$d\hat{\ell}(t) = \left[-\nu\Lambda^2\hat{\ell} + P_\ell B(\hat{\ell},\hat{\ell})\right]dt + \sigma(\hat{\ell})dW(t,\xi)\ .$$

This process is a standard elliptic diffusion as we have removed the memory term. We define $\hat{v}_{A_1}(x,B) = \hat{Q}_+(x,[B]\cap A_1^+) = \mathbb{P}\{\hat{\ell} \in [B]\cap A_1^+\}$. We know from the standard theory that $\hat{v}_{A_1}(x,B) = \int_B \hat{p}(x,y)d\lambda$ for some smooth density $\hat{p}(x,y)$. Here λ is the Lebesgue measure on $\mathbb{L}_\ell^2$, which is finite dimensional. In addition we know that $\hat{p}(x,\cdot)$ is positive in the interior of the ball $\{y : |y|^2 \le C_1\}$ and that for any $p > 1$

$$\sup_{x^{(1)},x^{(2)}\in A_0}\int\left[\frac{\hat{p}(x^{(1)},y)}{\hat{p}(x^{(2)},y)}\right]^p \hat{p}(x^{(2)},y)d\lambda(y) < \infty\ . \tag{30}$$

We can uniformly bound the moments of the density ratio because K_0 is strictly less than C_1 and hence we never get close to the boundary.

For a moment let us assume that we can show that $v_{A_1}(x,B)$ has a density $p(x,y)$ relative to $\lambda(y)$ and is absolutely continuous with respect to $\hat{v}_{A_1}(x,B)$. In addition, suppose that for any $p \geq 1$,

$$\sup_{x\in A_0} \int \left[\frac{\hat{p}(x,y)}{p(x,y)}\right]^p \hat{p}(x,y)d\lambda(y) < \infty \tag{31}$$

Assuming this we can prove the result through repeated applications of the Holder's inequality. For brevity, we write p_i for $p(x^{(i)},y)$ and $\hat{p}_i$ for $\hat{p}(x^{(i)},y)$ where $i=1,2$.

$$\int \left[\frac{p_1}{p_2}\right]^p p_1 d\lambda = \int \left[\frac{\hat{p}_1}{p_2}\right]^p \left[\frac{p_1}{\hat{p}_1}\right]^{p+1} \hat{p}_1 d\lambda \leq \left(\int \left[\frac{\hat{p}_1}{p_2}\right]^{2p} \hat{p}_1 d\lambda\right)^{\frac{1}{2}} \left(\int \left[\frac{p_1}{\hat{p}_1}\right]^{2p+2} \hat{p}_1 d\lambda\right)^{\frac{1}{2}}$$

bounding the second factor by (31) gives

$$\leq C\left(\int \left[\frac{\hat{p}_1}{\hat{p}_2}\right]^{2p+1} \left[\frac{\hat{p}_2}{p_2}\right]^{2p} \hat{p}_2 d\lambda\right)^{\frac{1}{2}} \leq C\left(\int \left[\frac{\hat{p}_1}{\hat{p}_2}\right]^{4p+2} \hat{p}_2 d\lambda\right)^{\frac{1}{4}} \left(\int \left[\frac{\hat{p}_2}{p_2}\right]^{4p} \hat{p}_2 d\lambda\right)^{\frac{1}{4}} .$$

Since the second factor in the final expression is again bouned by (31) and the first factor by (30), we obtain the quoted result.

To complete the proof, we need to demonstrate (31). This will be accomplished through Lemma B.1. Since the difference between the equations for ℓ and $\hat{\ell}$ is the G term, we need to show that it is uniformly bounded for paths in A_1^+ and initial conditions in A_0. By Lemma (E.1), we know that we just need a uniform bound on $|u(t)|^2_{\mathbb{L}^2} = |\ell(t)|^2_{\mathbb{L}^2} + |h(t)|^2_{\mathbb{L}^2}$ over the interval. The initial high mode is uniformly bounded because we start in A_0. The low modes is bounded over the time interval because we also stay in A_1. Combining these two facts with Lemma D.1, shows that the high modes also stay bounded in $\mathbb{L}^2$ over the time interval. Combining all of this give the uniform bound on $|G|$ as was needed. □

5 Conclusion

This work is part of a close collaboration between the author, Weinan E, and Yakov Sinai. There support and encouragement have been invaluable. The author also acknowledges extremely useful discussions with Amir Dembo, George Papanicolaou, and Andrew Stuart. The author is also grateful for the hospitality of the Warwick Mathematics Institute. Their hospitality during October and November of 2000 provided the environment where this (more) readable version of the coupling construction could finally be written up. The NSF is also thanked for their support through grant DMS-9971087. Lastly, the author thanks a diligent reviewer who pointed out a number of inadequacies, both big and small, in an earlier version of this paper and whose comments pushed the author to rework some sections to improve their clarity.

A Energy Estimates

Lemma A.1. *There exist positive constants C_0 and C_1 so that*

$$\mathbb{E}\,|u(t)|_{\mathbb{L}^2}^2 \leq e^{-2\nu t}\mathbb{E}\,|u(0)|_{\mathbb{L}^2}^2 + \frac{\mathcal{E}_0^*}{2\nu}\left(1-e^{-2\nu t}\right)$$
$$\mathbb{E}\,|\Lambda u(t)|_{\mathbb{L}^2}^2 \leq e^{-2\nu t}\mathbb{E}\,|\Lambda u(0)|_{\mathbb{L}^2}^2 + \frac{\mathcal{E}_1^*}{2\nu}\left(1-e^{-2\nu t}\right)$$

For any $p > 1$,

$$\mathbb{E}\,|u(t)|_{\mathbb{L}^2}^{2p} \leq e^{-2p\nu t}\mathbb{E}\,|u(0)|_{\mathbb{L}^2}^{2p} + C_0\int_0^t e^{-2\nu(t-s)}\mathbb{E}\,|u(s)|_{\mathbb{L}^2}^{2(p-1)}\,ds$$
$$\mathbb{E}\,|\Lambda u(t)|_{\mathbb{L}^2}^{2p} \leq e^{-2p\nu t}\mathbb{E}\,|\Lambda u(0)|_{\mathbb{L}^2}^{2p} + C_1\int_0^t e^{-2p\nu(t-s)}\mathbb{E}\,|\Lambda u(s)|_{\mathbb{L}^2}^{2(p-1)}\,ds$$

In particular, this means that all moments of the energy and enstrophy are uniformly bounded by a constant times there initial value.

Proof of Lemma A.1. These are just the standard energy and Sobolev estimates in this setting. There is a little extra work to do in the stochastic setting. Proofs can be found in [VF88, Mat99, Mat98, EMS01]. Not all of these references consider the case when σ_k depends on on u. However, as we have assumed that each $\sigma_k(u)$ is uniformly bounded the same proofs apply. □

Lemma A.2. *Fix a $\epsilon_0 \in (0,1)$ and define*

$$E_t \stackrel{def}{=} |u(t)|_{\mathbb{L}^2}^2 + 2\nu\epsilon_0\int_0^t |\Lambda u(s)|_{\mathbb{L}^2}^2\,ds$$

There exist a $\gamma > 0$ so that whenever $|u(0)|_{\mathbb{L}^2}^2 \leq C_0$

$$\mathbb{P}\left\{\sup_{s\in[0,t]} E_s - \mathcal{E}_0^* s > C_0 + K\right\} \leq e^{-\gamma K}$$

for any K. Using this we see that for any positive K_0 and K_1 and $\delta \in (0,1]$, there is a positive C, independent of K_0, so

$$\mathbb{P}\left\{E_t \leq C_0 + K_0 + \mathcal{E}_0^* t + K_1 t^\delta \text{ for all } t \geq 0\right\} \geq 1 - Ce^{-\gamma K_0}$$

Furthermore the constant $C \to 0$ as $K_1 \to \infty$.

Proof of Lemma A.2.

$$\begin{aligned}|u(t)|_{\mathbb{L}^2}^2 + 2\nu\epsilon_0\int_0^t |\Lambda u(s)|_{\mathbb{L}^2}^2 = |u_0|_{\mathbb{L}^2}^2 + \int_0^t \mathcal{E}_0^*(u(s))ds - 2\nu(1-\epsilon_0)\int_0^t |\Lambda u(s)|_{\mathbb{L}^2}^2\,ds \\ + \int_0^t \langle u(s), \sigma(u(s))dW(s)\rangle_{\mathbb{L}^2} \\ \leq C_0 + \mathcal{E}_0^* t + N_t\end{aligned}$$

where $N_t = \frac{2\nu(1-\epsilon_0)}{\sigma_{max}^2}[M,M]_t + M_t$, $M_t = \int_0^t \langle u(s), \sigma(u(s))dW(s)\rangle_{\mathbb{L}^2}$ and $[M,M]_t$ is the quadratic variation of M_t. To prove the first estimate we need to control the probability that $\sup_s N_s > K$. In this setting, the exponential martingale inequality states that for positive γ and K,

$$\mathbb{P}\left\{\sup_{s\in[0,t]} M_s - \frac{\gamma}{2}[M,M]_s > K\right\} \le e^{-\gamma K} .$$

Hence the result follows with $\gamma = \frac{4\nu(1-\epsilon_0)}{\sigma_{max}^2}$

To prove the second result, define the events

$$B_n = \left\{\sup_{s\in[0,n]} N_s > K_0 + K_1(n-1)^\delta\right\}$$

and observe that

$$\mathbb{P}\left\{N_t \le K_0 + K_1 t^\delta\right\} \ge 1 - \mathbb{P}\left\{\bigcup_k B_k\right\} \ge 1 - \sum_k \mathbb{P}\left\{B_k\right\} .$$

Applying the first estimate to $\mathbb{P}\{B_k\}$, we see that the sum is finite. More exactly, there exist positive constants C,

$$\mathbb{P}\left\{N_t \le K_0 + K_1 n^\delta\right\} \ge 1 - \sum_n e^{-\gamma[K_0+K_1(n-1)^\delta]} \ge 1 - Ce^{-\gamma K_0} .$$

□

B Comparisons of Path Space Measures

Let $x(t)$ be the stochastic process in $\mathbb{R}^n$ on the time interval $[0,T]$ given by the Itô stochastic equation

$$dx(t) = f(t, x(\cdot))dt + g(t, x(\cdot))dW(t) . \tag{32}$$

$W(t)$ is a standard Brownian motion on $\mathbb{R}^m$. f and g are non-anticipating functionals from $\mathbb{R}_+ \times C([0,T];\mathbb{R}^n)$ into the $n\times m$-matrices and n-vectors respectively. (See [RY94, Pro90].) We assume that (32) is well defined pathwise up to time T with probability one and there exists a positive a so that $x^T gg^T x > a|x|^2$ for all $x \in \mathbb{R}^n$. Let

$$dy(t) = D(t, y(\cdot))dt + f(t, y(\cdot))dt + g(t, y(\cdot))dW(t)$$

where h is a non-anticipating n-vector valued function on $\mathbb{R}^n$. We will always start $x(t)$ and $y(t)$ from the same fixed initial condition. Let A be a subset of $C([0,T];\mathbb{R}^n)$ and let A_t be the restriction to $C([0,t];\mathbb{R}^n)$. For any $B \subset C([0,t];\mathbb{R}^n)$ define $P_t^x(B)$ and $P_t^y(B)$ by $\mathbb{P}\{x \in B\}$ and $\mathbb{P}\{y \in B\}$ respectively.

Lemma B.1. *In the above setting, assume there exists a* $D_* : \mathbb{R}_+ \to \mathbb{R}_+$ *such that*

$$\sup_{z(\cdot)\in A} |D(t, z(\cdot))| \le D_*(t)$$

for all $t \in [0, T]$*. Then for any* $p > 1$ *we have the following estimates:*

$$\mathbb{E}^y \left[\frac{dP_t^y}{dP_t^x}(y)\right]^p \mathbf{1}_{A_t}(y) \le \exp\left(\frac{a(p^2-p)}{2}\int_0^t |D_*(s)|^2 ds\right)$$

and

$$\mathbb{E}^y \left[\frac{dP_t^y}{dP_t^x}(y) - 1\right]^2 \mathbf{1}_{A_t}(y) \le \exp\left(a\int_0^t |D_*(s)|^2 ds\right) - 1$$

Proof of Lemma B.1. Define $\hat{D}(t, y(\cdot))$ equal to $D(t, y(\cdot))$ when $y \in A_t$ and zero otherwise. Let $\hat{y}(t)$ solve

$$d\hat{y}(t) = \hat{D}(t, \hat{y}(\cdot))dt + f(t, \hat{y}(\cdot))dt + g(t, \hat{y}(\cdot))dW(t) \ .$$

Observe that as long as $\hat{y}$ stays in A then $y = \hat{y}$. Hence, if $P^y(B) = \mathbb{P}\{\hat{y} \in B\}$ then

$$\mathbb{E}^y \left[\frac{dP_t^y}{dP_t^x}(y)\right]^p \mathbf{1}_{A_t}(y) = \mathbb{E}^{\hat{y}} \left[\frac{dP_t^{\hat{y}}}{dP_t^x}(\hat{y})\right]^p \mathbf{1}_{A_t}(\hat{y}) \le \mathbb{E}^{\hat{y}} \left[\frac{dP_t^{\hat{y}}}{dP_t^x}(\hat{y})\right]^p \ .$$

By Girsanov's theorem,

$$\begin{aligned}\left[\frac{dP_t^{\hat{y}}}{dP_t^x}(\hat{y})\right]^p &= \exp\left(-p\int_0^t [g^{-1}(s,\hat{y})D(s,\hat{y})]dW(s) - \frac{p}{2}\int_0^t |g^{-1}(s,\hat{y})D(s,\hat{y})|^2 ds\right) \\ &\le M_p(t)\exp\left(\frac{a(p^2-p)}{2}\int_0^t |D_*(s)|^2 ds\right)\end{aligned}$$

where

$$M_p(t) = \exp\left(-p\int_0^t [g^{-1}(s,\hat{y})D(s,\hat{y})]dW(s) - \frac{p^2}{2}\int_0^t |g^{-1}(s,\hat{y})D(s,\hat{y})|^2 ds\right) \ .$$

Because $M_p(t)$ has expected value one, we obtain the first of the quoted estimate. Equally we have

$$\mathbb{E}^{\hat{y}} \left[\frac{dP_t^{\hat{y}}}{dP_t^x}(\hat{y}) - 1\right]^2 = \mathbb{E}^{\hat{y}} \left[\frac{dP_t^{\hat{y}}}{dP_t^x}(\hat{y})\right]^2 + 1 - 2\mathbb{E}^{\hat{y}}\frac{dP_t^{\hat{y}}}{dP_t^x}(\hat{y}) = \mathbb{E}^{\hat{y}} \left[\frac{dP_t^{\hat{y}}}{dP_t^x}(\hat{y})\right]^2 - 1$$

which implies the second estimate. □

C Coupling Estimates

For any two probability measure μ_1 and μ_2 on a space X, we can always write them relative to a common measure ν so that $d\mu_i = \psi_i d\nu$. Then we define the measures $(\mu_1 \wedge \mu_2)(\,\cdot\,)$, $(\mu_1 - \mu_2)^+(\,\cdot\,)$, and $(\mu_2 - \mu_1)^+(\,\cdot\,)$ respectively by the densities $(\psi_1 \wedge \psi_2)d\nu$, $(\psi_1 - \psi_2)^+ d\nu$, $(\psi_2 - \psi_1)^+ d\nu$ where $a \wedge b = \min(a, b)$ and $(a)^+$ is a if a is positive and zero otherwise. Notice that $\mu_1 = (\mu_1 \wedge \mu_2) + (\mu_1 - \mu_2)^+$. Also observe that if $|\cdot|_{TV}$ is the total variation norm then $|\mu_1 - \mu_2|_{TV} = 1 - (\mu_1 \wedge \mu_2)(X) = (\mu_1 - \mu_2)^+(X) = (\mu_2 - \mu_1)^+(X)$.

Lemma C.1. *Let μ_1 and μ_2 be two measures on a space X with $\mu_i(X) \le 1$. Assume that μ_1 is equivalent to μ_2 and that there exists a constant $C' > 0$ and $p > 1$ so that*

$$\int_X \left[\frac{d\mu_1}{d\mu_2}(x)\right]^{p+1} d\mu_2(x) = \int_X \left[\frac{d\mu_1}{d\mu_2}(x)\right]^{p} d\mu_1(x) < C'$$

then

$$\int_X \left|1 \wedge \frac{d\mu_1}{d\mu_2}(x)\right| d\mu_2(x) \ge \left[1 - \frac{1}{p}\right] \left(\frac{\mu_1(X)^p}{pC'}\right)^{\frac{1}{p-1}} .$$

Notice that this lower bound is strictly positive if $\mu_1(X) > 0$ (or equivalently $\mu_2(X) > 0$).

Proof of Lemma C.1. Let $f(x)$ denote $\frac{d\mu_1}{d\mu_2}(x)$. Then

$$\begin{aligned}\int_X |1 \wedge f(x)|\, d\mu_2(x) =& \int_X \left(f - [f-1]^+\right) d\mu_2(x)\\ =&\mu_1(X) - \int_X \frac{[f-1]^+}{f} d\mu_1(x)\end{aligned}$$

Fixing any $M > 0$, we define $B_M = \{x : f(x) > M\}$ and B_M^c as its compliment. By the Chebyshev's inequality, $\mu_1(B_M) \le C' M^{-p}$. Continuing, we have

$$\begin{aligned}\int_X |1 \wedge f(x)|\, d\mu_2(x) =&\mu_1(X) - \int_{B_M} \frac{[f-1]^+}{f} d\mu_1(x) - \int_{B_M^c} \frac{[f-1]^+}{f} d\mu_1(x)\\ \ge&\mu_1(X) - \frac{M-1}{M}\mu_1(X) - \frac{C'}{M^p} = \frac{\mu_1(X)}{M} - \frac{C'}{M^p} .\end{aligned}$$

Optimizing over the choice of M gives the result. □

D Control of High Modes By Low Modes

Lemma D.1. *If $h(t)$ is the solution to* (19) *with some low mode forcing $\ell \in C\big([0,t], \mathbb{L}^2_\ell\big)$, then $\sup_{s\in[0,t]} |h(s)|_{\mathbb{L}^2}$ is bounded by a constant depending on $|h(0)|_{\mathbb{L}^2}$ and $\int_0^t |\ell|^4_{\mathbb{L}^2}\, ds$.*

Proof of Lemma D.1. This lemma follows from standard estimates on the nonlinearity followed by Gronwall's inequality. Its proof can be found in Lemma C.1 from [EMS01]. □

E B Estimates

Lemma E.1. *Let $\{e_k, k \in \mathbb{Z}^2\}$ be a basis for $\mathbb{L}^2$. Consider a splitting of $\mathbb{L}^2 = \mathbb{L}^2_\ell + \mathbb{L}^2_h$. Let N^+ be $\sup\{|k| : \exists\, e_k$ with $e_k \in \mathbb{L}^2_\ell\}$ and P_ℓ be the projector onto $\mathbb{L}^2_\ell$. If $u, v \in \mathbb{L}^2$ then*

$$|P_\ell B(u,v)| \leq C(N^+)^3 \, |u|_{\mathbb{L}^2} \, |v|_{\mathbb{L}^2}$$

Proof of Lemma E.1. This Lemma is taken from [EMS01]. It is a simple consequence of well known result in [CF88]. We recapitulate the proof as it is short.

In the periodic setting, P_ℓ, P_{div}, and $(-\Delta)^s$ all are simply Fourier multipliers and hence commute with one other. Recall that $B(u,v) = P_{div}(u \cdot \nabla)v$ and hence,

$$\begin{aligned}
|P_\ell B(u,v)| &= \sup_{\substack{w \in \mathbb{L}^2 \\ |w|=1}} |\langle P_\ell B(u,v), w\rangle_{\mathbb{L}^2}| = \sup_{\substack{w \in \mathbb{L}^2 \\ |w|=1}} |\langle B(u,v), P_\ell w\rangle_{\mathbb{L}^2}| \\
&= \sup_{\substack{w \in \mathbb{L}^2 \\ |w|=1}} |\langle B(u, P_\ell w), v\rangle_{\mathbb{L}^2}| \leq C \, |u|_{\mathbb{L}^2} \, |v|_{\mathbb{L}^2} \sup_{\substack{w \in \mathbb{L}^2 \\ |w|=1}} \left|\Lambda^3 P_\ell w\right|_{\mathbb{L}^2} \\
&\leq C(N^+)^3 \, |u|_{\mathbb{L}^2} \, |v|_{\mathbb{L}^2} \sup_{\substack{w \in \mathbb{L}^2 \\ |w|=1}} |w|_{\mathbb{L}^2} \leq C(N^+)^3 \, |u|_{\mathbb{L}^2} \, |v|_{\mathbb{L}^2}
\end{aligned}$$

□

References

[BKL01] J. Bricmont, A. Kupiainen, and R. Lefevere. Ergodicity of the 2D Navier-Stokes equations with random forcing. *Comm. Math. Phys.*, 224(1):65–81, 2001. Dedicated to Joel L. Lebowitz.

[CF88] Peter Constantin and Ciprian Foiaş. *Navier-Stokes Equations.* University of Chicago Press, Chicago, 1988.

[DPZ96] Giuseppe Da Prato and Jerzy Zabczyk. *Ergodicity for Infinite Dimensional Systems.* Cambridge, 1996.

[Dud76] R. M. Dudley. *Probabilities and metrics.* Matematisk Institut, Aarhus Universitet, Aarhus, 1976. Convergence of laws on metric spaces, with a view to statistical testing, Lecture Notes Series, No. 45.

[Dur96] Richard Durrett. *Probability: theory and examples.* Duxbury Press, Belmont, CA, second edition, 1996.

[EH] J. P. Eckmann and M. Hairer. Uniqueness of the invariant measure for a stochastic pde driven by degenerate noise. preprint.

[EM01] Weinan E and Jonathan C. Mattingly. Ergodicity for the Navier-Stokes equation with degenerate random forcing: finite-dimensional approximation. *Comm. Pure Appl. Math.*, 54(11):1386–1402, 2001.

[EMS01] Weinan E, J. C. Mattingly, and Ya G. Sinai. Gibbsian dynamics and ergodicity for the stochastic forced navier-stokes equation. *Comm. Math. Phys.*, 224(1), 2001.

[Fer97] Benedetta Ferrario. Ergodic results for stochastic Navier-Stokes equation. *Stochastics and Stochastics Reports*, 60(3–4):271–288, 1997.

[Fla94] Franco Flandoli. Dissipativity and invariant measures for stochastic Navier-Stokes equations. *NoDEA*, 1:403–426, 1994.

[FM95] Franco Flandoli and B. Maslowski. Ergodicity of the 2-D Navier-Stokes equation under random perturbations. *Comm. in Math. Phys.*, 171:119–141, 1995.

[Har56] T. E. Harris. The existence of stationary measures for certain Markov processes. In *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. II*, pages 113–124, Berkeley and Los Angeles, 1956. University of California Press.

[Has80] R. Z. Has′minskiĭ. *Stochastic stability of differential equations*. Sijthoff & Noordhoff, Alphen aan den Rijn, 1980. Translated from the Russian by D. Louvish.

[KS] Sergi Kuksin and Armen Shirikyan. Private comunication. December 2000.

[KS00] Sergei Kuksin and Armen Shirikyan. Stochastic dissipative PDEs and Gibbs measures. *Comm. Math. Phys.*, 213(2):291–330, 2000.

[Mat98] Jonathan C. Mattingly. *The Stochastically forced Navier-Stokes equations: energy estimates and phase space contraction*. PhD thesis, Princeton University, 1998.

[Mat99] Jonathan C. Mattingly. Ergodicity of 2D Navier-Stokes equations with random forcing and large viscosity. *Comm. Math. Phys.*, 206(2):273–288, 1999.

[MSH01] J. C. Mattingly, A.M. Stuart, and D. J. Higham. Ergodicity for SDEs and approximations: Locally lipschitz vector fields and degenerate noise. *Submited*, 2001.

[MT93] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, 1993.

[Pro90] Philip Protter. *Stochastic Integration and Differential Equations: a new approach*. Springer-Verlag, 1990.

[RY94] Daniel Revuz and Marc Yor. *Continuous martingales and Brownian motion*, volume 293 of *Grundlehren der Mathematischen Wissenschaften*. Springer-Verlag, Berlin, second edition, 1994.

[VF88] M.J. Vishik and A.V. Fursikov. *Mathematical Problems of Statistical Hydrodynamics*. Kluwer Academic Publishers, 1988. Updated version of Russian original of same name.