EXPANDING ON AVERAGE DIFFEOMORPHISMS OF SURFACES: EXPONENTIAL MIXING

JONATHAN DEWITT AND DMITRY DOLGOPYAT

ABSTRACT. We show that the Bernoulli random dynamical system associated to a expanding on average tuple of volume preserving diffeomorphisms of a closed surface is exponentially mixing.

CONTENTS

Date: October 12, 2024.

1. Introduction

1.1. **The main result.** In this paper, we prove exponential equidistribution and mixing results for expanding on average random dynamical systems. Suppose that M is a closed Riemannian surface with a smooth area, and $(f_1, \dots, f_m)$ is a tuple of diffeomorphisms in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. We then define a random dynamical system, where at each time step we choose uniformly at random an index $i \in \{1, \dots, m\}$ and apply f_i to M. We call this the (uniform Bernoulli) random dynamical system on M associated to the tuple $(f_1, \dots, f_m)$. A realization of the randomness is then given by a word from $\Sigma = \{1, \dots, m\}^{\mathbb{N}}$. As usual, we equip Σ with the distance $d(\omega', \omega'') = 2^{-k}$ where $k = \max\{N : \omega'_n = \omega''_n \text{ for } n < N\}$. We let $\sigma \colon \Sigma \to \Sigma$ denote the left shift and let μ the uniform Bernoulli product measure on Σ.

For such random dynamical systems, mixing does not hold for all tuples $(f_1, \dots, f_m)$. We will introduce an additional hypothesis. We say that a tuple $(f_1, \dots, f_m)$ is *expanding on average* if there exists $\lambda > 0$ and $n_0 \in \mathbb{N}$ such that for all $v \in T^1M$, the unit tangent bundle of M,

$$\frac{1}{n_0}\mathbb{E}\left[\ln \|Df_\omega^{n_0}v\|\right] \geq \lambda > 0. \tag{1.1}$$

Note that (1.1) is a C^1-open condition on the tuple $(f_1, \dots, f_m)$, so in principle it could be checked on a computer (cf. [Chu20]).

The main result of our paper is that the systems satisfying (1.1) enjoy exponential mixing.

Theorem 1.1. *(Quenched Exponential Mixing) Suppose that M is a closed surface and that $(f_1, \dots, f_m)$ is an expanding on average tuple of diffeomorphisms in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. Let $\beta \in (0,1)$ be a Hölder regularity. There exists $\eta > 0$ such that for a.e. $\omega \in \Sigma$, there exists C_ω such that for any $\phi, \psi \in C^\beta(M)$,*

$$\left| \int \phi\psi \circ f_\omega^n \, d\,\mathrm{vol} - \int \phi \, d\,\mathrm{vol} \int \psi \, d\,\mathrm{vol} \right| \leq C_\omega e^{-\eta n} \|\phi\|_{C^\beta} \|\psi\|_{C^\beta} \tag{1.2}$$

where $f^i_{\sigma^j(\omega)} = f_{\omega_{j+i}} \cdots f_{\omega_{j+1}}$. Further, there exists $D_1 > 0$ such that

$$\mu(\omega : C_\omega \geq C) \leq D_1 C^{-1}. \tag{1.3}$$

In fact, the tail bound (1.3) implies a related result, annealed exponential mixing for the associated skew product. We give the proof of the following in §11.4.

Corollary 1.2. *(Annealed Exponential Mixing) Let M be a closed surface, let $(f_1, \dots, f_m)$ be an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$, and $\beta \in (0,1)$ be a Hölder regularity. Let $F\colon \Sigma \times M \to \Sigma \times M$ be the skew product defined by*

$$F(\omega, x) = (\sigma(\omega), f_{\omega_0}(x)).$$

Then F is exponentially mixing, that is, there exist $\bar{\eta} > 0$, D such that for any $\Phi, \Psi \in C^\beta(\Sigma \times M)$,

$$\left| \iint \Phi(\Psi \circ F^n) \, d\mu \, d\,\mathrm{vol} - \iint \Phi \, d\mu \, d\,\mathrm{vol} \iint \Psi \, d\mu \, d\,\mathrm{vol} \right| \leq De^{-\bar{\eta} n} \|\Phi\|_{C^\beta} \|\Psi\|_{C^\beta}.$$

Before we proceed to discussing the relationship of this work with the existing literature, we will look at some examples of systems satisfying (1.1).

Remark 1.3. *Although we have written this paper for a finite tuple $(f_1, \dots, f_m)$ of diffeomorphisms to emphasize the discreteness of the noise, one can consider random dynamics generated by any probability measure μ on $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. Similar arguments to the ones we present here imply the analogous conclusions hold for random dynamics generated by a measure μ with compact support on $\mathrm{Diff}^2_{\mathrm{vol}}(M)$, where M is a closed surface.*

1.2. **Examples.** There are a number of sources of tuples $(f_1, \dots, f_m)$ that are expanding on average. The random dynamics arising from such tuples may exhibit uniform or non-uniform hyperbolicity. One of the simplest and archetypal examples is the following.

Example 1.4. Suppose that $(A_1, \dots, A_m)$ is a tuple of matrices in $\mathrm{SL}(2, \mathbb{Z})$ satisfying the hypotheses of Furstenberg's theorem, namely the tuple is strongly irreducible and contracting. Then the Bernoulli random product of these matrices has a positive top Lyapunov exponent. It follows from the proof of Furstenberg's theorem, see, e.g. [BL85, Thm. III.4.3], that there exists N and $\lambda > 0$ such that for all unit vectors $v \in \mathbb{R}^2$,

$$N^{-1}\mathbb{E}\left[\ln \|A_\omega^N v\|\right] \geq \lambda > 0.$$

Each $A_i \in \mathrm{SL}(2,\mathbb{Z})$ acts on $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$, and the associated random dynamics on $\mathbb{T}^2$ is uniformly expanding on average. Because this is an open condition, we see that any volume preserving perturbation of the A_i is also uniformly expanding. Thus, our theorem applies to a class of non-linear systems that do not exhibit any uniform hyperbolicity.

In addition, the expanding on average property generalizes to many other random walks on homogeneous spaces, see for example [EL, Def. 1.4], which uses this property to study stiffness of stationary measures of random walks on homogeneous spaces.

Expanding on average systems also arise as perturbations of isometric systems.

Example 1.5. Perhaps the first example where this condition was considered for nonlinear diffeomorphisms was the paper of Dolgopyat and Krikorian [DK07]. Suppose that $(R_1,\dots,R_m)$ is a tuple of isometries of S^2 that generates a dense subgroup of $\mathrm{SO}(3)$. Then [DK07] shows that there exists k_0 such that if $(f_1,\dots,f_m)$ is a sufficiently C^{k_0} small volume preserving perturbation of $(R_1,\dots,R_m)$, and the tuple $(f_1,\dots,f_m)$ has a stationary measure with non-zero Lyapunov exponents, then $(f_1,\dots,f_m)$ is expanding on average. See also DeWitt [DeW24].

Other work has explored how ubiquitous expanding on average systems are, in some cases studying whether expanding on average systems can be realized by perturbing a known system of interest.

Example 1.6. Chung [Chu20] gives a proof that certain random perturbations of the standard map are expanding on average (see also [BXY17, BXY18] which studies the size of Lyapunov exponents for perturbations of the standard map with a large coupling constant). [Chu20] also presents convincing numerical simulations showing that certain actions on character varieties are expanding on average as well.

There are also some results that construct expanding on average systems densely in a weak* sense.

Example 1.7. The paper [Pot22] says that for every open set $\mathcal{U} \subseteq \mathrm{Diff}^{\infty}_{\mathrm{vol}}(M)$, where M is a surface, there exists a finitely supported measure on $\mathcal{U}$ that is expanding on average. This result was generalized to higher dimensions in [ES23].

1.3. **Relationship with other works.** Exponential mixing plays the central role in the study of statistical properties of dynamical systems. In particular, multiple exponential mixing implies several probabilistic results including the Central Limit Theorem [Che06, BG20], Poisson Limit Theorem [DFL22],and the dynamical Borel Cantelli Lemma [Gal10] among others. Further, exponential mixing was recently shown to imply Bernoullicity [DKRH24].

For deterministic systems, however, robust exponential mixing has been only established for a limited class of systems: uniformly hyperbolic systems in both smooth and piecewise smooth settings [CM06, Via99, You98], or for partially hyperbolic systems where all Lyapunov exponents in the central direction have the same sign [dCJ02, CV13, Dol00]. Here we say that a certain property holds robustly if it holds for a given system as well as for its small perturbations. In contrast, if additional symmetries are present then there are many other cases where exponential mixing is known, see [GS14, KM96, Liv04, TZ23]. There are also checkable conditions for exponential mixing in the nonuniformly hyperbolic setting, see [You98, You99]. However, except for the aforementioned examples, these conditions hold for individual systems rather than open sets. On the other hand KAM theory tells us that away from (partially) hyperbolic systems one has open sets of non-ergodic systems, so one cannot expect chaotic behavior to be generic.

The situation is different for random systems. In fact, if the supply of random maps is rich enough then one show that exponential mixing and other statistical properties hold generically. Such results are known for stochastic flows of diffeomorphisms [DKK04] as well as for random deterministic shear flows [BCZG23]. It is therefore natural to ask how large should the set of random diffeomorphisms must be so that the corresponding random dynamical system exhibits random behavior. The following conjecture is formulated in [DK07].

Conjecture 1.8. For each closed manifold M with volume and regularity $k \geq 1$, there exists m, such that the space of tuples $(f_1, \ldots, f_m)$ that are stably ergodic is open and dense in $\left(\mathrm{Diff}^k_{\mathrm{vol}}(M)\right)^m$.

The point of this conjecture is that only a tiny bit of randomness, perhaps even the minimum amount, should be sufficient to ensure robust ergodic and statistical properties for dynamical system. Consequently, the situation where the driving measure has uniformly small, finite support on $\mathrm{Diff}^2_{\mathrm{vol}}(M)$ is the most interesting, and hardest case to consider this question. The obvious approach to this conjecture is to first to show that an open and dense set of tuples is expanding on average.

Other papers have significantly extended the properties of expanding on average systems. One of the first is [BRH17], which shows a strong stiffness property of these systems: any stationary measure for the Markov process that is not finitely supported is volume [BRH17, Thm. 3.4]. Thus, in some sense, volume is the only measure whose statistical properties are interesting to study. The only statistical property beyond ergodicity studied before for expanding on average systems is large deviations for ergodic sums established in [Liu16, Thm. 4.1.1]. Our paper provides an additional contribution to this topic by showing that expanding on average systems enjoy exponential mixing. In fact, Conjecture 1.8 provides an additional motivation for this work, because it shows that should the conjecture be true, then exponential mixing is a generic property for random dynamical systems.

Some work has been done towards showing that uniform expansion is a generic property. In particular, [OP22] shows that one may obtain positive integrated Lyapunov exponent for conservative random systems on surfaces. This work differs from the papers [Pot22] and [ES23] as [OP22] does not require an arbitrarily large number of diffeomorphisms to obtain its result.

Returning to deterministic systems, it is natural to ask for conditions for strong statistical properties to hold in a robust way. Optimal conditions are not yet well understood. While there are strong indications that at least a dominated splitting is necessary [Pal00], the best available results pertain to partially hyperbolic systems. A well known conjecture of Pugh and Shub [Shu06] states that stably ergodic systems contain an open and dense subset of partially hyperbolic systems. Currently the best results on this problem are due to [BW10] which can be consulted for a detailed discussion on this subject. In fact, the methods of Pugh and Shub also give the K-property [BW10]. Going beyond the K-property remains an outstanding challenge even in the partially hyperbolic setting. In view of the strong consequences of exponential mixing it is natural to conjecture the following.

Conjecture 1.9. Exponential mixing holds for an open and dense set of volume preserving partially hyperbolic systems.

Currently there are two possible ways to attack this conjecture. The first one is based on the theory of weighted Banach spaces, [AGT06, CL22, GL06, Tsu01, TZ23]. To describe the second approach recall that the papers [Via08, AV10] show that partially hyperbolic systems often have non-zero exponents. It is therefore natural to see if one could try to extend the methods used in proving exponential mixing in non-uniformly hyperbolic systems to handle partially hyperbolic setting. As mentioned above, this approach was successful in handling the case there the central exponents have the same sign. In the present paper we consider a skew product with a shift in the base and where the Lyapunov exponents in the central direction have different signs. We hope that a similar approach could be useful for studying more general skew products, and hopefully could provide a blueprint for studying mixing in partially hyperbolic systems.

In summary, the present work is the first step in extending mixing to a large class of smooth systems both random and deterministic, and we hope that various extensions will be addressed in future works.

Acknowledgments: The first author was supported by the National Science Foundation under Award No. DMS-2202967. The second author was supported by the National Science Foundation under award No. DMS-2246983. The authors are grateful to Matheus Manzatto de Castro for comments on an earlier version of the manuscript.

2. Setting and basic definitions

2.1. **Random dynamics and skew products.** In this section, we will state some basic definitions that will be used throughout the paper. Although we introduce many of these definitions and notations here, we will recall and reintroduce them when they are used; this section is just an overview.

We begin by recalling the main definition of our setup.

Definition 2.1. We say that a tuple $(f_1, \dots, f_m) \in \mathrm{Diff}^1(M)$ is *expanding on average* if there exists some $n_0 \in \mathbb{N}$ and $\lambda_0 > 0$ such that for all $v \in T^1M$,

$$\mathbb{E}\left[n_0^{-1} \ln \|Df_\omega^{n_0} v\|\right] \geq \lambda_0 > 0. \tag{2.1}$$

Throughout the paper, $(f_1, \dots, f_m)$ typically denotes an uniformly expanding on average tuple of volume preserving diffeomorphisms of a closed surface M. However, in some cases, we merely are referring to a tuple and do not make use of any further assumptions.

We write (Σ, σ) for the one sided shift on m symbols, i.e. $\Sigma = \{1, \dots, m\}^{\mathbb{N}}$ with σ being the left shift. We endow this space with the measure μ, which is the uniform Bernoulli measure on Σ. Write $\hat{\Sigma}$ and $\hat{\mu}$ for the two-sided shift and the invariant Bernoulli measure over μ.

We may view the random dynamics in two ways. First, as a Markov process on M. The second way, as mentioned in the statement of Corollary 1.2, is as the skew product $F\colon \Sigma \times M \to \Sigma \times M$. This skew product preserves the product measures $\mu \otimes \mathrm{vol}$. When we say that the tuple $(f_1, \dots, f_m)$ is *ergodic*, we mean that the skew product F is ergodic for the measure $\mu \otimes \mathrm{vol}$. This is equivalent to the absence of almost surely invariant Borel subsets of M of intermediate measure. See [Kif86] for more discussion of the relationship between the skew product and the random dynamics on M.

For a word $\omega \in \Sigma$, we write $f_\omega^n\colon M \to M$ for the composition $f_{\omega_n} \cdots f_{\omega_1}$. We use the same notation for finite words ω. For a sequence of linear maps $(A_i)_{1 \leq i \leq n}$, we write $A^i = A_i \cdots A_1$. We do not always start this product with the first matrix, so we also have the notation

$$A_i^k = A_{i+k} \cdots A_{i+1}.$$

Note that this is compatible with the notation $f_{\sigma^j(\omega)}^i = f_{\omega_{j+i}} \cdots f_{\omega_{j+1}}$ from above.

2.2. **Stable subspaces.** For a sequence of linear maps, we will frequently use the singular value decomposition when it is defined. If we have a sequence of matrices $A_1, A_2, \dots$ then, when it is defined, we write E_n^s for the most contracted singular direction of A^n. We usually apply this to the sequence of linear maps $D_x f_\omega^n$. We write $E_i^s(\omega, x)$ for most contracted singular direction of $D_x f_\omega^i$, and we write $E_i^u(\omega, x)$ for the most expanded singular direction of $D_x f_\omega^i$, should these directions be well defined. Often we will suppress the x and ω and just write E_i^s, other times we will write $E_\omega^s(x)$.

Throughout the paper we will consider sets Λ_n^ω which are the sets of points $x \in M$ that are (C, λ, ϵ)-tempered for the word ω up until time n, where temperedness is defined in §4.1. These points are essentially the finite time analogue of a Pesin block, c.f. [BP07].

2.3. **Stable manifolds.** The most important dynamical objects we will consider are the stable manifolds and fake stable manifolds. Given a point $x \in M$, we define its stable manifold to be the set of points

$$W^s(\omega, x) = \{y \in M : d(f_\omega^n(x), f_\omega^n(x)) \text{ exponentially fast}\}.$$

Note that the stable manifold depends on ω. We denote a segment of length 2δ centered at x in $W^s(\omega, x)$ by $W_\delta^s(\omega, x)$. The properties of these "true" stable manifolds are discussed in Section 5. For general information about stable manifolds in random dynamical systems, see [LQ95].

As alluded to above, we will not only work with the stable manifolds, but also with finite time versions of stable manifolds. We will denote by $W_{n,\delta_0}^s(\omega, x)$ the time n fake stable manifold of x for the word ω restricted to segment of radius δ_0 centered at x. The point of the fake stable manifolds is that up to time n, they have similar contraction properties to an actual stable manifold. In the limit, they converge to the true stable manifold. Their definition is somewhat technical, but a detailed treatment of the fake stable manifolds is given in Appendix B which essentially concerns itself with a quantified, finite time version of Pesin theory.

An important application of stable manifolds, fake or otherwise, is their holonomy. Suppose that we have two curves γ_1, γ_2 and a locally defined lamination $\mathcal{W}$ such that each leaf of $\mathcal{W}$ intersects γ_1 and γ_2 at a unique point. Let I_1 and I_2 be the points of intersection of $\mathcal{W}$ with γ_1 and γ_2. Then $\mathcal{W}$ defines a *holonomy* map $H^{\mathcal{W}} \colon I_1 \to I_2$ by carrying the unique point of intersection with a particular plaque of the lamination to the corresponding point in the other curve.

An important property that such a holonomy may satisfy is *absolutely continuity* with respect to volume, which means that it carries Riemannian volume of γ_1 restricted to I_1 to a measure equivalent to the restriction to I_2 of Riemannian volume on γ_2. These properties will be discussed in more detail in Appendix B.

2.4. **Norms.** In this paper, we will use many estimates from calculus.

First we consider the norms of curves. An unparametrized curve in a manifold does not come equipped with any C^1 norm, as the C^1 norm of a curve is dependent on parametrization. Consequently, we will always view such a curve with its arclength parametrization. For $x \in \gamma$, we may consider the norm of the second derivative of γ at the origin when we view γ as a graph over its tangent in an exponential chart. We then define $\|\gamma\|_{C^2}$ as the supremum of this norm over all $x \in \gamma$. Note that this is essentially the same thing as the supremum of the extrinsic curvature of γ at x over all points $x \in \gamma$.

Throughout the proof, we will be interested in studying the log Hölder norms of some densities along curves. We will be slightly unconventional and write $\|\ln \rho\|_{C^\alpha}$ for the Hölder constant of $\ln \rho$, where ρ is a density. Note that this doesn't include an estimate on $\| \ln \rho\|_\infty$, as such a norm usually contains. This is because the magnitude of the density is infrequently the important things in our arguments.

When we work in coordinates, we will write $\|\phi\|_i$ as the supremum of all the ith partial derivatives of the function ϕ. For example, if $\phi\colon \mathbb{R}^2 \to \mathbb{R}$, then we define

$$\|\phi\|_2 = \sup_{x\in\mathbb{R}^2} \max\left\{\left|\frac{d^2\phi}{dxdy}\right|, \left|\frac{d^2\phi}{dx^2}\right|, \left|\frac{d^2\phi}{dy^2}\right|\right\}.$$

2.5. **Probability facts.** In the course of the paper we will some facts from probability, which we state here for the convenience of readers who are familiar with dynamics but not as much with probability. Sometimes we will write something like $\mathbb{P}_\omega(A)$ for the measure $\mu(A)$ when we are thinking probabilistically. Also, we will often write $\mathbb{E}\left[\ldots\right]$ when we are taking expectations with respect to μ, as μ is the measure driving the random dynamics.

The following concentration in equality is very useful for us.

Theorem 2.2. [Ste97, Thm. 1.3.1] *(Azuma-Hoeffding inequality) Suppose that $X_1, X_2, \ldots$ is a martingale difference sequence. Then*

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i\right| \ge \lambda\right) \le 2\exp\left(\frac{-\lambda^2}{2\sum_{i=1}^n \|X_i\|_{L^\infty}^2}\right). \tag{2.2}$$

3. Outline of the paper

3.1. **Quenched and annealed properties.** The main technical result of this paper is a type of "annealed" coupling theorem, Proposition 7.7. From this theorem we deduce after a small amount of additional work, quenched exponential equidistribution (Proposition 11.9) as well as quenched exponential mixing, which, in turn, implies the annealed exponential mixing (see Corollary 1.2).

Before proceeding, let us recall what is meant, in the probabilistic sense, by an *annealed* as opposed to a *quenched* limit theorem for a random dynamical system defined by Bernoulli random application of maps $(f_1, \ldots, f_m)$. In an *annealed* limit theorem, we average over the entire ensemble whereas in a *quenched* limit theorem one obtains a limit theorem for almost every realization of the random dynamics. For example, in the case of equidistribution consider $\phi\colon M \to \mathbb{R}$ a Hölder observable and ν a probability measure on M, such as a curve with density. Then annealed equidistribution says:

$$\frac{1}{m^n} \sum_{\omega^n \in \{1,\ldots,m\}^n} \int \phi \circ f_\omega^n \, d\nu \to \int \phi \, d\,\mathrm{vol},$$

whereas quenched equidistribution says that for almost every $\omega \in \Sigma^{\mathbb{N}}$ with respect to the Bernoulli measure μ on Σ,

$$\int \phi \circ f_\omega^n \, d\mu \to \int \phi \, d\,\mathrm{vol}.$$

Note that the annealed result follows from the mixing of the skew product studied in §6.2.

While the two notions are not always equivalent, our annealed coupling theorem comes with such fast rates that by the Fubini theorem, we can deduce quenched limit theorems. This reduction happens in Section 11.

3.2. **Description of the key step.** The main results of this paper follow from our annealed exponentially fast coupling proposition, Proposition 7.7, which says the following. Suppose we have two standard pairs $\hat{\gamma}_1$ and $\hat{\gamma}_2$. Each standard pair is a C^2 curve γ_i along with a density ρ_i defined along γ. Suppose that $\omega \in \Sigma$ is a random word. We say that two points $x \in \gamma_1$ and $y \in \gamma_2$ are "coupled" at time k if:

(1) $f^k_\omega(x) \in W^s_{loc}(\sigma^k(\omega), f^k_\omega(y))$,
(2) The stable manifold $W^s_{loc}(\sigma^k(\omega), f^k_\omega(y))$ contracts uniformly exponentially quickly, so that $f^k_\omega(x)$ and $f^k_\omega(y)$ attract uniformly exponentially fast, independent of x, y, ω.

In other words, after two points couple at time k they attract uniformly quickly. In fact, in our coupling procedure if x and y couple at time k then $f^k_\omega(x)$ and $f^k_\omega(y)$ both lie in a uniformly (C, λ, ϵ)-tempered stable manifold (see Definition 5.1). Proposition 7.7 constructs a coupling which occur exponentially quickly in the sense that the set of points where the coupling time is greater than k has exponentially small measure.

The first step towards constructing the coupling is to show that for two "nice" standard pairs $\hat{\gamma}_1$ and $\hat{\gamma}_2$ that are quite close, there exist uniform $\epsilon_0, \epsilon_1 > 0$ such that with ϵ_0 probability at least ϵ_1 proportion of the mass of $\hat{\gamma}_1$ couples at time 0. Namely, with ϵ_0 probability, the stable manifolds W^s_ω intersect $\hat{\gamma}_1$ and $\hat{\gamma}_2$ in sets of uniformly large measure, thus those points can be coupled. This fact implies that a positive proportion of the mass on $\hat{\gamma}_1$ can be coupled at the first attempt.

The complement of the pairs that couple is the disjoint union of a potentially large number of very small curves. For these "leftover" curves we will wait a potentially long time for them to grow and smoothen and then equidistribute at small scale so that we can try coupling them again. We refer to this growth and smoothening as "recovery" and the equdistribution as "precoupling." As a positive proportion of the remaining mass gets coupled during each attempt at coupling, we expect only an exponentially small amount of mass to remain uncoupled after n attempts.

The actual argument is much more complicated for a fairly simple reason: we cannot determine if two points x and y lie in the same stable manifold until we have seen the entire word ω. However, we do not want to look into the future at the entire word ω since then we would loose the Markov character of dynamics and would not be able to use many estimates that rely on the Markov property. Consequently, we define a "stopping" time for each pair (x, ω) which tells us when to "give up" on trying to couple during the current attempt and switch to recovery. For the moment, we regard the coupling argument as having three main steps:

(1) (Local Coupling) Attempt to couple two uniformly smooth nearby curves $\hat{\gamma}_1$ and $\hat{\gamma}_2$.
(2) (Recovery) Show that pieces of curve that fail to couple recover quickly so that their image become long and smooth.
(3) (Precoupling) There is a time N_0 such that given two long smooth curves we can divide them into subcurves such that for most of the subcurves their images N_0 units of time later are close to each other, so we can then try to locally couple them again.

We now describe the outline of the rest of the paper and how its different sections relate to the three main steps described above.

The first goal of the paper is show that for any point $x \in M$ that for most words $\omega \in \Sigma$ the stable manifolds $W^s(\omega, x)$ have good properties including good distribution of their tangent vector, controlled C^2 norm, and that they contract quickly. To do this, we will need to obtain good estimates on Df^n_ω. We show that for typical words ω, Df^n_ω has a putative stable direction that has all of the properties that the stable direction of a Pesin regular point would have.

We formalize these properties with our notion of (C, λ, ϵ)-temperedness, which is described in detail in §4.1. We remark, however, that this notion is weaker than the usual notion of ϵ-temperedness used in Pesin theory. We show that there exist $\lambda, \epsilon > 0$ such that for almost every word ω that the trajectory will exhibit $(C(\omega), \lambda, \epsilon)$-temperedness for some $C(\omega) > 0$. Further, we obtain estimates for the tail of $C(\omega)$. We then also study the distribution of $E^s_\omega(x)$, the stable direction for the word ω at the point x and obtain estimates on the regularity of this measure, which show that the distribution of $E^s(\omega)$ and hence the stable manifolds is not concentrated in any particular direction, see Proposition 4.11. This discussion occupies Section 4. Through the application of Azuma's inequality, we are able to show that a typical trajectory exhibits temperedness.

In Section 6, we study the mixing properties of the skew product map F. The proofs rely on the properties of stable manifolds that are recalled in Section 5. Mixing plays a crucial role in the Finite Time Mixing Proposition given in Section 9. This plays an important role at the precoupling stage.

Section 7 contains the precise statement of the main coupling Proposition 7.7. We then divide the proof into three main parts: the Local Coupling Lemma 7.10, the Coupled Recovery Lemma 7.9, and the Finite Time Mixing Proposition 7.11 which corresponds to steps (1)–(3) in the outline above. Lemma 7.9 is proven in Section 8, Proposition 7.11 is proven in Section 9, and Lemma 7.10 is proven in Section 10.

Finally, in Section 11 we derive our main results from the main coupling proposition: we derive Theorem 1.1 and Corollary 1.2 from Proposition 7.7.

The paper contains two appendices. Appendix A describes how the smoothness of a curve which is transversal to the stable direction improves under the dynamics, while Appendix B discusses fake stable manifolds and their holonomy. In particular, we show that these objects converge exponentially fast to true stable manifolds and holonomies respectively. While the estimates in the appendices are similar to several results in Pesin theory, we provide the proofs in our paper since we could not find exact references in the existing literature. This is partially due to the fact that we put a greater emphasis to the finite time estimates because we want to preserve the Markov property of the dynamics and hence cannot base our coupling algorithm on the knowledge of the future behavior of orbits.

3.3. **Mixing in hyperbolic dynamics.** We now compare our work with strategies used in other works. Historically the first mixing results for hyperbolic systems relied on symbolic dynamics, see [Bow75, Rue78, Sin72, PP90]. Currently the most flexible realization of this approach is via symbolic dynamics given by Young towers ([You98]). Later, several methods working directly with the hyperbolic systems were developed. In particular, we would like to mention weighted Banach spaces developed in [GL06] (see [Bal00] for a review) as well as the coupling approach developed in [You99]. We note that most hyperbolic systems could be analyzed by each of these methods but a different amount of work is required in different cases. For example, a recent paper [DL23] constructs weighted Banach spaces suitable for the billiard dynamics. However, these spaces are necessarily complicated reflecting the complexity of billiards systems.

In our work, we use the coupling approach. This method was originally used in [You99] to handle symbolic systems, while the modifications which allow working directly on the phase space are due to [Dol00, CM06]. The two papers mentioned above implemented the coupling methods for systems with dominated splitting. In our case, we have to deal with the general non-uniformly hyperbolic situation and this significantly expands the potential applications of the coupling method.

An attractive feature of our result is that we make only one assumption (1.1) which is, in fact, open. Our result is an example of a successful implementation of the line of research asking which dynamical properties follow just from existence of a hyperbolic set with controlled geometry. This direction is exemplified by a conjecture of Viana [Via98], which asks if the existence of positive measure hyperbolic set implies existence of a physical measure. While several important recent results obtained progress on this question (see [BO21, Bur24, BCS23, CLP22] as well as [BCS22] which deals with a measure of maximal entropy), much less is known about qualitative properties. In the present (and a follow up) paper we are able to get a full package of statistical properties starting from a simple assumption (1.1).

Below we list key ingredients of our approach since similar ideas could be useful in studying other hyperbolic systems.

(1) Using martingale large deviation bounds, we demonstrate an abundance of times where the orbit of a given vector is backward tempered.
(2) Using two dimensionality and volume preservation, we promote exponential growth of the norm to existence of a hyperbolic splitting.
(3) Using Pesin theory we show that hyperbolic set cannot have gaps of too small a size since these gaps would be filled with orbits of slightly weaker hyperbolicity.
(4) We use fake stable manifolds and quantitative estimates on their convergence to construct a finite time "fake" coupling.
(5) Using a Mañe type argument we show that a fake coupling converges quickly to a real coupling for most trajectories.

Finally, we would like to mention that recently a different approach to quenched mixing based on random Young towers has been developed, see [ABR22, ABRV23]. So far, the authors have proved the existence of random towers for relatively simple systems where hyperbolicity is uniform at least in one direction. It might be possible to obtain exponential mixing in our case by verifying the conditions of [ABRV23], however, this would not simplify our analysis. Indeed the main ingredients of the Young towers is the following: the existence of a positive measure horseshoe, an exponential tail on the return time, and a finite time mixing estimate. The last ingredient is already established in our paper. To construct a large horseshoe would require estimates similar to our local coupling lemma of Section 10, while having an exponential tail on return times would be similar to our recovery lemma of Section 8. In addition there several technical properties of Young tower whose verification would require additional space and effort. For this reason we prefer to give a direct proof of exponential mixing in our setting rather than deducing our result by a lengthy verification of the conditions of the deep recent work of [ABRV23].

4. Estimates on the growth of vectors and temperedness

In this section, we study infinitesimal properties of uniformly expanding random dynamical systems. The main results of this section are a proof that the sequence of linear maps $Df_{\omega_0}, Df_{\omega_1}, \dots, Df_{\omega_n}$ applied along the trajectory of a point x typically has a splitting with most of the same properties as a point in a Pesin block has. Moreover, we give quantitative estimates on the angle between the vectors in the splitting, as well as the probability that the splitting experiences a renewal.

4.1. **Tempered vectors and sequences of linear maps.** In this subsection we discuss some notions of tempering for sequences of linear maps. We remark that typical notions of tempering used in Pesin theory involve both lower and upper bounds, i.e. they involve a statement like $e^{\lambda-\epsilon} \le \|A|_{E^u}\| \le e^{\lambda+\epsilon}$. We will only take one of these two bounds to avoid having to do more estimates than necessary. Further, the version of tempering used in Pesin theory is often adapted so that the value of λ is a particular Lyapunov exponent for a particular measure. In such a context, a tempered splitting will have expansion at rate $e^{\lambda-\epsilon}$ rather than at rate e^{λ}, as we have below. Compare for example, with the definition of (λ, μ, ϵ)-tempered in [BP07, Def. 1.2.]. In the language of this section, points that are (λ, μ, ϵ)-tempered in the sense of [BP07], have a splitting that is $(C, \lambda - \epsilon, \epsilon)$-tempered in our sense.

Before we get to our ultimate notion of a tempered splitting, Definition 4.2, we first record several estimates and introduce intermediate notions.

Definition 4.1. Consider a finite or infinite sequence of linear maps $(A_n)_{n\in I}$ between a sequence of normed 2-dimensional vector spaces V_i, where I is either $\mathbb{N}$ or a set of the form $\{1, \dots, n\}$, and $A_i \colon V_i \to V_{i+1}$.

(1) We say that (A_n) has (C, λ, ϵ)-*subtempered* norms when

$$\|A^{i+j}\| \geq e^{C} e^{\lambda i} e^{-\epsilon j} \|A^j\|,$$

for all $i \geq 1$, $j \geq 0$, with $i + j \in I$.

(2) We say that a vector v is (C, λ, ϵ)-*subtempered* for the sequence of linear transformations A_i if

$$\|A_k^m v^k\| \geq e^{C} e^{\lambda m} e^{-\epsilon k}, \tag{4.1}$$

where $A_k^m = A_{k+m} \cdots A_{k+1}$ and $v^k = A^k v / \|A^k v\|$, for all $k, m \in \mathbb{N}$ with $k + m \in I$.

(3) We say that the vector v is (C, λ, ϵ)-*supertempered* if

$$\|A_k^m v^k\| \leq e^{C} e^{\lambda m} e^{\epsilon k}, \tag{4.2}$$

for all m, k and v^k as above.

(4) Similarly, we may speak of a vector $v \in T_x M$ being sub or super tempered for a sequence of diffeomorphisms $(f_n)_{n\in I}$ if it sub or super tempered for the sequence of differentials $D_x f_1, D_{f_1(x)} f_2, \dots$, etc.

Finally, we say that a sequence of maps has an (C, λ, ϵ)-tempered splitting if there exists a pair of directions e^u and e^s such that the action of the maps is $(-C, \lambda, \epsilon)$-subtempered on e^u and $(C, -\lambda, \epsilon)$-supertempered on e^s. In addition, we impose a lower bound on the angle between these two directions. Note that we do not require the angle itself to be tempered in the sense that it locally decays slowly: we just require that it stay bounded below by a slowly decaying function.

Definition 4.2. We say that a finite or infinite sequence $A_1, \dots, A_n$ of linear maps $A_i \colon V_i \to V_{i+1}$ of 2-dimensional inner product spaces has a (C, λ, ϵ)-*tempered* splitting if there exists a pair of unit vectors $e^s, e^u \in V_1$ such that

$$\|A_k^m (A^k e^u)\| / \|A^k e^u\| \geq e^{-C} e^{\lambda m} e^{-\epsilon k}, \tag{4.3}$$

$$\|A_k^m (A^k e^u)\| / \|A^k e^s\| \leq e^{C} e^{-\lambda m} e^{+\epsilon k}, \tag{4.4}$$

$$\angle(A^k e^s, A^k e^u) \geq e^{-C} e^{-\epsilon k}. \tag{4.5}$$

Similarly, we say that this sequence of maps has a *reverse tempered* splitting, if the sequence of maps $A_n^{-1}, \dots, A_1^{-1}$ has a tempered splitting.

In the rest of this section we will show that typically the sequence of differentials along a random orbit has a tempered splitting.

4.2. **Temperedness of sums of real valued random variables.** In order to study the temperedness of vectors, we will first study additive sequences of real random variables. This will be sufficient for our purposes because one may think of the norm of a vector acted upon by matrices as the sum of random variables of the form $\ln \|Av\|\|v\|^{-1}$.

In what follows, we will be studying tempered sequences of sums of real valued random variables. The results of this subsection will be used in the proof of Proposition 4.16, which says that tempered times occur exponentially fast.

Definition 4.3. If $X_1,\dots,X_n$ is a finite or infinite sequence of real numbers then we say that this sequence is (C,λ,ϵ)-*tempered* if for each $0\le j<k\le n$, we have that

$$\sum_{i=j+1}^{k} X_i-\lambda(k-j)+j\epsilon\ge C. \tag{4.6}$$

We also say that a finite sequence $X_1,\dots X_n$ is (C,λ,ϵ)-*reverse tempered* if the sequence $X_n,\dots,X_1$ is (C,λ,ϵ)-tempered.

Note that for fixed $\lambda,\epsilon>0$ every finite sequence is (C,λ,ϵ)-tempered for a sufficiently negative choice of C. Further, note that this condition is harder to satisfy for large positive C, and easier to satisfy for very negative C.

We are interested in finding tempered times for sequences of random variables.

Proposition 4.4. *Fix constants $c>\lambda_0>\lambda_1>0$ and $\epsilon>0$. Then there exist $D_1,D_2>0$ such that the following hold. Suppose that $X_1,X_2,\dots$ is a submartingale difference sequence with respect to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ such that*

(1) $|X_i|\le c$;
(2) $\mathbb{E}[X_i|\mathcal{F}_{i-1}]\ge\lambda_0$.

Then the temperedness constant of the random sequence has an exponential tail. Namely, for $C\ge 0$,

$$\mathbb{P}(X_1,X_2,\dots,\ \textit{is not } (-C,\lambda_1,\epsilon)\textit{-tempered})\le D_1\exp(-D_2C). \tag{4.7}$$

Under the same assumptions on a finite sequence, (4.7) *holds with the same constants.*

Proof. For a fixed C, for the sequence to be $(-C,\lambda_1,\epsilon)$-tempered, for each pair of indices $0\le j<k$ the following inequality must be satisfied:

$$X_k+\dots+X_{j+1}-(k-j)\lambda_1+j\epsilon\ge -C. \tag{4.8}$$

To estimate the probability of this event consider $\chi_{k+1}=\mathbb{E}[X_{k+1}|\mathcal{F}_k]$, and let $\hat{X}_k=X_{k+1}-\chi_{k+1}$. Then the sequence $\hat{X}_k$ is a martingale difference sequence. Then,

$$\mathbb{P}(X_k+\dots+X_{j+1}-(k-j)\lambda_1+j\epsilon\le -C)=\mathbb{P}(\hat{X}_k+\dots+\hat{X}_{j+1}+\sum_{i=j+1}^{k}\chi_i-(k-j)\lambda_1+j\epsilon\le -C)$$

$$\le\mathbb{P}\left(\left|\sum_{i=j+1}^{k}\hat{X}_i\right|\ge\left|-\sum_{i=j+1}^{k}\chi_i+(k-j)\lambda_1-j\epsilon-C\right|\right)\le\mathbb{P}\left(\left|\sum_{i=j+1}^{k}\hat{X}_i\right|\ge|-(k-j)(\lambda_0-\lambda_1)-j\epsilon-C|\right)$$

because we know that the term in the right hand absolute value is negative and $\chi_i\ge\lambda_0>\lambda_1$. Then by Azuma's inequality (Thm. 2.2),

$$\mathbb{P}\left(X_k+\dots+X_{j+1}-(k-j)\lambda_1+j\epsilon\le -C\right)\le 2\exp\left(-\frac{(m(\lambda_0-\lambda_1)+j\epsilon+C)^2}{2mc^2}\right) \tag{4.9}$$

$$\leq 2\exp\left(-\frac{m(\lambda_0-\lambda_1)^2+2(j\epsilon+C)(\lambda_0-\lambda_1)}{2c^2}\right),$$

where $m = k - j$. Summing over j and m we obtain that there exist $D_1, D_2 > 0$ independent of n such that:

$$\sum_{k\geq j+1}^{n} \mathbb{P}(X_k + \cdots + X_{j+1} - (k-j)\lambda_1 + j\epsilon \leq -C) \leq D_1 \exp(-D_2 C), \tag{4.10}$$

which gives the needed conclusion. □

We now estimate the probability that a sequence of random variables as above first fails to be tempered at a time n. This will be used to ensure that failure times in the local coupling lemma have an exponential tail.

Proposition 4.5. *Fix constants $c > \lambda_0 > \lambda_1 > 0$ and $\epsilon > 0$. Then there exists $\eta > 0$ such that the following holds. For each C there exists D_1 such that if $X_1, X_2, \ldots$ is a submartingale difference sequence with respect to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ and*

(1) $|X_i| \leq c$;
(2) $\mathbb{E}[X_i|\mathcal{F}_{i-1}] \geq \lambda_0$,

then if $\mathcal{S}$ is the first n such that $X_1, X_2, \ldots, X_n$ is not (C, λ_1, ϵ)-tempered then:

$$\mathbb{P}(\mathcal{S} \geq n) \leq D_1 e^{-\eta n}.$$

Proof. To obtain a proof of the proposition we show that except on a set of exponentially small probability, the sequence $X_1, \ldots, X_n$ satisfies better estimates than (C, λ_1, ϵ)-temperedness requires for the constraints related on X_{n+1}. In fact, these estimates are so much better than what is needed, that regardless of what X_{n+1} is the sequence will remain (C, λ_1, ϵ)-tempered as long as $X_1, \ldots, X_n$ is (C, λ_1, ϵ)-tempered. Hence the sequence fails to be tempered for the first time at time $n+1$ with exponentially small probability.

We claim that there exist $\eta, D_1 > 0$ such that with probability at least $1 - D_1 e^{-n\eta}$, for all $0 \leq j < n$,

$$\sum_{i=j+1}^{n} X_i - \lambda_1(n-j) + j\epsilon \geq C + (n-j)(\lambda_0 - \lambda_1)/2. \tag{4.11}$$

We now estimate the probability that (4.11) holds for each $0 \leq j < n$. This is the same as estimating the probability that

$$\sum_{i=j+1}^{n} X_i < \lambda_1(n-j) - j\epsilon + C + (n-j)(\lambda_0 - \lambda_1)/2.$$

Note that this is the same inequality as (4.8), with $(n-j)(\lambda_0 - \lambda_1)/2$ added to the constant C appearing there. Thus (4.9) gives

$$\mathbb{P}\left(\sum_{i=j+1}^{n} X_i < \lambda_1(n-j) - j\epsilon + C + \frac{(n-j)(\lambda_0-\lambda_1)}{2}\right) \leq 2\exp\left(-\frac{((n-j)(\lambda_0-\lambda_1)/2 + j\epsilon + C)^2}{2(n-j)c^2}\right)$$

As at least one of j and $n-j$ exceeds $n/2$ in size, we see that there exists $a > 0$ such that

$$\mathbb{P}\left(\sum_{i=j+1}^{n} X_i < \lambda_1(n-j) - j\epsilon + C + (n-j)(\lambda_0-\lambda_1)/2\right) \leq e^{-an}.$$

Hence there exists $D_1 > 0$ such that

$$\sum_{j=0}^{n-1} \mathbb{P}\left(\sum_{i=j+1}^{n} X_i < \lambda_1(n-j) - j\epsilon + C + (n-j)(\lambda_0 - \lambda_1)/2\right) \leq D_1 e^{-(a/2)n}.$$

Thus we see that there is a set of probability $1 - D_1 e^{-(a/2)n}$ such that the inequalities (4.11) all hold. In particular as long as n is sufficiently large, for a realization $X_1, \ldots, X_n$ in this set, it follows that $X_1, \ldots, X_n, X_{n+1}$ is necessarily also (C, λ_1, ϵ)-tempered if $X_1, \ldots, X_n$ is.

This implies that the probability of $X_1, X_2 \ldots$ failing to be (C, λ_1, ϵ)-tempered for the first time at time n is at most $D_1 e^{-(a/2)n}$, and the proposition follows. □

4.3. **Tempered splittings from tempered norms.** In this subsection, we show that one may obtain a tempered splitting for a sequence of matrices in $\mathrm{SL}(2, \mathbb{R})$ when the norms of the matrix products are themselves tempered. Namely, we show that if the norms of a product of matrices has subtempered norm in the sense of Definition 4.1, then the product has a hyperbolic splitting. The proof consists of several steps. The first step is to show that there is a stable subspace on which the product's action is super-tempered.

As before, we write $A^n = A_n \cdots A_1$. We denote by s_n the most contracted singular direction of A^n and by u_n the most expanded singular direction. Recall that for $A \in \mathrm{SL}(2, \mathbb{R})$ we have $\|As\| = \|A\|^{-1}$ where s is a unit vector in the most contracted singular direction.

Before proceeding to the next proof, we see how the most contracted singular direction changes as we compose more matrices. Note that the following computation does not use any temperedness assumptions. Define α_n as follows:

$$s_n = \cos \alpha_n s_{n+1} + \sin \alpha_n u_{n+1}. \tag{4.12}$$

Then we can compute that

$$\|A^{n+1} s_n\| = \sqrt{\|A^{n+1}\|^{-2} \cos^2 \alpha_n + \|A^{n+1}\|^2 \sin^2 \alpha_n} \geq \|A^{n+1}\| \sin \alpha_n.$$

But we also have the estimate:

$$\|A^{n+1} s_n\| \leq \|A_{n+1}\| \|A^n(s_n)\| = \|A_{n+1}\| \|A^n\|^{-1}.$$

Thus

$$\sin \alpha_n \leq \frac{\|A_{n+1}\|}{\|A^{n+1}\| \|A^n\|}. \tag{4.13}$$

We now observe that if the sequence $(A_n)_{n \in \mathbb{N}}$ has a well defined stable direction E^s, then $s_n \to E^s$ and we can estimate their distance by

$$\angle(E^s, s_n) \leq D \sum_{m \geq n} \alpha_m. \tag{4.14}$$

This is good because we expect this sum to be dominated by its first term in the presence of non-trivial Lyapunov exponents.

Now consider a sequence of matrices $A_1, A_2, \ldots$ whose norm is (C, λ, ϵ)-tempered and such that each matrix has norm bounded above by $\Lambda > 0$. If we have $\|A^n v\| \geq e^C e^{n\lambda}$ for some unit vector v, then

$$\angle(E^s, s_n) \leq D \sum_{m \geq n} e^{-2C} \Lambda e^{-2m\lambda} \leq e^{D' - 2C} \Lambda e^{-2n\lambda}, \tag{4.15}$$

for some D' depending only on λ.

Proposition 4.6. *Suppose that $C_0, \lambda, \epsilon, \Lambda > 0$ are fixed. Then there exist D and $N \in \mathbb{N}$ such that if $A_1, \ldots, A_n$, $n \geq N$ is a sequence of matrices in $\mathrm{SL}(2,\mathbb{R})$ with (C, λ, ϵ)-subtempered norms. Then:*

(1) There exist perpendicular vectors s and u so that $(A_i)_{1\le i\le n}$ has a $(\max\{0, -3C\} + D, \lambda - 2\epsilon, 3\epsilon)$ tempered splitting in the sense of Definition 4.2. In the case that $\|A^n\| > 1$, we may take s and u to be the most contracted and expanded singular directions of A^n, respectively.

(2) In the case of an infinite sequence $(A_i)_{i\in\mathbb{N}}$ with subtempered norms there exists an orthogonal pair of unit vectors s and u that defines such a splitting. Further, there exists a unique one dimensional subspace E^s such that any non-zero $v \in E^s$ that satisfies $\limsup_{n\to\infty} n^{-1}\ln\|A^n v\| < 0$ is in E^s.

(3) Finally, there exists $N_0(C) = \lceil (C + \ln(2))/\lambda \rceil$ and D' such that for $n \geq N_0$ and $m_2 \geq m_1 \geq N_0$, and any (C, λ, ϵ)-tempered sequence of matrices $(A_i)_{1\le i\le n}$ as above, A^{m_1} and A^{m_2} have unique contracted singular directions $E^s_{m_1}$ and $E^s_{m_2}$ and moreover,

$$\angle(s_{m_1}, s_{m_2}) \leq e^{-4C+D'} e^{-2(\lambda-\epsilon)m_1}.$$

The analogous statement also holds for $n = \infty$.

Proof. If $\|A^n\| = 1$, choose arbitrarily a vector s_n. Otherwise, let s_n be a unit vector most contracted by A^n. Let s_m be the most contracted vector for A^m. If s_m does not exist because $\|A^m\| = 1$, then there is no most contracted direction, and we instead set $s_m = s_n$. Let u_n be a unit vector in the orthogonal complement of s_n. We show that u_n and s_n define a tempered splitting. This requires estimating three things: the contraction of s_n, the growth of u_n, and the decay of the angle between them.

We now proceed with the proof of (1). First, we will show that the action on the vector s_n is super-tempered. Define α_m as in (4.12). Then there exists some D_1 such that

$$\sin\alpha_m \leq D_1 \frac{\|A_m\|}{\|A^m\|\|A^{m+1}\|}. \tag{4.16}$$

Indeed for indices m where s_m and s_{m+1} are both defined by the actual most contracting directions, this follows as in (4.13). Otherwise, note that one of A^m or A^{m+1} has norm 1, hence the right hand side is uniformly bounded below by $e^{-2\Lambda}$, and thus there exists such a D_1.

From (4.16), it is immediate that there exists $D_2 > 0$ such that

$$\angle(s_m, s_n) \leq D_2 \sum_{m\le j<n} \frac{\|A_j\|}{\|A^j\|\|A^{j+1}\|}. \tag{4.17}$$

From (C, λ, ϵ)-subtempered norms we have for all $m + l \leq n$,

$$\|A^{m+l}\| \geq e^C e^{\lambda l}\|A^m\| e^{-\epsilon m}. \tag{4.18}$$

Combining (4.17) and (4.18), and the uniform bound $\|A\| \leq e^\Lambda$, we get

$$\angle(s_m, s_n) \leq D_2 e^{-2C+2\Lambda}\|A^m\|^{-2} e^{2\epsilon m} \sum_{0\le l<n-m} e^{-2\lambda l} \leq D_2 D_\lambda e^{-2C+2\Lambda}\|A^m\|^{-2} e^{2\epsilon m} \tag{4.19}$$

Hence there exists $D_3 > 0$ such that for all $0 \leq m \leq n$,

$$\begin{aligned}\|A^m s_n\| &\leq \|A^m\|^{-1} + \sin\angle(s_n, s_m)\|A^m\| \leq \|A^m\|^{-1} + D_3 D_\lambda e^{-2C+2\Lambda}\|A^m\|^{-1} e^{2\epsilon m} \\ &\leq (1 + D_3 D_\lambda e^{-2C+2\Lambda} e^{2\epsilon m})\|A^m\|^{-1}. \end{aligned} \tag{4.20}$$

We now check that s_n is supertempered. This is more complicated. Write $\hat{s}_n^k$ for $A^k s_n/\|A^k s_n\|$. For all $j+k \le n$, we have

$$\|A_k^j \hat{s}_n^k\| \|A^k s_n\| = \|A^{j+k} s_n\|.$$

Thus

$$\|A_k^j \hat{s}_n^k\| \le \|A^{j+k} s_n\| \|A^k s_n\|^{-1}.$$

Applying (4.20) with $m = j+k$ we get

$$\|A_k^j \hat{s}_n^k\| \le (1 + D_3 D_\lambda e^{-2C+2\Lambda} e^{2\epsilon(j+k)}) \|A^{j+k}\|^{-1} \|A^k s_n\|^{-1} \tag{4.21}$$

By subtemperedness, $\|A^{j+k}\| \ge e^C e^{j\lambda} e^{-k\epsilon} \|A^k\|$, thus

$$\|A_k^j \hat{s}_n^k\| \le e^{-C} e^{-j\lambda} e^{k\epsilon} (1 + D_3 D_\lambda e^{-2C+2\Lambda} e^{2\epsilon(j+k)}).$$

Hence there exists D_4 such that

$$\|A_k^j \hat{s}_n^k\| \le e^{-\min\{-C,-3C\}+D_4} e^{-j(\lambda-2\epsilon)} e^{3k\epsilon}. \tag{4.22}$$

Thus s_n is $(\max\{0,-3C\}+D_4, \lambda-2\epsilon, 3\epsilon)$-supertempered.

Next we estimate how fast the angle between s_n and $u_n = (s_n)^\perp$ decays. This will lead to a growth estimate on u_n. Consider the angle θ_m between $A^m s_n$ and $A^m u_n$. Because the maps are in $\mathrm{SL}(2,\mathbb{R})$,

$$1 = \|A^m s_n\| \|A^m u_n\| \sin\theta_m. \tag{4.23}$$

Hence by (4.20),

$$\sin\theta_m \ge \frac{1}{\|A^m s_n\| \|A^m\|} \ge (1 + D_3 D_\lambda e^{-2C+2\Lambda} e^{2\epsilon m})^{-1}. \tag{4.24}$$

For $0 \le D_3 D_\lambda e^{-2C+2\Lambda} e^{2\epsilon m} \le 1$,

$$\sin\theta_m \ge 1/2. \tag{4.25}$$

Otherwise, as $1/(1+x) \ge 1/(2x)$ for $x \ge 1$,

$$\sin\theta_m \ge (2D_3)^{-1} D_\lambda^{-1} e^{2C-2\Lambda} e^{-2\epsilon m}. \tag{4.26}$$

In both cases, we see that there exists D_5 such that

$$\sin\theta_m \ge e^{\min\{2C,0\}-D_5} e^{-2\epsilon m}. \tag{4.27}$$

Finally, we estimate the rate of growth of u_n. First, note that because s_n and u_n are orthogonal, applying (4.24) and (4.20) to (4.23) gives

$$\|A^m u_n\| = (\sin\theta_m)^{-1} \|A^m s_n\|^{-1} \ge 1 \cdot (1 + D_3 D_\lambda e^{-2C+2\Lambda} e^{2\epsilon m})^{-1} \|A^m\|.$$

Then letting $\hat{u}_n^k = A^k u_n/\|A^k u_n\|$, we can estimate $\|A_k^j \hat{u}_n^k\|$ as before:

$$\|A_k^j \hat{u}_n^k\| = \|A^{j+k} u_n\| \|A^k u_n\|^{-1} \tag{4.28}$$

$$\ge (1 + D_3 D_\lambda e^{-2C+2\Lambda} e^{2\epsilon(j+k)})^{-1} \|A^{j+k}\| \|A^k\|^{-1} \tag{4.29}$$

$$\ge (1 + D_3 D_\lambda e^{-2C+2\Lambda} e^{2\epsilon(j+k)})^{-1} e^C e^{-\epsilon k} e^{\lambda j} \|A^k\| \|A^k\|^{-1} \tag{4.30}$$

$$= (1 + D_3 D_\lambda e^{-2C+2\Lambda} e^{2\epsilon(j+k)})^{-1} e^C e^{-\epsilon k} e^{\lambda j}. \tag{4.31}$$

If $D_3 D_\lambda e^{-2C+2\Lambda} e^{2\epsilon(j+k)} < 1$, then

$$\|A_k^j \hat{u}_n^k\| \ge \frac{1}{2} e^C e^{-\epsilon k} e^{\lambda j}. \tag{4.32}$$

Otherwise, as $1/(1+x) \geq 1/(2x)$ for $x \geq 1$, we see that there exists $D_5 > 0$ such that:

$$\|A_k^j \hat{u}_n^k\| \geq (2D_3)^{-1} D_\lambda^{-1} e^{2C-2\Lambda} e^{-2\epsilon(j+k)} e^C e^{-\epsilon k} e^{\lambda j} \geq e^{D_5} e^{3C-2\Lambda} e^{-3\epsilon k} e^{(\lambda-2\epsilon)j}. \tag{4.33}$$

So, we see that there exists D_6 such that

$$\|A_k^j \hat{u}_n^k\| \geq e^{\min\{C,3C\}+D_6} e^{(\lambda-2\epsilon)j} e^{-3\epsilon k}, \tag{4.34}$$

which shows that u_n is $(\max\{0,-3C\}+D_6, \lambda-2\epsilon, 3\epsilon)$-subtempered.

We can now conclude by reading off the constants for the splitting we just obtained from equations (4.22), (4.27), and (4.34) and comparing with Definition 4.2. Thus there is D_7 depending only on $\lambda, \Lambda, \epsilon$, such that s_n and u_n define a subtempered splitting with constants:

$$D_7 = (\max\{0,-3C\}+D_7, \lambda-2\epsilon, 3\epsilon). \tag{4.35}$$

This finishes the proof of the first conclusion of the proposition.

The proof of (2) is straightforward, similar to part (1), and very similar to a usual proof of Osceledec theorem [Via14, Ch. 4], so we omit it.

Item (3) also follows from the above proof once we know that N is large enough that the stable subspace is well defined. This certainly holds if $n \geq \lceil (C+\ln(2))/\lambda \rceil$ since then $\|A^n\| \geq 2$. Then from equation (4.19) and temperedness of the norm, if $m_1 \leq m_2$, we have that

$$\begin{aligned}\angle(s_{m_1}, s_{m_2}) &\leq D_2 D_\lambda e^{-2C+2\Lambda} \|A^{m_1}\|^{-2} e^{2\epsilon m_1} \\ &\leq D_2 D_\lambda e^{-2C+2\Lambda} e^{2\epsilon m_1} (e^{-2C} e^{-2m_1\lambda}) \leq e^{-4C+D_8} e^{-2(\lambda-\epsilon)m_1},\end{aligned}$$

for some D_8, which gives item (3). □

4.4. **Tempered splittings for expanding on average diffeomorphisms.** In this subsection, we apply the above developments to describe hyperbolicity of expanding on average random dynamical systems. There are two main results, the first is Proposition 4.8, which is a quantitative estimate on the probability that $D_x f_\omega^n$ has a (C,λ,ϵ)-tempered splitting. The second estimate is Proposition 4.14, which controls the stable direction for this splitting.

To begin, we estimate the probability that the sequence $\|D_x f^n\|$ is tempered.

Proposition 4.7. *For a closed surface M, suppose that $(f_1, \dots, f_m)$ is a uniformly expanding on average tuple in $\mathrm{Diff}_{\mathrm{vol}}(M)$ with constants n_0 and λ_0. Then for all $0 < \lambda_1 < \lambda_0$ and all sufficiently small $\epsilon > 0$, there exists $D, \alpha > 0$ such that for all $x \in M$,*

$$\mu(\{\omega : \|D_x f_\omega^n\| \text{ is not } (-C,\lambda_1,\epsilon)\text{−subtempered}\}) \leq De^{-\alpha C}. \tag{4.36}$$

Proof. This follows from the estimates on temperedness obtained for submartingales. Essentially, for a fixed $v \in T_x^1 M$, $X_n = \|D_x f_\omega^{nn_0} v\|$ is a submartingale with respect to a filtration $\mathcal{F}_n$ generated by the coordinates of ω, and $\mathbb{E}[X_n|\mathcal{F}_{n-1}] \geq \lambda_0$. Thus Proposition 4.4 gives that for all sufficiently small $\epsilon > 0$, and $0 < \lambda_1 < \lambda_0$, there exist $D_1, D_2 > 0$ such that:

$$\mathbb{P}(\|D_x f^{nn_0}\| \text{ is not } (-C,\lambda_1,\epsilon)\text{-tempered}) \leq D_1 e^{-D_2 C}.$$

Then to obtain temperedness along the entire sequence, not just times of the form nn_0, note that we have a uniform bound on the norm and conorm of all $\|D_x f_{\omega_i}\|$, $1 \leq i \leq m$. □

Since a tempered sequence of norms implies the existence of a tempered splitting by Proposition 4.6, the following is immediate.

Proposition 4.8. *Suppose that M is a closed surface and $(f_1,\ldots,f_m)$ is uniformly expanding on average tuple of diffeomorphisms in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$ with expansion constant λ_0. Then for all $0<\lambda_1<\lambda_0$, and sufficiently small $\epsilon>0$, there exists $D,\alpha>0$ such that for all $x\in T^1M$,*

$$\mu(\{\omega : D_xf^n_\omega \text{ does not have a } (C,\lambda,\epsilon)-\text{tempered splitting}\})\le De^{-\alpha C}. \tag{4.37}$$

In particular, for all $x\in M$ and almost every ω, $D_xf^n_\omega$ has a well defined one-dimensional stable subspace $E^s_\omega(x)$.

Below, it will be important to consider the probability that a trajectory that is (C,λ,ϵ)-tempered suddenly fails to be tempered. In order to quantify this we will introduce an auxiliary quantity for (C,λ,ϵ)-tempered orbits of length n. We call this the *cushion* of the orbit and it measures how far the inequalities from Definition 4.1(1) are from failing.

Definition 4.9. If the sequence of matrices $A_1,\ldots,A_n$ is (C_0,λ,ϵ)-tempered, then we define its *cushion* U to be

$$U=\min_{0\le k<n}\left[\ln\|A^n\|-\ln\|A^k\|-C_0-(n-k)\lambda+\epsilon k\right]$$

Note that a trajectory can have such a large cushion that whatever happens at the next iterate, the trajectory will not fail to be tempered. The cushion reflects the only inequalities relevant to tempering that the term A_{n+1} would affect, should it be added to the sequence.

The following proposition is a large deviations estimate that says that typically the cushion is quite large.

Proposition 4.10. *For a closed surface M, suppose that $(f_1,\ldots,f_m)$ is an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$ with expansion constant $\lambda_0>0$. For fixed C_0, let $U(n,\omega,x)$ be the cushion of D_xf^n when viewed as a (C_0,λ,ϵ)-tempered trajectory.*

Then for any C_0, $\lambda<\lambda_0$, and $\epsilon>0$, there exist $\delta,\eta,D>0$ such that

$$\mathbb{P}(U(n,\omega,x)<n\delta|D_xf^n_\omega \text{ is } (C_0,\lambda,\epsilon)\text{-tempered})\le De^{-\eta n}.$$

Proof. The proof is straightforward: we are just estimating the difference between $\ln\|D_xf^n_\omega\|$ and $\ln\|D_xf^i_\omega\|$.

Note that in order for a given trajectory to fail to have a cushion of size $\bar\epsilon n$, it needs to be the case that for each $0\le k\le n$, that

$$\bar\epsilon n>\ln\|D_xf^n_\omega\|-\ln\|D_xf^k_\omega\|-C_0-\lambda(n-k)+\epsilon k. \tag{4.38}$$

Call this event $\Omega_{n,k}$. Note that this event is a subset of the event that

$$\bar\epsilon n+C_0\ge\ln\|D_xf^n_\omega\|-\ln\|D_xf^k_\omega\|-\lambda(n-k)$$

As before, $\ln\|D_xf^n_\omega\|-\ln\|D_xf^k_\omega\|-\lambda(n-k)$ is a submartingale with differences bounded by some $\Lambda>0$. Hence as $\bar\epsilon n+C_0$ is positive for n sufficiently large, it is less than the expectation of $\ln\|D_xf^n_\omega\|-\ln\|D_xf^k_\omega\|-\lambda(n-k)$. Thus Azuma's inequality gives

$$\begin{aligned}\mathbb{P}(\Omega_{n,k})&\le\mathbb{P}\left(\left|\ln\|A^n\|-\ln\|A^k\|-\mathbb{E}\left[\ln\|A^n\|-\ln\|A^k\|\right]\right|>\bar\epsilon n+C_0\right)\\&\le 2\exp\left(-\frac{(\bar\epsilon n+C_0)^2}{2\Lambda n}\right)\le C_1\exp\left(-\frac{\bar\epsilon}{2\Lambda}n\right).\end{aligned}$$

Summing over k, we find that the probability that at least one of the inequalities (4.38) fails for $1\le k\le n$ is exponentially small, which gives the result. □

Next, we study the distribution of the stable subspaces in an expanding on average system. We obtain two estimates. First, we obtain an estimate on the distribution of all stable subspaces through a point, Proposition 4.11. Second, in Proposition 4.14, we show that the empirical distribution of stable subspaces converges quickly to the actual distribution of the true stable subspaces.

Proposition 4.11. *Suppose that M is a closed surface and that $(f_1, \dots, f_m)$ is an expanding on average tuple of diffeomorphisms in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. Then there exist constants $C, \alpha > 0$ such that if ν^s_x denotes the distribution of stable subspaces through the point x, then for each $v \in \mathbb{P}T_xM$,*

$$\nu^s_x(\{z \mid d(z, v) \leq \epsilon\}) \leq C\epsilon^\alpha,$$

where d is the angle between those points and $\mathbb{P}(T_xM)$ denotes the projectivization of T_xM.

Naturally, before proceeding with the proof, we must show for $v \in T^1M$ that the norm of $D_xf^n_\omega v$ along a typical trajectory does grow exponentially. In fact, we show that even slow exponential growth is quite unlikely.

Lemma 4.12. *In the setting of Proposition 4.11, suppose that* (1.1) *holds with constants $n_0 \in \mathbb{N}$ and $\lambda_0 > 0$. Then there exist $\gamma, C > 0$ such that if $v \in T^1M$, then*

$$\mathbb{P}_\omega(\|Df^n_\omega v\| \leq e^{\lambda_0 n/3}) \leq Ce^{-\gamma n}. \tag{4.39}$$

Proof. First, note that by considering the Taylor expansion of e^{-t}, that for sufficiently small t and all $v \in T^1M$,

$$\mathbb{E}\left[e^{-t\ln\|Df^{n_0}_\omega v\|}\right] \leq (1 - (n_0\lambda_0/2)t).$$

Next, observe that writing $\overline{v}$ for $v/\|v\|$,

$$\begin{aligned}\mathbb{E}\left[e^{-t\ln\|Df^{2n_0}_\omega v\|}\right] &= \mathbb{E}\left[e^{-t\ln\|Df^{n_0}_\omega v\|}e^{-t\ln\|Df^{n_0}_{\sigma^{n_0}(\omega)}(\overline{Df^{n_0}_\omega v})\|}\right] \\ &\leq \mathbb{E}\left[e^{-t\ln\|Df^{n_0}_\omega v\|}(1 - (n_0\lambda_0/2)t)\right] \leq (1 - (n_0\lambda_0/2)t)^2,\end{aligned}$$

where we have used the independence of $\sigma^{n_0}\omega$ from ω_i for $i < n_0$. Similarly, by boundedness of the C^1 norm of the f_i, we see inductively that there exists $D > 0$ such that for all n,

$$\mathbb{E}\left[e^{-t\ln\|Df^n_\omega v\|}\right] \leq D\,(1 - (n_0\lambda_0/2)t)^{n/n_0} \leq e^{-n\lambda_0/2},$$

since $1 - t/2 < e^{-t}$ for small t. By Markov's inequality

$$\begin{aligned}&\mathbb{P}(\|Df^n_\omega v\| \leq e^{\lambda_0 n/3}) \leq \mathbb{P}(e^{-t\ln\|Df^n_\omega v\|} \geq e^{-t\lambda_0 n/3}) \\ &\leq \frac{\mathbb{E}\left[e^{-t\ln\|Df^n_\omega v\|}\right]}{e^{-t\lambda_0 n/3}} \leq D\frac{(1 - (n_0\lambda_0/2)t)^{n/n_0}}{e^{-t\lambda_0 n/3}} \leq De^{-n\lambda_0 t/2 + \lambda_0 nt/3} \leq De^{-n\lambda_0 t/6}. \qquad \square\end{aligned}$$

For $v \in T^1M$, let $B_\epsilon(v)$ be the set of directions w with $\sin(\angle(v, w)) \leq \epsilon$ and Λ be the maximum of the norm of $\|D_xf_i\|$ over the set of all $1 \leq i \leq m$ and $x \in M$.

Lemma 4.13. *For all $\sigma > 0$ sufficiently small there exist $0 < \theta < 1$ such that for any $v \in \mathbb{P}(T_xM)$ and sufficiently small $\epsilon > 0$, if $-\frac{\lambda_0}{6\Lambda}\ln(\epsilon) \leq n \leq -\frac{\lambda_0}{3\Lambda}\ln(\epsilon)$, and*

$$\delta = \max_{u \in B_\epsilon(v)} \sin\angle(Df^n_\omega u, Df^n_\omega v),$$

then

$$\mathbb{P}(\delta \leq \epsilon^{1+\sigma} \text{ and for all } u \in B_\epsilon(v),\ \|Df^n_\omega u\| \geq 2^{-1}e^{n\lambda_0/3}\|u\|) \geq 1 - \epsilon^\theta.$$

Proof. By Lemma 4.12, for each n we have $\|Df_\omega^n v\| \geq e^{\lambda_0 n/3}$ on a set of measure $1 - Ce^{-\gamma n}$. Then for any unit vector u with $\sin(\angle(v,u)) \leq \epsilon$,

$$\|Df_\omega^n u\| \geq \|Df_\omega^n v\| - \|Df_\omega^n (u-v)\| \geq e^{\lambda_0 n/3} - \epsilon e^{\Lambda n} \geq e^{\lambda_0 n/3}/2,$$

as long as ϵ is sufficiently small and n satisfies $n \leq -\frac{\lambda_0}{3\Lambda}\ln(\epsilon)$.

Since the f_i are volume preserving, the areas of the triangles between vectors are preserved. Since all vectors in $B_\epsilon(v)$ are stretched, we see that

$$\sin\angle(Df_\omega^n v, Df_\omega^n u) = \epsilon \|Df_\omega^n v\|^{-1}\|Df_\omega^n u\|^{-1} \leq 2\epsilon e^{-(2/3)\lambda_0 n}.$$

But if $n \geq -\frac{\lambda_0}{6\Lambda}\ln(\epsilon)$ and ϵ is sufficiently small, then $\sin\angle(Df_\omega^n v, Df_\omega^n u) \leq 2\epsilon e^{-\frac{2}{3}\lambda_0\frac{\lambda_0}{6\Lambda}(-\ln(\epsilon))}$. Thus we see that for sufficiently small ϵ and $\sigma > 0$ that for n satisfying

$$-\frac{\lambda_0}{6\Lambda}\ln(\epsilon) \leq n \leq -\frac{\lambda_0}{3\Lambda}\ln(\epsilon)$$

it holds that $\sin\angle(Df_\omega^n v, Df_\omega^n u) \leq \epsilon^{1+\sigma}$ for all ω in a set of size $1 - C\epsilon^{-\gamma n}$. □

Proof of Proposition 4.11. Using Lemma 4.13 we may now conclude. Fix some $\sigma > 0$ as in the lemma, $\Lambda/(6\lambda_0) < \alpha < \Lambda/(3\lambda_0)$ and let $\epsilon > 0$ be small enough that the lemma applies. Let $\epsilon_1 = \epsilon$ and then define $\epsilon_k = \epsilon^{(1+\sigma)^k}$. Let $b_k = \lfloor -\alpha(1+\sigma)^k \ln(\epsilon)\rfloor$ and $n_k = \sum_{k=0}^{k-1} b_k$ be an increasing sequence of times. By our choice of α we may apply the lemma to each additional block of iterations of f_ω of length b_k with $\epsilon = \epsilon_k$. We then define:

$$\eta_k^\omega(\epsilon, v) = \max_{w \in B_{\epsilon_k}(Df_\omega^{n_{k-1}} v)} \sin\angle(Df_\omega^{b_k} w, Df_\omega^{b_k} v),$$

$$\tau_k^\omega(\epsilon, v) = \inf_{w \in B_{\epsilon_k}(Df_\omega^{n_{k-1}} v)} \|Df^{b_k}_{\sigma^{n_{k-1}}\omega} w\|.$$

Lemma 4.13 asserts that for every v and k that

$$\mathbb{P}(\eta_k^\omega(\epsilon_k, v) \leq \epsilon_k^{1+\sigma} \text{ and } \tau_k^\omega(\epsilon_k, v) \geq 2^{-1}e^{\lambda_0(n_k - n_{k-1})/3}) \geq 1 - \epsilon_k^\theta.$$

As the dynamics is IID and the above estimate is independent of the vector $v \in \mathbb{P}(TM)$, we see that there exists $C > 0$ such that:

$$\mathbb{P}\left(\text{for all } k\ \eta_k^\omega(\epsilon, v) \leq \epsilon_k \text{ and } \tau_k^\omega(\epsilon, v) \geq \frac{e^{\lambda_0 n_k/3}}{2}\right) \geq \prod_{i=1}^{\infty}\left(1 - \epsilon_k^\theta\right) \geq 1 - C\epsilon^\theta. \tag{4.40}$$

By Proposition 4.8, at the point x almost every word ω has a well defined stable subspace $E^s_\omega(x)$. If a vector $v \in T^1_x M$ satisfies (4.40), then for any $w \in B_\epsilon(v)$, $\|Df_\omega^{n_k} w\| \geq e^{\lambda_0 n_k/3}2^{-k}$, which grows rapidly in k as long as ϵ was chosen sufficiently small. Thus this vector cannot be in $E^s_\omega(x)$. Thus $\mathbb{P}(E^s_\omega(x) \in B_\epsilon(v)) \leq C\epsilon^\theta$, and we are done. □

Next we check that if we consider the distribution of stable subspaces for finite time realizations of the dynamics that the distribution of the finite time stable subspaces converges quickly to the stationary stable distribution. Essentially this should be true for the same reason that it is true for IID matrix products. The proof is a slight extension of the argument that appears above.

Proposition 4.14. *Suppose that M is a closed surface and $(f_1, \ldots, f_m)$ is an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. There exist c_0, C, θ such that for any $x \in M$ and $v \in T^1_x M$, if*

$N_0 \geq c_0 |\ln(\epsilon)|$ *the following holds. Let* $E^s_n(\omega)$ *be the maximally contracted subspace of the product* $D_x f^n_\omega$*. Then:*

$$\mathbb{P}(\textit{for some } n > N_0, E^s_n(\omega) \in B_\epsilon(v) \textit{ or } E^s_n(\omega) \textit{ does not exist}) \leq C\epsilon^\theta. \tag{4.41}$$

Proof. The proof of the above fact is essentially a corollary of the estimates obtained in the proof of Lemma 4.13.

We apply that same proof and choose sufficiently small $0 < \sigma < \lambda_0/(3\Lambda)$ where λ_0 and Λ are as in that proposition, as are b_k and n_k. Then we find that there exists C, θ such that for all sufficiently small $\epsilon > 0$, we have equation (4.40), so for $\epsilon_k = \epsilon^{(1+\sigma)^k}$,

$$\mathbb{P}(\text{for all } k\ \delta^\omega_k(\epsilon, v) \leq \epsilon_k \text{ and } \tau^\omega_k(\epsilon, v) \geq 2^{-1} e^{\lambda_0 n_k/3}) \geq 1 - C\epsilon^\theta. \tag{4.42}$$

This shows as before that at the times n_k, that we have the estimate

$$\|Df^{n_k}_\omega w\| \geq e^{\lambda_0 n_k/3} 2^{-k}$$

for all $w \in B_\epsilon(v)$ on a set of measure $1 - C\epsilon^\theta$. In particular, as we chose σ quite small, for $k \geq 2$, we see that for any time n from n_{k-1} to n_k, that

$$\|Df^n_\omega w\| \geq \|Df^{n_{k-1}}_\omega w\| e^{-(n-n_{k-1})\Lambda} \geq e^{n_{k-1}\lambda_0/3-(n-n_{k-1})\Lambda}.$$

But by choice of σ, that exponent is at least

$$((1+\sigma)^{k-1}\lambda_0/3 - ((1+\sigma)^k - (1+\sigma)^{k-1})\Lambda)\ln(\epsilon) = (1+\sigma)^{k-1}(\lambda_0/3 - \sigma\Lambda)\ln(\epsilon) > 0.$$

Thus from the definition of the n_k in Lemma 4.13, we see that on a set of probability $1 - C\epsilon^\theta$ for any $n > n_1 = |\alpha(1+\sigma)\ln(\epsilon)|$, that $E^s_n(\omega)$ does not lie in $B_\epsilon(v)$ and the result follows. □

4.5. **Reverse tempered sequences.** We are interested in reverse tempered times since they are key for proving smoothing lemmas. The main result of this subsection is Proposition 4.18, which shows that the waiting time until a reverse tempered time occurs has an exponential tail.

The following lemma estimates how much the temperedness of a sequence improves when we prepend entries on it. Note that by reversing the order of the sequence, this gives the corresponding estimate for reverse temperedness.

Lemma 4.15. *Suppose that* $a_1, \ldots, a_n$ *is a* (C, λ_0, ϵ) *tempered sequence and* $b_1, \ldots, b_m$ *is a* $(D, \lambda_1, \epsilon/2)$ *tempered sequence where* $\lambda_1 - \lambda_0 > \epsilon$*, then* $b_1, \ldots, b_m, a_1, \ldots, a_n$ *is*

$$(\min\{D, m\epsilon/2 + C + D, m\epsilon + C\}, \lambda_0, \epsilon)$$

tempered sequence.

Proof. Let $c_1, \ldots, c_{m+n}$ denote the new joined sequence and let C' be the (λ_0, ϵ) temperedness constant for this sequence. Each pair of indices $0 \leq j < k \leq n+m$ gives a constraint on the constant of temperedness:

$$C' = \min_{0 \leq j < k \leq n+m} j\epsilon + \sum_{i=j+1}^{k} (c_i - \lambda_0). \tag{4.43}$$

Note that the only pairs of indices that offer a non-trivial constraint are those with at least one of $j+1, k \geq m+1$. The constraint arising from a pair of indices with $j, k \leq m$, is certainly satisfied as long as the temperedness constant is at most D. This leaves two cases.

For a pair of indices $j < m < k$, we obtain the constraint that

$$C' \leq j\epsilon + \sum_{i=j+1}^{m} (b_i - \lambda_0) + \sum_{i=m+1}^{k} (a_i - \lambda_0). \tag{4.44}$$

But by temperedness, we can bound the right hand side below:

$$j\epsilon + \sum_{i=j+1}^{m} (b_i - \lambda_0) + \sum_{i=m+1}^{k} (a_i - \lambda_0) \geq D + \frac{j\epsilon}{2} + (m-j)(\lambda_1 - \lambda_0) + C \geq m\epsilon/2 + D + C.$$

If both $j+1, k \geq m+1$, then as the sequence $a_1, \ldots, a_m$ is already (C, λ_0, ϵ)-tempered, the constraint on these entries of the sequence improves by $m\epsilon$ as they are now additionally offset by m from 0. So, they give the constraint $C' \leq C + m\epsilon$.

Taking the minimum over the three bounds above gives the result. □

Using the above, we will now prove that for submartingale difference sequences the renewals of backward temperedness have exponential tails.

Proposition 4.16. *(Exponential return times to the tempered set) Fix $c > \lambda_0 > \lambda > 0$ and pick $0 < \epsilon < (\lambda_0 - \lambda)/3$. There exist $C_0, D_1, D_2 > 0$ such that the following holds. Let $X_1, X_2, \ldots$ be a submartingale difference sequence with respect to a filtration $(\mathcal{F}_n)_{n\in\mathbb{N}}$ such that for all $n \in \mathbb{N}$,*

(1) $|X_n| < c$;
(2) $\mathbb{E}[X_n|\mathcal{F}_{n-1}] \geq \lambda_0$.

Fix $N \in \mathbb{N}$ and let T denote the first time k after N such that $X_1, \ldots, X_{N+k}$ is (C_0, λ, ϵ)-reverse tempered. Then

$$\mathbb{P}(T > N + k) \leq D_1 e^{-D_2 k}. \tag{4.45}$$

Proof. The proof has essentially two steps. First, in the following claim, we study how long it takes for a sequence with bad temperedness constant to recover. This happens with linear speed because we are studying a submartingale sequence with $\mathbb{E}[X_n|\mathcal{F}_{n-1}]$ uniformly bounded away from zero. We estimate how fast the reverse-temperedness constant improves as we append blocks of a fixed size Δ_0. As a sequence of length N might have a bad temperedness constant, to obtain the result we then apply the tail estimate on the temperedness constant for sequences of length N. As each of these things has an exponential tail, we obtain the result.

The main claim is the following.

Claim 4.17. *There exist C_0 and $A, B > 0$ independent of N, such that if $X_1, \ldots, X_N$ is (R, λ, ϵ)-tempered and T is the first time greater than N that is (C_0, λ, ϵ)-reverse tempered, then*

$$\mathbb{P}(T > N + k | X_1, \ldots, X_N \textit{ is } (R, \lambda, \epsilon)\textit{-tempered}) \leq A e^{R - Bk}.$$

Proof. Let $\lambda_1 = (\lambda + \lambda_0)/2$ and denote by $B_{i,\Delta}$ the backwards $(\lambda_1, \epsilon/2)$-temperedness constant of the sequence $X_{i+1}, \ldots, X_{i+\Delta}$. By Proposition 4.4, there exist A_2, B_2 (independent of i and Δ) such that for $C \geq 0$,

$$\mathbb{P}(X_{i+1}, \ldots, X_{i+\Delta} \text{ is not } (-C, \lambda, \epsilon)\text{-tempered}) \leq A_2 e^{-B_2 C}.$$

As this tail on the temperedness constant is independent of i and Δ, we see that there exists Δ_0 sufficiently large and $\delta > 0$ such that for any $i \in \mathbb{N}$,

$$\mathbb{E}[\Delta_0 \epsilon/2 + B_{i,\Delta_0} | \mathcal{F}_i] > \delta > 0. \tag{4.46}$$

We now check how much appending a block of length Δ_0 improves temperedness. Let C_i' denote the backwards (λ, ϵ)-temperedness constant of the sequence

$$X_1, \dots, X_N, X_{N+1}, \dots, X_{N+i\Delta_0}.$$

and let D_i denote the $(\lambda_1, \epsilon/2)$ backwards tempered constant of the sequence

$$X_{N+(i-1)\Delta_0+1}, \dots, X_{N+i\Delta_0}.$$

Then by Lemma 4.15,

$$C_{i+1}' = \min\{D_{i+1}, \epsilon\Delta_0/2 + D_{i+1} + C_i', \epsilon\Delta_0 + C_i'\}.$$

We also define $\hat{C}_0 = C_0'$ and

$$\hat{C}_{i+1} = \min\{\epsilon\Delta_0/2 + D_{i+1} + \hat{C}_i, \epsilon\Delta_0 + \hat{C}_i\}.$$

Note that by (4.46) there exists $\delta > 0$ depending only on $c, \lambda, \lambda_1, \epsilon$, such that

$$\mathbb{E}\left[\hat{C}_{i+1} | \mathcal{F}_{N+i\Delta_0}\right] - \hat{C}_i \geq \delta > 0. \tag{4.47}$$

Suppose that we define T so that we decide to stop when $C_i' \geq -\epsilon\Delta_0/2$. Observe that if $i+1$ is the first index such that $\hat{C}_{i+1} \geq 0$ then because

$$\hat{C}_{i+1} \geq \epsilon\Delta_0/2 + D_{i+1} + \hat{C}_i,$$

and $\hat{C}_i < 0$ we must have that $D_{i+1} \geq -\epsilon\Delta_0/2$. Thus

$$C_{i+1}' \geq \min\{D_{i+1}, \epsilon\Delta_0/2 + D_{i+1} + C_i', \epsilon\Delta_0 + C_i'\} \geq -\epsilon\Delta_0/2. \tag{4.48}$$

Let $C_0 = -\epsilon\Delta_0/2$. Thus if k is the first index such that $\hat{C}_k \geq 0$, then $T < n + \Delta_0 k$. Thus we need to obtain a bound for the first time $\hat{C}_i \geq 0$.

We now bound the tail on the first time $\hat{C}_i \geq 0$. Note that $\hat{C}_i$ is a submartingale. Further let M be an upper bound on $\left|C_{i+1}' - C_i'\right|$ over all i (an upper bound exists because $|X_i| < c$). Let $\chi_i = \mathbb{E}\left[\hat{C}_i | \mathcal{F}_{n+(i-1)\Delta_0}\right] \geq \delta > 0$. Then $\beta_i = \hat{C}_{i+1} - \chi_i$ is a martingale difference sequence. We now estimate:

$$\mathbb{P}(\hat{C}_k \leq 0) \leq \mathbb{P}\left(-R + \sum_{i=1}^{k} \beta_k \leq -\sum_{i=0}^{k-1} \chi_i\right) \leq \mathbb{P}\left(\sum_{i=1}^{k} \beta_k \leq -k\delta + R\right)$$

Thus for $k \geq R/\delta$, by Azuma's inequality (Theorem 2.2),

$$\mathbb{P}(\hat{C}_k \leq 0) \leq 2\exp\left(-\frac{(k\delta - R)^2}{2kM^2}\right) \leq 2\exp\left(-\frac{k\delta^2}{2M^2} + \frac{R\delta}{M^2} - \frac{R^2}{2kM^2}\right) \leq 2\exp\left(-k\frac{\delta^2}{2M^2} + R\left(\frac{\delta}{M^2}\right)\right).$$

If $\delta/M^2 \leq 1$, then we are already done with $B = \delta^2/(2M^2\Delta_0)$. Otherwise, if $\delta/M^2 > 1$, then for $k \geq 2R/\delta$, which is the only range where the bound is less than 1, the right hand side is bounded above by

$$2\exp\left(-k\frac{\delta^2}{2M^2} + R\left(\frac{\delta}{M^2}\right)\right) \leq 2\exp\left(R - k\frac{\delta^2}{2M^2}\frac{M^2}{\delta}\right),$$

and thus the estimate holds with $B = \delta/(2\Delta_0)$ in this case as well. This finishes the proof of the claim. □

Let A, B and C_0 be as in the claim. From Proposition 4.4, there exists D_1, D_2 such that for all $C \geq 0$,

$$\mathbb{P}(X_1, \ldots, X_N \text{ is } (-C, \lambda, \epsilon)\text{-tempered}) \geq 1 - D_1 \exp(-D_2 C).$$

From the claim we know that if $X_1, \ldots, X_N$ is $(-C, \lambda, \epsilon)$-tempered and T is the waiting time for a future (C_0, λ, ϵ)-tempered time, then

$$\mathbb{P}(T > N + k) \leq A e^{C - Bk}.$$

Combining these two estimates we see that

$$\begin{aligned}\mathbb{P}(T > N + k) \leq & \mathbb{P}(X_1, \ldots, X_N \text{ is } (-Bk/2, \lambda, \epsilon)\text{-tempered and } T > N + k) \\ & + \mathbb{P}(X_1, \ldots, X_N \text{ is not } (-Bk/2, \lambda, \epsilon)\text{-tempered}) \\ & \leq A \exp(Bk/2 - Bk) + D_1 \exp(-D_2 Bk/2) \\ & \leq A \exp(-Bk/2) + D_1 \exp(-D_2 Bk/2).\end{aligned}$$

The conclusion is now immediate. □

The above results imply that expanding on average diffeomorphisms have frequent reverse tempered times.

Proposition 4.18. *Suppose that $(f_1, \ldots, f_m)$ is an expanding on average tuple of diffeomorphisms in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. There exist $\lambda > 0$ such that for all sufficiently small $\epsilon > 0$, there exists C_0, C, α such that for all $x \in M$ and $N \in \mathbb{N}$, if we let $T(x)$ be the first (C_0, λ, ϵ)-reverse tempered time for $\|D_x f^n_\omega\|$ that is greater than or equal to N, then*

$$\mathbb{P}(T(x) \leq N + k) \geq 1 - C e^{-\alpha k},$$

and $D_x f_\omega^{T(x)}$ has a well defined splitting into maximally expanded and contracted singular directions.

Proof. $X_n = \|D_x f^{n n_0}\|$ is a submartingale satisfying the hypotheses of Proposition 4.16, hence X_n satisfies the required estimate on reverse tempered times. The last claim follows from Proposition 4.6. □

Proposition 4.18 shows that there is a uniformly large density subset of points such that $D_x f^n_\omega$ is reverse tempered. We now show that the stable direction of the resulting tempered splitting does not lie too close to any particular vector v.

Lemma 4.19. *Suppose that $(f_1, \ldots, f_m)$ is an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$, for M a closed surface. There exist D, α, c_0 and $C, \lambda > 0$ such that for all sufficiently small $\epsilon > 0, x \in M$ and interval $I \subset T^1_x M$, if $n \geq c_0 \ln |I|$, where $|I|$ is the length of I, if $T(x)$ is the first time greater than n that the sequence $D_x f^n_\omega$ has a (C, λ, ϵ) reverse tempered splitting, denoting the most contracted direction of $D_x f^n_\omega$ by E^s_T,*

$$\mathbb{P}(E^s_T \in I | T(x) \leq n + k) \leq C |I|^\alpha.$$

Proof. This probability equals $\dfrac{\mathbb{P}(E^s_T \in I \text{ and } T(x) \leq n + k)}{\mathbb{P}(T(x) \leq n + k)}$. By Proposition 4.18, the denominator is at least $1 - C_1 e^{-k C_2}$, for some C_1, C_2. If c_0 is as in Proposition 4.14, then for $n \geq c_0 \ln |I|$, then the numerator is bounded above by $\mathbb{P}(E^s_T \in I) \leq C_3 |I|^\alpha$. □

5. Stable manifolds of expanding on average systems

In this section we show Proposition 5.3, which says that with probability $1-C^{-\alpha}$ a point has a stable manifold of length at least C. The proof has two parts. First we state a abstract proposition that gives the existence of a stable manifold with good properties through a point x provided that there exists a tempered hyperbolic splitting along the orbit of x. We then estimate the probability that this criterion holds.

In §2.3 we introduced the stable manifolds for the random dynamics. We now introduce a quantitative property of them that will be of use later.

Definition 5.1. We say that a stable manifold $W^s(\omega, z)$ is (C, λ, ϵ)-tempered if the length of $W^s(\omega, z)$ is at least C^{-1} and the points in the stable manifold attract uniformly quickly: for $x, y \in f^n_\omega(W^s_{C^{-1}}(\omega, z))$,

$$d_{f^{n+m}_\omega(W^s_{C^{-1}}(\omega,z))}(f^m_{\sigma^n(\omega)}(x), f^m_{\sigma^n(\omega)}(y)) \le Ce^{-\lambda m}e^{\epsilon n}.$$

Now we give a quantitative estimate on the number of stable curves of a given C^2 norm and length. This result follows from a careful reading of the construction of stable manifolds in the book of Liu and Qian [LQ95], in particular, Theorem III.3.1, which constructs stable manifolds of random dynamical systems lying in a certain type of Pesin block that the authors denote by $\Lambda^{l,r}_{a,b,k,\epsilon}$. In the case that the random dynamics only arises from a finite collection of diffeomorphisms (i.e. has bounded C^2 norm), the constraint from the r parameter does not matter—r essentially measures how small a neighborhood of x one must look at for the map in an exponential chart to be uniformly close to its derivative. In our setting, once we pick sufficiently large $r_0 > 0$ there is no constraint. The number k is our case also does not matter—it specifies the dimension of the splitting we are considering.

In the 2-dimensional setting a point $x \in M$ lies in $\Lambda^{l,r}_{a,b,k,\epsilon}$ for the sequence of diffeomorphisms $f_1, f_2, \ldots$ if, writing $f^{n+k}_n = f_{n+k}\cdots f_{n+1}$, we have an invariant splitting along the trajectory $E^s_{f^n(x)} \oplus E^u_{f^n(x)}$ such that for the reference metric on the manifold we have that:

$$\left|Df^{n+k}_n(f^n(x))|_{E^s}\right| \le le^{\epsilon n}e^{(a+\epsilon)k}$$
$$\left|Df^{n+k}_n(f^n(x))|_{E^u}\right| \ge l^{-1}e^{-\epsilon n}e^{(b-\epsilon)k}$$
$$\angle(E^s_{f^n_1(x)}, E^u_{f^n_1(x)}) \ge l^{-1}e^{-\epsilon n}.$$

This is defined at the beginning of [LQ95, Sec. 3]. In the language we have been using above, a $(-C, \lambda, \epsilon)$-tempered trajectory belongs to the set $\Lambda^{e^C, r_0}_{\lambda,-\lambda,1,\epsilon}$. From [LQ95, Thm. III.3.1], we may now deduce the following proposition.

Proposition 5.2. *Suppose that $(f_1, \ldots, f_m)$ is a tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$, where M is a closed surface. Fix $\lambda, \epsilon > 0$. Then there exist constants D_1, D_2 such that if (ω, x) is a $(-C, \lambda, \epsilon)$-tempered trajectory, then $W^s_\omega(x)$ exists and is at least $D_1 e^{-2C}$ long. Further, on this interval, its C^2 norm is at most $D_2 e^{6C}$ (when viewed as a graph over its tangent space at x). Moreover these estimates are $e^{7\epsilon}$-tempered along the trajectory.*

Proof. From the above discussion, a (C, λ, ϵ)-tempered point lies in $\Lambda^{e^C, r_0}_{\lambda,-\lambda,1,\epsilon}$. So, we just need to recover the estimates from the proof of [LQ95, Thm. III.3.1]. In fact these estimates are stated there. As we are keeping λ, ϵ fixed, the conclusion will follow once we compute the quantities α_n and β_n appearing in that theorem given our particular choices. Although [LQ95] only shows the stable manifolds are $C^{1,1}$, the estimates provided there on the Lipschitz constant of the derivative is enough for controlling the C^2 norm because we know that the stable manifolds are in fact as smooth as the dynamics, which is C^2 [Arn98, Rem. 7.3.20].

First we explain how to estimate β_n, which controls the norm. The first quantity that gets defined in the proof is $c_0 = 4Ar'e^{2\epsilon}$. Here, A is the quantity appearing in the proof of [LQ95, Lem. 1.3], which is equal to $4(l^2)(1-\epsilon^{-2\epsilon})^{-1/2}$. Thus $c_0 \le C_1 e^{2C}$. Therefore the quantity $D = (1-e^{-2\epsilon})^{-3}(1+e^{-2\epsilon})^2 c_0 e^{-a}$ on p. 66 of [LQ95] is at most $C_2 e^{2C}$. Hence β_n, which is defined on p. 68 of [LQ95] as $2DA^2 e^{7\epsilon n}$ and controls the norm of the stable curve, is at most $C_3 e^{6C} e^{7\epsilon n}$.

The length of the curve given by the quantity α_n defined on p. 68 of [LQ95] where it is defined to be $A^{-1} r_0 e^{-5\epsilon n}$. From the definition of A given above, this is bounded below by $C_4 e^{-2C} e^{-5\epsilon n}$. We are done. □

We then estimate the probability that a stable manifold is (C, λ, ϵ)-tempered.

Proposition 5.3. *Suppose that $(f_1, \ldots, f_m) \in \mathrm{Diff}_{\mathrm{vol}}(M)$ is a uniformly expanding on average tuple, where M is a closed surface. Then there exists $\lambda, \epsilon, \alpha > 0$ such that for all $C > 0$*

$$\mu(\{\omega : W^s_\omega(x) \text{ is not } (C, \lambda, \epsilon)\text{-}\textit{tempered}\}) \le C^{-\alpha}.$$

Proof. As the maps $f_1, \ldots, f_m$ are uniformly $C^{1+\text{Hölder}}$ and uniformly expanding, the trajectory is $(-C, \lambda, \epsilon)$-tempered with probability $1 - De^{-\alpha C}$ by Proposition 4.8. This stable curve is at least $D_1 e^{-2C}$ long from Proposition 5.2. The contracting of the stable manifold required by Definition 5.1 then follows from a standard graph transform argument, appearing in Chapter 7 of [BP07] or [LQ95, Lem. 3.2], or from keeping track of the contraction in the graph transform arguments in §A.4. □

6. Exactness of the skew product

We now consider measure theoretic properties of the skew product $F\colon \Sigma \times M \to \Sigma \times M$. We begin with the most basic property, ergodicity, in Proposition 6.1. Then we show that this system is exact in Proposition 6.5. As exactness implies mixing, this proposition plays a key role in the proof of finite time mixing in Section 9 where it is used in the proof of fiberwise mixing in Proposition 9.1.

6.1. Ergodicity.

The ergodicity of expanding on average systems has been known since [DK07, Section 10]. We need an extension of this result. Consider the diagonal skew product

$$F_k\colon \Sigma \times M^k \to \Sigma \times M^k \quad \text{given by} \quad (\omega, x_1, \ldots, x_k) \mapsto (\sigma(\omega), f_{\omega_0}(x_1), \ldots, f_{\omega_0}(x_k)). \tag{6.1}$$

Note that F_k preserves the measure $\mu \otimes \mathrm{vol}^k$.

Proposition 6.1. *Suppose that $(f_1, \ldots, f_m)$ is an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$ for M a closed surface. Then for each $k \in \mathbb{N}$, F_k is ergodic with respect to $\mu \otimes \mathrm{vol}^k$.*

We will not include a full proof of the above proposition as the result for $F = F_1$ is explained quite clearly in [Chu20, §3.2] as well as [Liu16, Lem. 4.41]. For $k > 1$, the result can be deduced along similar lines. No higher dimensional dynamics is needed because the dynamics is a product and hence all dynamical constructs, like stable manifolds, are just products of the constructs for the system F_1.

The proof of Proposition 6.1 relies implicitly on the following lemma which will be important in Section 6.2 as well. For $x \in M$, we let $B_\delta(x)$ denote the ball of radius δ centered at x.

Lemma 6.2. *Suppose that $(f_1, \dots, f_m)$ is an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. Then there exist $0 < \delta_1 < \delta_2$ and $\lambda, \epsilon, C_0, \epsilon_0 > 0$ such that for all $x \in M$ there exist two positive measure subsets $V_1, V_2 \subseteq \Sigma$ and a pair of transverse cones $\mathcal{C}_1, \mathcal{C}_2$ defined on $B_{\delta_2}(x)$ by parallel transport of cones based at x such that the following holds. Let Λ_ω denote the set of (C_0, λ, ϵ)-tempered points in $B_{\delta_1}(x)$ under the dynamics defined by ω, and set*

$$Q^\omega(x) = \bigcup_{y \in \Lambda_\omega \cap B_{\delta_1}(x)} W^s_{\delta_2}(\omega, y).$$

Then

(1) For $i \in \{1, 2\}$, $\omega_i \in V_i$, and $y \in \Lambda_{\omega_i}$ the stable manifold $W^s_{\delta_2}(\omega, y)$ is uniformly contracting and tangent to $\mathcal{C}_i$.

(2) For $i \in \{1, 2\}$ and $\omega_i \in V_i$, the laminations by stable manifolds satisfy the usual absolute continuity properties:

(AC 1) *If $K \subseteq M$ is a Borel set, and for almost every $y \in \Lambda_{\omega_i}$ the Riemannian leaf measure of $K \cap W^s_{\delta_2}(\omega_i, y)$ is zero, then $\mathrm{vol}(Q^{\omega_i} \cap K) = 0$.*

(AC 2) *If T is a transversal to $\mathcal{C}_i$ and $K \subseteq M$ is a Borel set, and for a positive measure subset of $z \in T$, $W^s_{\delta_2}(\omega_i, z) \cap K$ has positive leaf measure, then $\mathrm{vol}(K) > 0$.*

(3) For $i \in \{1, 2\}$ and $\omega_i \in V_i$, $\mathrm{vol}(Q^{\omega_i} \cap B_{\delta_1}(x)) > .99 \,\mathrm{vol}(B_{\delta_1}(x))$.

This lemma is implicit in Chung [Chu20] and Liu [Liu16], and further can be deduced from the propositions we prove below. In particular, our Propositions 10.12 and B.13 contain the needed claims. Lemma 6.2 allows a random version of the Hopf argument where the stable manifolds for different words $\omega \in \Sigma$ play the role of the stable and unstable manifolds in the usual Hopf argument. This can be used to prove Proposition 6.1. We will not repeat this argument here as it is adequately explained in the sources mentioned.

6.2. **Strong mixing.** Here we show that for $k \geq 1$ the skew product $F_k \colon \Sigma \times M^k \to \Sigma \times M^k$ defined in (6.1) is strong mixing for the measure $\mu \otimes \mathrm{vol}^k$. We will use this property later. A good reference for many of the properties discussed in this section is [Roh67].

Definition 6.3. An endomorphism T of a Lebesgue space $(M, \mathcal{B}, \mu)$ is *exact* if $\bigcap_{n=0}^{\infty} T^{-n}\mathcal{B} = \mathcal{N}$, the trivial sub-sigma algebra of M.

An invertible map, i.e. an automorphism, T of a Lebesgue space $(M, \mathcal{B}, \mu)$, is called a K-*automorphism* if there exists a sub-sigma algebra $\mathcal{K} \subset \mathcal{B}$ such that:

(1) $\mathcal{K} \subset T\mathcal{K}$; (2) $\bigvee_{n=0}^{\infty} T^n \mathcal{K} = \mathcal{B}$; (3) $\bigcap_{n=0}^{\infty} T^{-n}\mathcal{K} = \{\emptyset, M\}$.

Both exact systems and K-automorphisms are strong multiple mixing [Roh64, p. 17, 27], [Roh67, 15.2]. Further, an endomorphism is exact if and only if its natural extension is a K-automorphism [Roh64, p. 27].

We now describe how one may show that an automorphism $T \colon (M, \mu) \to (M, \mu)$ is exact. The *Pinsker partition* of M is the finest measurable partition $\pi(T)$ of M that has zero entropy. This means that any other measurable partition with zero entropy is coarser, mod 0, than $\pi(T)$. It turns out that T is a K-automorphism if the Pinsker partition of T trivial, i.e. $\pi(T) = \{\emptyset, M\}$, see [Roh67, 13.1,13.10]. In fact, the conditions enumerated in the definition of K-automorphism above essentially say that the Pinsker partition is trivial.

A useful fact for studying the Pinsker partition is the following.

Lemma 6.4. *(see* [BP07, p. 288], [Roh67, 12.1]*) If a measurable partition η satisfies $T\eta \geq \eta$ and $\bigvee_{n=0}^{\infty} T^n\eta = \epsilon$, the partition into points, then $\bigwedge_{n=0}^{\infty} T^{-n}\eta \geq \pi(T)$.*

Here we use the standard notation for partitions where we write $\mathcal{A} \leq \mathcal{B}$ if $\mathcal{A}$ is coarser than $\mathcal{B}$. An example of a partition satisfying the hypotheses of Lemma 6.4 is the partition of a shift space Σ into local stable sets, $W^s_{loc}(\omega) = \{\eta : \omega_i = \eta_i \text{ for } i \geq 0\}$.

We now show for $k \geq 1$ that the map F_k defined above is mixing.

Proposition 6.5. *Let $(f_1, \ldots, f_m)$ be an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$ for M a closed surface. Then the associated skew product $F\colon \Sigma \times M \to \Sigma \times M$ is exact, and hence strong mixing of all orders, for the measure $\mu \otimes \mathrm{vol}$. The same holds for $F_k\colon \Sigma \times M^k \to \Sigma \times M^k$.*

Proof. To show exactness and hence strong mixing of F, we will show that the natural extension of the skew product $F\colon \Sigma \times M \to \Sigma \times M$ has the K-property. As before, we denote by $\hat{\Sigma}$ the two sided shift, so that the natural extension of F is $\hat{F}\colon (\hat{\Sigma} \times M, \hat{\mu} \otimes \mathrm{vol}) \to (\hat{\Sigma}, \hat{\mu} \otimes \mathrm{vol})$, where $\hat{\mu}$ is the Bernoulli measure on $\hat{\Sigma}$. Note that the measure on the natural extension has this simple description because each f_i preserves volume.

We begin by showing that modulo 0, any element of the Pinsker partition is of the form $\hat{\Sigma} \times U$ where $U \subseteq M$. The local stable sets of the words $\omega \in \hat{\Sigma}$, form a measurable partition of $\hat{\Sigma}$ indexed by the elements of Σ. Further, the sets $\{W^s_{loc}(\omega) \times \{x\}\}_{x \in M}$ form a measurable partition of $\hat{\Sigma} \times M$. If we let η denote this partition, then $\bigwedge_{n=0}^{\infty} F^{-n}\eta$ is the partition into sets of the form $\hat{\Sigma} \times \{x\}$, where $x \in M$. By Lemma 6.4, we see that $\pi(\hat{F}) \leq \{\hat{\Sigma} \times \{x\} : x \in M\}$. Note that this shows that the atoms of the Pinsker partition of $\hat{F}$ are of the form $\Sigma \times A$ where A are the atoms of a partition of M. We denote this partition by $\mathcal{P}$ and the atom containing a point $x \in M$ by $\mathcal{P}(x)$.

We now show that the Pinsker partition is even coarser by using the dynamics in the fiber; in fact our goal is to show that $\pi(\hat{F})$ has an atom with positive mass. From Liu and Qian, there is a measurable partition of $\hat{\Sigma} \times M$ subordinate to the partition into full stable leaves [LQ95, Proposition VI.5.2] where each atom is a non-trivial curve in a stable leaf. This shows that for almost every $x \in M$ and almost every ω, that Lebesgue almost every $y \in W^s(\omega, x)$ is in $\mathcal{P}(x)$. (This uses AC1 for the stable lamination.) Let G^{ω_i} be the subset of Q^{ω_i} of points y such that $W^s_{\delta_2}(\omega_i, y)$ satisfies that almost every $z \in W^s_{\delta_2}(\omega_i, y)$ is in $\mathcal{P}(y)$. Note that there there is a subset $\bar{V}_i$ of full measure in V_i such that for $\omega_i \in \bar{V}_i$, G^{ω_i} has full measure in Q^{ω_i}. Now for $\omega_2 \in \bar{V}_2$ and $z \in G^{\omega_2}$, consider the intersection of a leaf $W^s_{\delta_2}(\omega_2, z)$ with G^{ω_1}, where $\omega_1 \in \bar{V}_1$. Suppose that for some such z the set $G^{\omega_1} \cap W^s_{\delta_2}(\omega_2, z)$ has positive measure. Then by definition of G^{ω_1}, almost every $y \in G^{\omega_1}$ has $W^s_{\delta_2}(\omega_1, y)$ saturated with points in $\mathcal{P}(z)$, and hence by AC2, $\mathcal{P}(z)$ has positive measure. Thus the Pinsker partition has a positive measure atom. If there were no such point z, then for almost every $z \in G^{\omega_2}$, the intersection $G^{\omega_1} \cap W^s(z, \omega_2)$ has zero leaf measure. Thus by AC1, $Q^{\omega_2} \cap Q^{\omega_1} \cap B_{\delta_1}(x)$ has measure zero. But as Q^{ω_1} and Q^{ω_2} each take up .99 proportion of the volume of $B_{\delta_1}(x)$, this is impossible. Thus we see that there is a positive volume atom of $\mathcal{P}$. Let $\Sigma \times A$ be this atom of $\pi(\hat{F})$ of positive measure.

As $\hat{F}$ is ergodic, it must cyclically permute a finite number of these positive measure sets. Because $\hat{F}$ is expanding on average, every power of $\hat{F}$ is also expanding on average. Hence, by Proposition 6.1, every power of F is ergodic. Thus the Pinsker partition has only a single non-trivial element, hence $\pi(\hat{F})$ is trivial. Hence $\hat{F}$ is a K-automorphism and so F is exact.

For the higher "diagonal" skew products F_k, the proof proceeds along very similar lines. As before, one has stable and unstable manifolds in each of the factors of M^k and hence through any particular point $(x_1, \ldots, x_k) \in M^k$, one has the stable/unstable manifold that is the product of the stable manifolds $W^{s/u}_{loc}(\omega, x_i)$. Hence in the extended system the stable an unstable foliations are transverse as before. By using these, one can similarly deduce that the Pinsker partition is finite. Further, from Proposition 6.1 every power of F_k is ergodic, which, as before implies that the Pinsker partition is trivial and thus the K-property holds for $\hat{F}_k \colon \hat{\Sigma} \times M^k \to \hat{\Sigma} \times M^k$. □

7. Coupling

In this section we present our main technical tool: the coupling lemma. We divide its proof into several steps according to the plan from Section 3. Accordingly, this section contains the outline of the rest of the paper.

7.1. **Standard pairs and standard families.** The proof of exponential mixing in this paper proceeds by showing that if μ_1 and μ_2 are two measures with smooth densities and ψ is a Hölder function then $\mu_1(\psi \circ f^n_\omega x) - \mu_2(\psi \circ f^n_\omega x)$ is exponentially small. Taking μ_2 to be vol and μ_1 to be the measure with density ϕ we obtain Theorem 1.1. Unfortunately, the set of measures whose densities satisfy a certain bound on their Hölder norm is not invariant by the dynamics, since compositions worsen Hölder regularity. So we need to consider a larger class of measures: the measures that are convex combinations of measures on (unstable) curves. This leads to notions of standard pairs and standard families that we now recall. We refer to [CM06, Chapter 7] for a detailed discussion of these notions.

Definition 7.1. A *standard pair* in a Riemannian manifold M is an arclength parametrized C^2 curve $\gamma \colon [a, b] \to M$ of bounded length along with a log-Hölder density ρ defined along γ (or equivalently $[a, b]$). We denote the pair of the curve and density by $\hat{\gamma}$ for emphasis.

There are two different ways of thinking about standard pairs. The first is that a standard pair is literally a pair of a curve and a density as in Definition 7.1. The second way is that we think of $\hat{\gamma} = (\gamma, \rho)$ as a "thickened" version of the underlying curve γ where the "thickness" is given at a point x by $\rho(x)$. More precisely, we may think of $\hat{\gamma}$ as a subset of $[a, b] \times [0, \max \rho]$ comprising the points (c, y) where $y \le \rho(c)$. We will often write $x \in \hat{\gamma}$ when referring to a point in this set associated to $\hat{\gamma}$. By thinking of the standard pair in this manner, we can imagine geometrically subdividing the pair into pieces. This type of subdivision is frequently used below.

Each standard pair defines a measure on M given for continuous $\psi \colon M \to \mathbb{R}$ by the formula

$$\hat{\rho}_\gamma(\psi) = \int_\gamma \psi(x)\rho(x)dx \tag{7.1}$$

where dx denotes the arclength parametrization of γ.

A standard curve comes with a notion of regularity. The regularity of $\hat{\gamma}$ is determined by the C^2 norm of γ as well as the C^2 norm of the density along γ. We recall now some notions from §2.4. Recall that we define the C^2 norm, $\|\gamma\|_{C^2}$, of the curve γ as the supremum of its second derivative as a graph over its tangent space in exponential charts.

Definition 7.2. Suppose that $\hat{\gamma}$ is a C^2 standard pair consisting of a curve γ and a density ρ. We say that $\hat{\gamma}$ is *R-good* if

(1) The length of γ is at least e^{-R}.

(2) The C^2 norm of γ is at most e^R.

(3) The density of ρ satisfies $\|\ln \rho\|_{C^\alpha} \le e^R$, where we measure distance with respect to the arclength parameter of γ. Recall that C^α only means the Hölder constant of the function.

We say that a standard pair $\hat{\gamma}$ is *R-regular* when at least (2) and (3) are satisfied.

Note that a larger R corresponds to a less regular curve.

Definition 7.3. For a standard pair $\hat{\gamma} = (\gamma, \rho)$, we say that $x \in \gamma$ has an *R-good neighborhood*, if there is a subcurve $\gamma' \subseteq \gamma$ containing x such that $(\gamma', \rho|_{\gamma'})$ is R-good.

Note that if x is in an R-good neighborhood of $\hat{\gamma}$, this does not imply that x is centered in long neighborhood. The point x might still be quite close to the edge. Later we will also deal with points x that are *centered* in an R-good neighborhood, meaning that the segments on either side of x form R-good neighborhoods.

Definition 7.4. A standard family is a collection of standard pairs $\{\hat{\gamma}_\theta\}_{\theta\in\Lambda}$ indexed by points from a probability space (Λ, λ).

Thus in the case that λ is atomic we just have a finite collection of standard pairs (counted with weights).

We say that a standard family is *R-good* if each standard pair that comprises it is R-good. We will only consider standard families where the goodness is bounded below.

Given a standard family $\{\gamma_\theta\}_{\theta\in\Lambda}$ we can associate a measure by integrating the measures corresponding to individual standard pairs with respect to the factor measure λ. For a function $\psi\colon M \to \mathbb{R}$, we set

$$\hat{\rho}_\Lambda(\psi) = \int_\Lambda \hat{\rho}_{\gamma_\theta}(\psi) d\lambda(\theta) \tag{7.2}$$

where $\hat{\rho}_{\gamma_\theta}$ is defined by (7.1).

A particularly useful property of standard families is that they can represent volume. It is straightforward to check that a standard pair representing volume exists by using charts.

Proposition 7.5. *Given a closed smooth manifold M endowed with a volume, there exists some $C > 0$ and a C-good standard family P_{vol} such that the associated measure represents volume on M, i.e. for any continuous function*

$$\int \phi \, dP_{\text{vol}} = \int \phi \, d \operatorname{vol}.$$

Below we will use a naïve estimate saying that the goodness of a standard pair can deteriorate at most exponentially quickly.

Proposition 7.6. *Suppose that $(f_1, \dots, f_m)$ are C^2 diffeomorphisms of a closed manifold. Then there exists $C, \eta > 0$ such that for any standard pair $\hat{\gamma}$ that is R-good and any $\omega \in \Sigma$, $f^n_\omega(\hat{\gamma})$ is $\max\{C + R + n\eta, C + n\eta\}$-good.*

Proof. The condition that the length of the curve can shrink at most exponentially fast is clear from the uniform bound on the derivative. The fact about the C^2 norm of curve follows immediately from Lemma A.9. This leaves the estimate on the density, which follows from Lemma A.7 because the C^2 norm of f^n_ω grows at most exponentially. □

Note that the representation (7.2) (including the representation of the volume from Proposition 7.5) is highly non-unique. One type of non-uniqueness that we shall often exploit in our proof is the possibility to divide a standard pair into pieces. To do so we partition the underlying curve γ into multiple disjoint subcurves $\gamma_1, \dots, \gamma_n$. We then obtain a subdivision of (γ, ρ) from the restrictions $(\gamma_1, \rho|_{\gamma_1}), \dots, (\gamma_n, \rho|_{\gamma_n})$. We give each piece unit mass for the indexing measure λ. Note that (γ, ρ) as well as the standard family $\{(\gamma_i, \rho|_{\gamma_i})\}_{1\le i\le n}$ both represent the same measure on M.

A more subtle type of subdivision occurs when we view a standard pair as a subset of $\gamma\times[0,\max\rho]$ and partition this subset in the vertical direction. Similarly, we will obtain a new standard family. But now the underlying curves of the family may not be disjoint. For a simple example, something we do multiple places in the local coupling argument is take a standard pair (γ,ρ), a number $\alpha\in(0,1)$, and subdivide this standard pair into $\{(\gamma,\alpha\rho),(\gamma,(1-\alpha)\rho)\}$ and give each piece mass 1 for the indexing measure λ. Alternatively, we could take $\hat{\gamma}_1=\hat{\gamma}_2=(\gamma,\rho)$ and allow the indexing measure assign them mass α and $1-\alpha$, which gives the same measure on M independent of α. Below, we will often think of this geometrically: we take the region associated to the standard pair in $\gamma\times[0,\max\rho)$ and slice it into regions. Projecting the Lebesgue measure on each region down to γ naturally defines a standard pair.

Next, if we have a standard family $\hat{\gamma}$ and a subfamily $\hat{\gamma}'$ of $\hat{\gamma}$ defined by some subdivision of $\gamma\times[0,\max\rho)$ as mentioned above, then we define $\hat{\gamma}\setminus\hat{\gamma}'$ to be the standard family defined by the complement of $\hat{\gamma}'$ in the subdivision.

7.2. **Main coupling proposition.** We now state the main technical result of the paper, from which the main mixing results of this paper are a consequence.

Proposition 7.7. *Suppose that $(f_1,\dots,f_m)$ is an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$, where M is a closed surface. There exists $\lambda>0$ such that for all sufficiently small $\epsilon>0$, there exist $C,\alpha>0$, such that for any R, a goodness of standard pairs, the following holds.*

Let $\hat{\gamma}_1$ and $\hat{\gamma}_2$ be two standard pairs with associated measures ρ_1 and ρ_2 of equal mass that are R-good. Then we have the measures $\mu\otimes\rho_i$ on $\Sigma\times\hat{\gamma}_i$, where μ is the Bernoulli measure on the one sided shift. There exists a coupling function $\Upsilon\colon\Sigma\times\hat{\gamma}_1\to\hat{\gamma}_2$, where for each ω the map $\Upsilon(\omega,\cdot)\colon\hat{\gamma}_1\to\hat{\gamma}_2$ is measure preserving, and a time $\hat{T}(\omega,x)$ such that

$$f_\omega^{\hat{T}(\omega,x)}(x)\in W^s_{\sigma^{\hat{T}(\omega,x)}\omega,C^{-1}}(f_\omega^{\hat{T}(\omega,x)}\Upsilon(\omega,x)),$$

and this stable manifold is uniformly (C,λ,ϵ)-tempered in the sense of Definition 5.1. Further

$$\mathbb{P}_{\omega,x}(\hat{T}(\omega,x)\geq n)\leq e^{\max\{R,0\}}e^{-\alpha n}.$$

The proof of this proposition is a combination of a local coupling lemma (Lemma 7.10) along with a recovery procedure.

When we attempt to couple two curves, we will insist that they are in a configuration that allows us to try and apply the Local Coupling Lemma (Lemma 7.10). What we mean by this is that the curves have controlled regularity and are sufficiently near to each other.

Definition 7.8. Let $\hat{\gamma}$ be a standard pair and $x\in\gamma$. We say that x is (C,δ)-*well positioned in* $\hat{\gamma}$ if $\hat{\gamma}$ is C-regular and x is δ distance away from the endpoints of γ, with distance measured along γ.

We say that two standard pairs $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are in a (C,δ,υ)-*configuration* if there exist x which is (C,δ)-well positioned in $\hat{\gamma}_1$, and y which is (C,δ)-well positioned in $\hat{\gamma}_2$ such that $d(x,y)<\upsilon$.

The proof of Proposition 7.7 proceeds along the following steps. We start with two C_0-good standard pairs, $\hat{\gamma}_1$ and $\hat{\gamma}_2$. Here C_0 is some uniform regularity appearing in Proposition 7.9 that we may obtain starting from an arbitrarily bad curve by waiting long enough.

(1) We prove that for a large proportion of words $\omega\in\Sigma$, the images $f^n_\omega(\hat{\gamma}_1)$ and $f^n_\omega(\hat{\gamma}_2)$ are mostly quite regular, and moreover, there is a large measure subset of the images that can be paired to form (C_1,δ,υ)-configurations for some C_1 that is worse that C_0. This relies on the mixing properties of our system studied in Section 6, and the needed conclusions are made precise in Proposition 7.11.

(2) We then run a "local" coupling argument on each tiny (C_1, δ, υ)-configuration. At each time step, we attempt to couple the remaining well tempered points using "fake" stable manifolds. This local coupling argument, Lemma 7.10, has a number of steps and draws on several intermediate estimates.

(a) There are $C, \lambda, \epsilon > 0$ and a cone field $\mathcal{C}_\theta$ that is uniformly transverse to both γ_1 and γ_2 such that the probability that any point is (C, λ, ϵ)-tempered and has E^s tangent to $\mathcal{C}_\theta$ is positive. Further, the probability that the tempering fails at time n is exponentially small.

(b) For a (C, λ, ϵ)-tempered point at time n, we see that there is a "fake" stable manifold W^s_n given by taking a curve nearly tangent to $Df^n_\omega(E^s_n)$ and pushing this curve backwards by $(Df^n_\omega)^{-1}$. (This construction is the subject of §B.4)

(c) There exist worse $(C', \lambda', \epsilon')$ such that for every (C, λ, ϵ)-tempered point x in γ_1, all points within distance$\|D_x f^n_\omega\|^{-(1+\sigma)}$ of are $(C', \lambda', \epsilon')$-tempered points at time n. (This is the content of Proposition 10.3). These $(C', \lambda', \epsilon')$-tempered points also have fake stable manifolds. We will try to couple these thickened neighborhoods of the (C, λ, ϵ)-tempered points with some neighborhoods in γ_2 determined by the fake stable holonomies. At the time when $D_x f^n_\omega$ fails to be (C, λ, ϵ)-tempered with E^s tangent to $\mathcal{C}_\theta$ we discard the point x and stop trying to couple it.

(d) For $(C', \lambda', \epsilon')$-tempered points, the holonomies of the fake stable manifolds W^s_n between γ_1 and γ_2 converge exponentially fast to the true, limiting stable holonomy. Moreover, the image of a point $x \in \gamma_1$ under H^s_n has fluctuations, as n changes, of size $\|D_x f^n_\omega\|^{-1.99}$, i.e. the distance between $H^s_n(x)$ and $H^s_{n+1}(x)$ in γ_2 is at most $\|D_x f^n_\omega\|^{-1.99}$. (This is proved in Proposition B.12.)

(e) The points we try to couple with on γ_2 are the image of the points on γ_1 under the fake stable holonomy H^s_n.

(f) By carefully choosing subdivisions of the standard pairs $\hat{\gamma}_1$ and $\hat{\gamma}_2$ we may discard mass from the standard pairs so that at the end of the procedure a positive proportion of the mass above each (C, λ, ϵ)-tempered point remains. The control on the size of the fluctuations of H^s_n relative to the lengths of the intervals of $(C', \lambda', \epsilon')$-tempered points containing the (C, λ, ϵ)-tempered points $\|D_x f^n_\omega\|^{-1.99} \ll \|D_x f^n_\omega\|^{-(1+\sigma)}$ allows us to ensure that we always have enough points on γ_2 to try to couple with.

(3) We prove that we may find simultaneous recovery times for a pair of R-good standard pairs (Proposition 7.9), so that if we have failed to couple and are left with a short standard subcurve of $\hat{\gamma}_1$ we can have this subcurve recover at the same time as a subcurve of $\hat{\gamma}_2$.

(4) Once we recover we will try to couple again using steps (1)–(3) above. Each time we try to couple, a positive amount of mass couples, and as the tail on the recovery time is exponential we do not spend too much time recovering.

7.3. **Statements of the lemmas for use during coupling.** We now state the main propositions and lemmas that are used in the proof of Proposition 7.7.

Lemma 7.9. *(Coupled Recovery Lemma) Let M be a closed surface and let $(f_1, \ldots, f_m)$ be an expanding on average tuple with entries in* $\mathrm{Diff}^2_{\mathrm{vol}}(M)$*. There exist $C_0, D_1, \alpha > 0$ such that if $\hat{\gamma}_1 = (\gamma_1, \rho_1)$ and $\hat{\gamma}_2 = (\gamma_2, \rho_2)$ are R-good standard families of equal mass then there is a pair of stopping times $\hat{T}_1$ and $\hat{T}_2$ defined on $\hat{\gamma}_1$ and $\hat{\gamma}_2$ with the following properties:*

(1) There is an exponential tail on the stopping time. Namely,

$$(\mu \otimes \rho_1)((\omega, x) \mid \hat{T}_1(\omega, x) > n) \leq D_1 e^{\max\{R,0\} - \alpha n}.$$

(2) If $z \in \hat{\gamma}_i$ is a point that stops at time n, and $B_i(z)$ is the connected component of z in the set $\{x \in \hat{\gamma}_i : \hat{T}_i(\omega, x) = n\}$, i.e the set of points $z \in \hat{\gamma}_i$ stopped at time n, then $f_\omega^{\hat{T}_i(z)}(B_i(z))$ is a C_0-good standard pair.

(3) For each $\omega \in \Sigma$, we always stop on the same amount of mass of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ at each time n. Specifically, for each ω and n, denote $S_i(\omega, n) = \{x \in \hat{\gamma}_i : \hat{T}_i(\omega, x) = n\}$. For each pair (ω, n) there is a measure preserving map $\Phi_n^\omega \colon S_1(\omega, n) \to S_2(\omega, n)$ carrying C_0-good connected components of $S_1(\omega, n)$ to C_0-good connected components of $S_2(\omega, n)$.

The following lemma is the most technical part of the coupling argument.

Lemma 7.10. *(Local Coupling Lemma) Suppose that $(f_1, \dots, f_m)$ is an expanding on average tuple. There exists $0 < \tau < 1$ such that for any $C_1 > 0$ there exists $\delta_0, L, D_1, D_2, \beta, C, \lambda, \epsilon > 0$ such that for any $0 < \delta' < \delta_0$ there exists δ_1 and $\epsilon_0, a_0 > 0$ such that for any two standard pairs $\hat{\gamma}_1$ and $\hat{\gamma}_2$ that are in a (C_1, δ', υ)-configuration with $\upsilon \le \tau\delta'$, we may couple a uniform proportion of the points on the two curves with an exponential tail on the points that do not couple.*

Specifically, for two C_1-good standard pairs $\hat{\gamma}_1, \hat{\gamma}_2$ of the same mass in a (C_1, δ', υ)-configuration with $\upsilon \le \tau\delta'$, there is a point $x \in M$, a ball $B_{\delta_0}(x) \subset M$ and connected components Γ_1 and Γ_2 of $\hat{\gamma}_1 \cap B_{\delta_1}(x)$ and $\hat{\gamma}_2 \cap B_{\delta_1}(x)$ such that Γ_1 and Γ_2 each contain a_0 proportion of the mass of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ respectively.

Further, there exist a pair of stopping times $\hat{T}_1(\omega, x)$ and $\hat{T}_2(\omega, x)$ defined on $\hat{\gamma}_1$ and $\hat{\gamma}_2$ such that if $B^{\hat{T}_i}(\omega, x) \subseteq \hat{\gamma}_i$ denotes the block of points stopped at the same time as x, then

(1) For all ω, n there exists $\Psi_n^\omega \colon \{x \in \hat{\gamma}_1 \colon \hat{T}_1(\omega, x) = n\} \to \{x \in \hat{\gamma}_2 : \hat{T}_2(\omega, x) = n\}$ such that if $\hat{T}_i(\omega, x) = n$, then $B(\omega, x)$ is an nL-good standard pair and Φ_n^ω carries $B(x)$ to an nL-good standard pair $B(\Phi_n^\omega(x)) \subseteq \hat{\gamma}_2$ of equal mass that is also stopped at time n.

(2) For each ω, the set of points in $\hat{\gamma}_1$ and $\hat{\gamma}_2$ where $\hat{T}_i = \infty$ are of equal measure and moreover these sets are intertwined by a measure preserving stable holonomy along uniformly (C, λ, ϵ)-tempered stable manifolds.

(3) There exists $D_1 > 0$ such that $(\mu \otimes \hat{\rho}^1)(\{(\omega, \hat{x}) : \hat{T}_1(\omega, \hat{x}) = n\}) \le D_1 e^{-\beta n}$. For $\hat{\gamma}_2$, we have a similar estimate, $(\mu \otimes \hat{\rho}^2)(\{(\omega, \hat{x}) : \hat{T}_2(\hat{x}) = n\}) \le D_1 e^{-\beta n}$.

(4) For all $x \in \Gamma_1$, the measure of words ω such that $\hat{T}_i(\omega, x) = \infty$ is at least ϵ_0.

In the lemma above, part (2) says that the points where $\hat{T}_i = \infty$ are coupled and such points attract exponentially fast. Part (4) says that the probability that the next coupling attempt is successful is at least ϵ_0. Part (3) says that the probability that "a point" stops and fails to couple at time n is exponentially small, while part (1) controls he regularity of the set of such points.

The following proposition says that there is a fixed time N_0 required for the C_0-good pairs produced by the coupled recovery lemma to get into position for the application of the local coupling lemma. The proof relies on the mixing properties from Section 6.

Proposition 7.11. *(Finite Time Mixing) Suppose $(f_1, \dots, f_m)$ is an expanding on average tuple as in Proposition 7.7. For any fixed $C_0 > 0$, there exist $C_1, C_2, \delta, \upsilon > 0$ such that the following holds.*

(1) $C_1, \delta, \upsilon > 0$ are such that a (C_1, δ, υ)-configuration satisfies the hypotheses of the Local Coupling Lemma 7.10 with $C_1 = C_1$, $\delta' = \delta$, and $\upsilon = \upsilon$.

(2) There exists $N_0 \in \mathbb{N}$ and $b_0 > 0$ such that for any C_0 regular standard pairs $\hat{\gamma}_1$ and $\hat{\gamma}_2$ of equal mass, for .99% of the words $\omega \in \{1, \dots, m\}^{N_0}$, there is a subdivision P_ω^1, P_ω^2 of the

standard families $f_\omega^{N_0}(\hat\gamma_1)$ *and* $f_\omega^{N_0}(\hat\gamma_2)$ *and subfamilies* Q_ω^1, Q_ω^2 *of* P_ω^1 *and* P_ω^2*, and a map* $\Psi\colon Q_\omega^1 \to Q_\omega^2$ *preserving measure such that the following hold.*

(a) *Each pair* $\hat\gamma \in Q_\omega^1$ *is associated by* Ψ *with a pair* $\Psi(\hat\gamma)$ *such that these pairs have equal mass and satisfy (1) above.*

(b) *The set* $Q^1 = \bigcup_{\omega\in\hat\Sigma}\{\sigma^{N_0}(\omega)\}\times Q_\omega^1$ *has measure* $b_0\rho_1(\hat\gamma)$ *with respect to* $\hat\mu\otimes\rho_1$*. The same holds for* Q^2*.*

(3) *The complement of* Q_ω^1 *in* $f_\omega^n(\hat\gamma_1)$ *is a standard family of* C_2*-good standard pairs. The same holds for* Q_ω^2*.*

As mentioned before, the proofs of these lemmas appear later in the paper. Lemma 7.9 is proven in Section 8, Proposition 7.11 is proven in Section 9, and Lemma 7.10 is proven in Section 10.

7.4. **Proof of the main coupling proposition.** We now show how to deduce the main coupling proposition, Proposition 7.7, from the various results stated in this section. We need a preliminary estimate showing that if we fail to couple then the whole failed attempt does not take too long. In the lemma below the recovery time is the sum of three terms:

(1) The time when we stop trying to locally couple as in Lemma 7.10 item (3);

(2) The time it takes for a point to recover so that it belongs to a C_0-good pair as in the Coupled Recovery Lemma 7.9;

(3) The fixed time N_0 where the point has a chance to enter a (C_1, δ, υ)-configuration according to Proposition 7.11.

The following lemma verifies that each trip through the coupling procedure has an exponential tail on its duration.

Lemma 7.12. *In the setting of Proposition 7.7, for each* C *there exist* $\hat C$ *and* $\bar r$ *such that if* $\hat\gamma_1$ *and* $\hat\gamma_2$ *are* C*-good standard pairs of equal mass, then*

$$(\mu\otimes\rho_1)((\omega,\hat x) : (\omega,\hat x) \text{ fails to couple and the recovery time is greater than } n) \le \hat C e^{-\bar r n}.$$

Proof. Take a small $\kappa > 0$ that will be specified below. First we try to locally couple, and then we recover. Let T be the recovery time and S be the time when we stop our attempt at coupling (ω, x). Then if $T \ge n$ then either:

(i) $S \ge \kappa n$ or (ii) $S \le \kappa n$ and the time it takes the corresponding part of the curve to recover is at least $(1-\kappa)n$.

The probability of the first event is exponentially small due to Proposition 7.10(3). In the second case since $S \le \kappa n$, it follows that (ω, x) belongs to κLn-good component. Thus by Proposition 7.9 the probability that the recovery takes more than $(1-\kappa)n$ time is less than $D_1 e^{(\kappa L - \alpha(1-\kappa))n}$ which is exponentially small if $\kappa < \alpha/(L+\alpha)$. □

The main coupling proposition is now easy to deduce because each coupling attempt couples a positive proportion of the remaining mass and, from Lemma 7.12, there is an exponential tail bound on how long a coupling attempt takes.

Proof of Proposition 7.7. Let $N(\omega, x) + 1$ be the number of total attempts at local coupling before (ω, x) couples. Let $\hat T(\omega, x)$ be the time when (ω, x) couples, and let $T_k(\omega, x)$ be its kth recovery time, i.e. the $k+1$st time we attempt to locally couple. As a positive amount of mass couples each time we apply the local coupling lemma, we see that there exists $\delta > 0$ such that

$$(\mu\otimes\rho_1)((\omega, x) : N(\omega, x) > k) \le e^{-k\delta}. \tag{7.3}$$

Next we show that for points that take k-attempts at local coupling to couple, that these attempts occur linearly fast. This will follow once we have a tail bound on T_k. By Lemma 7.12, T_1 has an exponential moment. In particular, $\sup \mathbb{E}\left[e^{tT_1}\right] = M(t)$ is finite for $t \le r$ where $r < \bar{r}$ and $\bar{r}$ is the constant from Lemma 7.12 and the supremum is taken over all pairs $\hat{\gamma}_1, \hat{\gamma}_2$ of C_1-good standard pairs which are in (C_1, δ, υ)-configurations as required by Lemma 7.10 and produced by Proposition 7.11.

Extend $T_k = T_N(\omega)$ if $k > n(\omega)$. A straightforward induction shows that $\mathbb{E}\left[e^{tT_k}\right] \le M(t)^k$. Thus by the Chernoff bound $(\mu \otimes \rho_1)(T_k \ge n) \le M(t)^k e^{-tn}$. In particular taking $t = r$, there is some $\beta > 0$ such that $(\mu \otimes \rho_1)(T_k \ge n | N = k) \le e^{\beta k} e^{-rn}$. Fix some small number α such that $0 < \beta\alpha < r/2$. Then

$$(\mu \otimes \rho_1)(T_N > n \text{ and } N \le \alpha n) \le (\mu \otimes \rho_1)(T_{\alpha n} > n) \le D_1 e^{-r/2n}.$$

By (7.3), with probability $1 - e^{-\delta\alpha n}$, a point (ω, x) couples after at most αn trials, and the result follows. □

8. Proof of the Coupled Recovery Lemma

8.1. **Recovery times.** In this subsection, we use the preceding lemmas to describe a recovery algorithm for the C^2 norm of an irregular curve and estimate the tail of the recovery time.

The next definition describes an iterate of f^n_ω that has a good enough splitting that $f^n_\omega(\gamma)$ will have a good neighborhood of a particular point. Note that a "good enough" splitting requires both a condition on the hyperbolicity as well as a condition on the angle between the curve γ and and the stable subspace. This definition will be used in the proof of the recovery lemma.

Definition 8.1. Fix a tuple of non-negative numbers $(C, \lambda, \epsilon, A, \epsilon', R)$. For a standard pair $\hat{\gamma}$, a point $x \in \gamma$ and a word $\omega \in \Sigma$, we say that n is a $(C, \lambda, \epsilon, A, \epsilon', R)$-*backwards good time* for x, γ, ω if $n = A \max\{R, 1\} + i$, for some $i \ge 0$ and

1. Df^n_ω has a (C, λ, ϵ)-reverse tempered splitting, for which we write E^s_m, E^u_m for the stable and unstable subspaces of this splitting in $T_{f^m_\omega(x)}M$.
2. $\angle(E^s_0, \dot{\gamma}(x)) \ge e^{-\epsilon' i}$.

The following lemma asserts that this type of backwards good time is sufficient to conclude that an R-good curve γ has its neighborhood of x smoothed by the random dynamics f^n_ω.

Note that the second condition in the lemma considers the situation where γ "recovers" in a neighborhood of x prior to time n. It is important in this case to know that from that point on, we can just restrict to the portion of the curve that has already recovered. This is useful because it helps us deal with situations where we wish to "stop" on certain parts of the curve and know that the parts we have stopped on will not be needed later when a different part of the curve recovers. Recall from Definition 7.2 that an R-regular curve has all the characteristics of R-good curves except that it is not required to be e^{-R} long.

Lemma 8.2. *Suppose M is a closed surface and that $(f_1, \ldots, f_m)$ is a tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. Then for any $\lambda > 0$, sufficiently small $\epsilon, \epsilon' > 0$, and any $C > 0$, there exists $A, C_0, C_1 > 0$ such that for any R-regular standard pair $\hat{\gamma} = (\gamma, \rho)$ and any $(C, \lambda, \epsilon, A, \epsilon', R)$-backwards good time n for $\omega \in \Sigma$ and $x \in \gamma$ if:*

1. *$\hat{\gamma}$ is R-good, or*
2. *there exists a time $0 \le m < n$ and a subinterval $I \subseteq \gamma$ such that $f^m_\omega(I)$ contains a neighborhood of $f^m_\omega(x)$ that is $e^{-C_1} e^{-.8\lambda(n-m)}$-long;*

then $f^n_\omega(\hat\gamma)$ contains a C_0-good neighborhood of $f^n_\omega(x)$. Moreover, if (2) holds, this neighborhood is contained in $f^n_\omega(I)$.

The above lemma follows immediately from the result below. The second paragraph of the statement of the lemma essentially says: if there is another point in γ that also experiences a recovery time, then we can stop on that recovering segment while still leaving enough of the curve γ so that x can still recover.

Lemma 8.3. *(Deterministic Recovery Lemma) Given a closed surface M and a tuple $(f_1,\dots,f_m)$ in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$, for any $\alpha,\lambda>0$ and all sufficiently small $\epsilon,\epsilon'>0$ and any $C>0$, there exist $C_0, A>0$ such that for any R-good standard pair $\hat\gamma=(\gamma,\rho)$, and any word ω such that time n is a $(C,\lambda,\epsilon,A,\epsilon',R)$-backwards good time for $x\in\gamma$, then there exists a neighborhood $B(x)\subseteq\gamma$ of size at most $e^{-.9\lambda n}$ such that $f^n_\omega(\hat B(x))$ is C_0-good, i.e. the pushforward of the standard pair $\hat\gamma$ restricted to $B(x)$ is C_0-good.*

Further, there exists C_1 such that for ω,x,γ as in the first part of the lemma, if $I\subseteq\gamma$ is an interval containing x and for some $1\le m<n$, $f^m_\omega(I)$ has length at least $e^{-C_1}e^{-.8\lambda(n-i)}$, then $f^n_\omega(I)$ contains a C_0-good neighborhood of $f^n_\omega(x)$.

Proof. We divide the proof into several steps. We begin by fixing some preliminaries. For the given (C,λ,ϵ), we apply Proposition A.13 with $e^{-i\epsilon'}=\theta$, which gives us the constants $\epsilon_0,\ell_{\max},D_2,\dots,D_8$ appearing in that proposition.
Step 1. (Length of $f^n_\omega\gamma$) By Proposition A.13(2), if

$$n\ge D_5+\frac{\max\{R,0\}-2\ln(e^{-i\epsilon'})}{.99\lambda}, \tag{8.1}$$

then $f^n_\omega\gamma$ contains a neighborhood γ_n of $f^n_\omega(x)$ of length $\ell_{\max}$. For ϵ' sufficiently small relative to λ, it follows that (8.1) holds as long as $n\ge A_1\max\{R,1\}+i$ for some A_1 depending only on D_5,λ,ϵ'.
Step 2. (C^2 estimate) By Proposition A.13(3)

$$\|\gamma_n\|_{C^2}<D_6e^{-2.9\lambda n}e^{D_7\ln\theta}\max\{\|\gamma\|_{C^2},1\}+D_8. \tag{8.2}$$

Thus there exists A_2,C_2 such that as long as $n\ge A_2\max\{R,1\}+i$, that $\|\gamma_n\|_{C^2}\le C_2$.
Step 3. (Smoothing the density) From Proposition A.13(4) applied to $D_9=C_2$ from the previous step, we see that there exists D_{10},D_{11} such that the following holds. If $\|\gamma_n\|_2<D_8$, then the pushforward of ρ along γ_n is given by:

$$\|\ln\rho_n|_{\gamma_n}\|_{C^\alpha}\le D_{10}e^{-.9\alpha\lambda n}e^{D_7\ln\theta}(1+\|\ln\rho\|_{C^\alpha}+\|\gamma\|_{C^2})+D_{11}. \tag{8.3}$$

In particular as long as $N\ge A_2\max\{R,1\}+i$, the above estimate holds. In the case that this estimate holds, then as $\|\ln\rho\|_{C^\alpha}$ and $\|\gamma\|_{C^2}$ are both at most e^R, we similarly see that there exists C_3 and A_3 such that if $n\ge A_3\max\{R,1\}+i$ then $\|\ln\rho_n|_{\gamma_n}\|_{C^\alpha}\le C_3$. Thus we see that there exists A such that the conclusion of the first paragraph holds.

For the claim in the second paragraph of the Lemma, we can apply Proposition A.13(2). The choice of A,C_0 in the first part of the proof imply that for such n, $\ell_{\max}$ is realized and thus by the final part of item (2) then the preimage of γ_n in $f^i_\omega\gamma$ has length at most $D_4e^{-.9\lambda(n-i)}$, thus if $f^i_\omega(I)$ has length at least $D_4e^{-.8\lambda(n-i)}$, then the image of $f^i_\omega(I)$ will have image that is a C_0 good neighborhood of $f^n_\omega(x)$. □

Next we show that the recovery times from the above lemma occur frequently.

Proposition 8.4. *Let M be a closed surface and suppose that $(f_1,\dots,f_m)$ is an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. There exists $\lambda>0$ such that for any $A>0$ and sufficiently* *small $\epsilon, \epsilon' > 0$, there exist $C > 0$ and $\alpha_3 > 0$ such that for any R-good standard pair $\hat{\gamma}$, if for $x \in \gamma$ we let $\hat{T}(\omega, x)$ be the first $(C, \lambda, \epsilon, A, \epsilon', R)$-backwards good time. Then*

$$(\mu \otimes \rho)((\omega, x) : \hat{T}(\omega, x) > A \max\{R, 1\} + i) \leq Ce^{-\alpha_3 i}. \tag{8.4}$$

The same holds for the analogous stopping time defined on an R-good standard family.

Proof. It suffices to prove this estimate at a single point x as we may then integrate the resulting estimate over all of $\hat{\gamma}$. From Proposition 4.18 there exist C_1, α_1 and $C, \lambda > 0$ such that for all sufficiently small $\epsilon > 0$ there exists $N \in \mathbb{N}$ such that if we let $S(\omega)$ be the stopping time that stops at the first (C, λ, ϵ)-reverse tempered time of $D_x f^n_\omega$ greater than any fixed $n \geq N$, then at that time there is a well defined splitting $T_x M = E^s_S \oplus E^u_S$ into maximally expanded and contracted singular directions, and

$$\mathbb{P}(S(\omega) > n + k) \leq C_1 e^{-\alpha_1 k}. \tag{8.5}$$

By Lemma 4.19 there exist $C_2, \alpha_2 > 0$ such that as long as $n \geq c_0 |\ln \theta|$,

$$\mathbb{P}(\angle(E^s_S, \dot{\gamma}(x)) < \theta | S \leq n + k) < C_2 \theta^{\alpha_2}. \tag{8.6}$$

Hence there exists $\alpha_3 > 0$ such that if S is the first time greater than $n = c_0 \epsilon' i$ that has a reverse tempered splitting, then

$$\mathbb{P}(\angle(E^s_S, \dot{\gamma}(x)) < e^{-\epsilon' i} | S \leq n + k) < C_2 e^{-\alpha_2 \epsilon' i}. \tag{8.7}$$

In particular, as long as ϵ' is sufficiently small relative to c_0, then $c_0 \epsilon' i < i/2$. Let S be the first (C, λ, ϵ)-reverse tempered time greater than $A \max\{R, 1\} + i/2$. Multiplying equations (8.5) and (8.7), we find that there exist $C_3, \alpha_3 > 0$ such that:

$$\mathbb{P}(S \leq A \max\{R, 1\} + i \text{ and } \angle(E^s_S, \dot{\gamma}) \geq e^{-\epsilon' i}) \geq 1 - C_3 e^{-\alpha_3 i}. \qquad \square$$

We now state without proof a more technical variant of the preceding lemma. It will be used in the proof of the coupled recovery lemma to allow "recovery times" for the hyperbolicity. We will divide the iterates of the system into blocks of size $\Delta q + \Delta$, where $\Delta, q \in \mathbb{N}$. Each block will be divided into two pieces one of length Δq and one of length Δ. We will only be interested in backwards good tempered times that occur in the second part of the block, which has length Δ. This is to ensure that there are large (temporal) gaps between possible recovery times. The following lemma shows that given this extra restriction on the backwards good times, we still have an exponential tail.

Proposition 8.5. *Let M be a closed surface and suppose that $(f_1, \ldots, f_m)$ is an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. There exists $\lambda > 0$ such that for any $A > 0$ and sufficiently small $\epsilon, \epsilon' > 0$, there exist $C > 0$ and $\alpha_4 > 0$ such that for all $\Delta, q \in \mathbb{N}$ and any R-good standard pair $\hat{\gamma}$, for any $N \geq A \max\{R, 1\}$, if for $x \in \gamma$ we let $\hat{T}(\omega, x)$ be the first time greater than equal to N such that*

$$\lceil A \max\{R, 1\} \rceil + j(q + 1)\Delta + q\Delta < \hat{T}(\omega, x) \leq \lceil A \max\{R, 1\} \rceil + (j + 1)(q + 1)\Delta,$$

for some $j > 0$ and $\hat{T}$ is a $(C, \lambda, \epsilon, A, \epsilon', R)$ backwards good time, then

$$(\mu \otimes \rho)((\omega, x) : \hat{T}(\omega, x) > N + i(q + 1)\Delta) \leq Ce^{-\alpha_4 i\Delta}. \tag{8.8}$$

8.2. **Coupled Recovery Lemma.** In this subsection, we prove the coupled recovery lemma, Lemma 7.9. In the statement we view the standard pair as the uniform distribution on the subset of $\gamma\times[0,\infty)$ of pairs (x,t) where $t\le \rho(x)$. We do this so that we may define stopping times for $\hat{\gamma}$ that stop on only part of the fiber over each point in γ. Additionally, in an abuse of notation, we will identify the density ρ with a measure that we also call ρ.

Proof of Lemma 7.9. After initial preliminaries, the proof divides into two parts. The first part is a coupled stopping procedure, which takes a word $\omega\in\Sigma$ and two standard pairs $\hat{\gamma}_1$ and $\hat{\gamma}_2$, and shows which parts of each curve get stopped as we follow the dynamics specified by ω so that we always stop on the same amount of mass of each pair. In the second part we show that with high probability the procedure from the first part actually stops on all but an exponentially small amount of $\hat{\gamma}_1,\hat{\gamma}_2$ in a linear amount of time. In the proof, we consider the case that $R>1$ as otherwise we can stop immediately and conclude.

We now fix some constants. By Proposition 8.5 there exists $\lambda>0$ such that for any $A>0$ and sufficiently small $\epsilon,\epsilon'>0$, there exists $C>0$ and $\alpha>0$ such that $(C,\lambda,\epsilon,A,\epsilon',R)$-backwards good times at the end of blocks of length $(q+1)\Delta$ occur exponentially fast after any time N greater than $A\max\{R,1\}$ for an R-good standard pair $\hat{\gamma}$, i.e. (8.8) holds.

We then apply Lemma 8.2, which shows that for this choice of $\lambda,C,\epsilon,\epsilon',A$, that any R-good standard pair $\hat{\gamma}$ and any $(C,\lambda,\epsilon,A,\epsilon',R)$-backwards good time to $x\in\hat{\gamma}$, $f^n_\omega(x)$ has a C_0-good neighborhood in $f^n_\omega(\hat{\gamma})$, i.e. the dynamics smoothens a neighborhood of x and makes it C_0 regular. Lemma 8.2 also gives the constant C_1 so that as long as $f^i_\omega(I)$ contains a neighborhood of $f^i_\omega(x)$ of size at least $e^{-C_1}e^{-.8\lambda(n-i)}$, then $f^{n-i}_{\sigma^i(\omega)}(f^i_\omega(I))$ contains a C_0-good neighborhood of $f^n_\omega(x)$.

For the rest of the proof we will not repeat $(C,\lambda,\epsilon,A,\epsilon',R)$-backwards good but just refer to such times as *tempered times* with this particular choice of constants being understood.

In the proof that follows, we divide the iterates of the system into blocks of size $(q+1)\Delta$. We will attempt to stop on a neighborhood of a point x when $D_xf^n_\omega$ has a tempered time in the interval $(\lceil AR\rceil+i(q+1)\Delta+q\Delta,\lceil AR\rceil+(i+1)(q+1)\Delta]$. This is the ith block, if there is such a tempered time, then we say that this is a *tempered block*. In the following, there will be points x that experience a tempered block ending at $\lceil AR\rceil+iq\Delta$ but that we do not stop because there was not enough mass stopping on the other curve to couple them. For these curves, we then wait for their next tempered time relative to the original curve. That we only allow stopping on the last Δ iterates of a block of length $(q+1)\Delta$ is to ensure that the hyperbolicity has enough time to stretch what remains of the recovered neighborhood of $f^{\lceil AR\rceil+i\Delta}_\omega(\gamma)$ so that it can recover to be a C_0-good curve at the tempered time.

In the proof we only try to couple recovered curves at the very last time in each block, whereas a curve may have a tempered time up to Δ iterates before then. If we have a C_0-good curve, $\hat{\gamma}$, and we apply the dynamics from $(f_1,\dots,f_m)$ at most Δ additional times, then there is some $C_0'\ge C_0$, so that the image of the curve will still be C_0' good even after those extra iterates. Consequently, for any $\alpha>0$, there exists $\delta(\alpha)>0$, such that if $\hat{\gamma}$ is a C_0' good curve, and we trim off the end segments of the curve of length $e^{-\delta}$, then we have lost at most $e^{-\alpha}$ proportion of the curve, where α is some number we will choose below. Further, note that as long as δ is sufficiently large, the trimmed off curves will be $e^{-\delta}$-good and that when we trim a C_0'-good curve, what remains will also still be δ-good.

The proof involves four additional parameters some of which were alluded to above, and which we choose to be sufficiently large that the following hold:

(1) There is an exponential tail on the wait for the first tempered block. For any $N \geq \lceil AR \rceil$, if $T(\omega, x)$ is the next tempered block after N, then

$$\mathbb{P}_\omega(T(\omega, x) \geq N + i(q+1)\Delta) \leq e^{-i\alpha}. \tag{8.9}$$

(2) We also fix a small constant $\beta > 0$. Then by possibly increasing Δ even further we can arrange that $\beta < \alpha/7$ and in addition have that α is greater than the cutoffs in Claims 8.6 and 8.7 below.

(3) We then choose q sufficiently large that $e^{-\delta} > e^{-C_1} e^{-.8\lambda q \Delta}$, where δ is the goodness of the recovered curve from above and depends on α and Δ.

Note that when picking the constants above, from the statement of Proposition 8.5 we first choose Δ to make $e^{-\alpha}$ arbitrarily small and both (1) and (2) hold. Then we increase q to ensure that (3) holds as well, which does not affect (1) or (2).

Part 1: Coupled Stopping Procedure. Fix a word $\omega \in \Sigma$. We begin with two standard pairs $\hat{\gamma}_1$ and $\hat{\gamma}_2$. We will let P^i_n be the subset of $\hat{\gamma}_i$ that has not been coupled after n attempts at coupled stopping, i.e. it consists of points that are not permanently stopped at time $\lceil AR \rceil + i(q+1)\Delta$. Note that P^i_n is naturally viewed as a standard family. We let I^i_j be the set of points in P^i_j whose $(j+1)$st block is a tempered block. For every point $x \in P^i_j$ its next stopping time $T(x, \omega)$ is defined to be the end of the next tempered block for that point. To simplify the notation, we write $N_0 = \lceil AR \rceil$.

An inductive assumption of the following procedure is the following:

$$\begin{gathered}\text{For any } \hat{\gamma} \in P^i_j, \text{ and } x \in \hat{\gamma},\ \hat{\gamma} \text{ is sufficiently long that if for some } k > j, \\ \text{the } k\text{th block is tempered, then } f^{(q+1)(k-j)\Delta}_{\sigma^{N_0+(q+1)j\Delta}(\omega)}(\hat{\gamma}) \text{ is } C_0'\text{-good.}\end{gathered} \tag{8.10}$$

For $i \in \{1, 2\}$, let $\widetilde{U}^i_j$ be the union of the C_0' good intervals of the points $x \in I^i_j$ at the end of the $(j+1)$st block; if two intervals within a single standard pair in P^i_j overlap, we take their union, so some intervals may be longer than $e^{-C_0'}$. Note that $\widetilde{U}^i_j$ is a C_0'-good standard family. Then for each standard pair $I \in \widetilde{U}^i_j$, we discard the interval of size $e^{-\delta}$ from the end of the interval. This gives us a new standard family $U^i_j \subseteq \widetilde{U}^i_j$. By choice of $\delta(\alpha)$ from above,

$$\rho_i(U^i_j) \geq (1 - e^{-\alpha})\rho_i(\widetilde{U}^i_j).$$

We now choose which of the subpairs in $\widetilde{U}^1_j$ and $\widetilde{U}^2_j$ to stop on for our fixed word ω. Suppose without loss of generality that U^1_j has less mass than U^2_j. We now stop on all points in U^1_j. We would like to stop on all the points in U^2_j, however U^2_j has too much mass compared with U^1_j. To compensate, we subdivide the standard family to create pieces with the appropriate height so that we can stop on a set of equal mass to U^1_j. First we subdivide $\hat{\gamma}_2$ vertically at height $\rho_1(U^1_j)(\rho_2(U^2_j))^{-1}\rho_2$ so that we keep over each point the same proportion of the mass. Call the two pieces of $\hat{\gamma}_2$ by A and B, where A is the piece with mass $\rho_1(U^1_j)(\rho_2(U^2_j))^{-1}\rho_2(\hat{\gamma}_2)$. Then if we take A' to be the restriction of the standard pair A to the points over U^2_j, this subpair satisfies that $\rho_2(A') = \rho_1(U^1_j)$. We stop on all points in A'. The map Φ in the statement of the proposition associates A' and U^1_j. The complement of these stopped sets A' and U^1_j then defines a pair of new standard families P^i_{j+1}.

In order for us to be able to proceed with this argument inductively, we must verify that the inductive assumption (8.10) still holds. From the second part of Lemma 8.2, as long as $x \in f^{N_0+(j+1)(q+1)\Delta}_\omega(\gamma)$ has length at least $e^{-\delta}$, and a point x experiences another tempered time $q\Delta$ iterates later, then by choice of q,

$$e^{-\delta} > e^{-C_1}e^{-.8\lambda q\Delta},$$

so by that lemma if there is a future tempered time $n > N_0 + (j+1)(q+1)\Delta + q\Delta$, then at that time the image of x will lie in a C_0-good pair. Note that as we only consider future tempered times that are at least $q\Delta$ past the point where the curve is $e^{-\delta}$ long that by our choice of constants and the last part of Lemma 8.2 the assumption (8.10) holds inductively.

This completes the description of the stopping procedure. We now turn to estimating the tail of the stopping time.

Part 2: Rate of Stopping. Let A_n^1 and A_n^2 be the pairs $(\omega, x) \subset \Sigma \times \hat{\gamma}_1$ and $\Sigma \times \hat{\gamma}_2$ that have not permanently stopped at time $n(q+1)\Delta$, i.e. after n attempts at coupled stopping they are still not stopped. Our goal now is to show that $(\mu \otimes \rho_1)(A_n^1)$ has an exponential tail. We begin with several claims. The idea is that if the amount of mass that has not stopped at time n is large, then this implies that a large proportion of points will have a tempered time very quickly. If a large proportion of each curve has a tempered time, then we can stop on these points and obtain the result.

In this part of the proof, we will write all stopping times as if we had reindexed things so that $N_0 = \lceil AR \rceil$ is time 0, $\lceil AR \rceil + (q+1)\Delta$ is time 1, etc, to avoid a mess of notation. Keep in mind from our choice of constants earlier that we can pick Δ as large as we like at the beginning of the proof to ensure that α is as large as we like below.

Claim 8.6. *For any $\beta > 0$, there exists $\alpha_0 \geq 2\beta$ such that for all $\alpha \geq \alpha_0$, if we have chosen the block size Δ as above to ensure an $e^{-n\alpha}$ tail on tempered times pointwise* (8.9)*, then if for some $n \in \mathbb{N}$ and all $i < n$, $(\mu \otimes \rho_1)(A_i^1) \leq e^{-i\beta}e^{\beta}$ and $e^{-n\beta} \leq (\mu \otimes \rho_1)(A_n^1) \leq e^{2\beta}e^{-n\beta}$, then at the end of the next block, $1 - e^{-\frac{99}{100}\alpha}$ proportion of the points (ω, x) in A_n^1 experience a tempered time.*

Proof. Let $T(\omega, x)$ denote the next tempered time for $(\omega, x) \in A_n^1$ then we wish to study a conditional probability $\mathbb{P}(T(\omega, x) > n+1 | (\omega, x) \in A_n^1)$, as this gives a bound on the probability that we stop at the next attempt. Then

$$\mathbb{P}(T(\omega, x) > n+1 | (\omega, x) \in A_n^1) = \frac{\mathbb{P}(T(\omega, x) > n+1 \text{ and } (\omega, x) \in A_n^1)}{\mathbb{P}(A_n^1)} \tag{8.11}$$

Let $B_j^n \subseteq A_n^1$ be the set of trajectories that have not had a tempered time since iterate j and hence are in A_n^1 for this reason. Thus $A_n^1 = \sqcup_{j=0}^n B_j^n$. Note that $B_j^n \subseteq A_j^1$ as these points certainly weren't stopped at time j. Hence

$$\mathbb{P}(T(\omega, x) > n+1 | (\omega, x) \in A_n^1) = \frac{\sum_{j=0}^n \mathbb{P}(T(\omega, x) > n+1 \text{ and } (\omega, x) \in B_j^n)}{\mathbb{P}(A_n^1)}$$

$$\leq \frac{\sum_{j=0}^n \mathbb{P}(T(\omega, x) > n+1 \text{ and } (\omega, x) \in A_j^1)}{\mathbb{P}(A_n^1)} \leq (\mathbb{P}(A_n^1))^{-1} \sum_{j=0}^n e^{-(n-j+1)\alpha} e^{-\beta j + 2\beta} \quad \text{by (8.8)}$$

$$\leq e^{2\beta} e^{n(\beta-\alpha)} e^{-\alpha} \sum_{j=0}^n e^{j(\alpha-\beta)} = e^{2\beta} e^{-\alpha} \sum_{j=0}^n e^{(n-j)(\beta-\alpha)} = e^{2\beta} e^{-\alpha} \sum_{j=0}^n e^{j(\beta-\alpha)}$$

$$\leq e^{2\beta} e^{-\alpha} (1 + 2e^{(\beta-\alpha)}) \leq e^{-\frac{99}{100}\alpha},$$

for α sufficiently large relative to β. This is the needed claim, so we are done. □

The following claim shows that if most of the remaining pairs (ω, x) are experiencing a tempered time at time n then we stop on a relatively large amount of mass at that step.

Claim 8.7. *There exists α_0 such that for all $\alpha \geq \alpha_0$, if B_n^1 and B_n^2 are the subsets of A_n^1 and A_n^2 having tempered times at time $n+1$ and if for $i \in \{1,2\}$,*

$$(\mu \otimes \rho_i)(B_n^i) \geq (1 - e^{-\alpha})(\mu \otimes \rho_i)(A_n^i), \tag{8.12}$$

then

$$(\mu \otimes \rho_i)(A_{n+1}^i) \leq e^{-\alpha/3}(\mu \otimes \rho_i)(A_n^i). \tag{8.13}$$

Proof. Let $\pi\colon \Sigma \times \hat{\gamma}_1 \to \Sigma$ denote the projection. Associated to A_n^1 and A_n^2 we have a measure $\widetilde{\mu}_n$ on Σ, given by

$$\widetilde{\mu}_n(X) = (\mu \otimes \rho_1)(\pi^{-1}(X) \cap A_n^1).$$

Note that if we had used A_n^2 to define $\widetilde{\mu}_n$, we would have obtained the same result.

Let $A_n^i(\omega)$ denote $\pi^{-1}(\{\omega\}) \cap A_n^i$. We claim that there is a set $X \subseteq \Sigma$ such that $\widetilde{\mu}_n(X) \geq (1 - e^{-\alpha/2})(\mu \otimes \rho_1)(A_n^i)$ and for $\omega \in X$, we have that

$$\rho_1(A_n^1(\omega) \cap B_n^1) \geq (1 - e^{-\alpha/2})\rho_1(A_n^1(\omega)). \tag{8.14}$$

Otherwise there would exist a set Y such that $\widetilde{\mu}_n(Y) > e^{-\alpha/2}(\mu \otimes \rho_1)(A_n^1)$ such that for $\omega \in Y$, equation (8.14) fails. Then by Fubini, we would find

$$(\mu \otimes \rho_1)(B_n^1) \leq ((1 - \widetilde{\mu}_n(Y)) + \widetilde{\mu}_n(Y)(1 - e^{-\alpha/2}))(\mu \otimes \rho_1)(A_n^1) < (1 - e^{-\alpha/2})(\mu \otimes \rho_1)(A_n^1),$$

which is impossible from our assumption (8.12).

Thus we may find a set $X_1 \subseteq X$ such that $\widetilde{\mu}_n(X_1) \geq (1 - e^{-\alpha/2})(\mu \otimes \rho_1)(A_n^1)$ and for $\omega \in X_1$, (8.14) holds. Similarly we may find a set X_2 such that the same holds for A_n^2. Then $\widetilde{\mu}_n(X_1 \cap X_2) \geq (1 - 2e^{-\alpha/2})\widetilde{\mu}_n(A_n^1)$ and for every point $\omega \in X_1 \cap X_2$, each curve in $A_n^i(\omega)$ has at least $1 - e^{-\alpha/2}$ proportion of its remaining mass recovering. As described in the first part of the proof, we then trim segments of length $e^{-\delta}$ off these subcurves, which by the choice of δ, leaves us with $(1 - e^{-\alpha})$ proportion of the remaining mass. Thus on each curve there is at least

$$(1 - e^{-\alpha/2})(1 - e^{-\alpha})(\mu \otimes \rho_1)(A_n^1(\omega))$$

mass to stop on. Hence by the estimate on the measure of such ω, we can stop on

$$(1 - 2e^{-\alpha/2})(1 - e^{-\alpha/2})(1 - e^{-\alpha})(\mu \otimes \rho_1)(A_n^1)$$

of the remaining mass. In particular, this implies that for sufficiently large α, that the unstopped mass remaining at the $(n+1)$th step satisfies:

$$(\mu \otimes \rho_1)(A_{n+1}^1) \leq e^{-\alpha/3}(\mu \otimes \rho_1)(A_n^1), \tag{8.15}$$

as desired. □

We can now conclude the desired rate of stopping. From our choice of constants, we have $\beta > 0$ sufficiently small and $\alpha > 0$ sufficiently large that $\beta < \alpha/7$ and both Claims 8.6 and 8.7 of the proof hold. As mentioned previously, from the choice of Δ at the beginning, we may take α as large as we like. Then we will show that for $n \in \mathbb{N}$,

$$(\mu \otimes \rho_1)(A_n^1) \leq e^{-n\beta}e^{\beta}. \tag{8.16}$$

We consider two cases depending on how much mass is left at time n.

(1) First, suppose that

$$(\mu \otimes \rho_1)(A_n^1) \leq e^{-n\beta} \tag{8.17}$$

Then certainly, $(\mu \otimes \rho_1)(A^1_{n+1}) \leq e^{\beta} e^{-(n+1)\beta}$.

(2) If at time n,

$$e^{-n\beta} \leq (\mu \otimes \rho_1)(A^1_n) \leq e^{2\beta} e^{-\beta n}, \tag{8.18}$$

and at all previous times $(\mu \otimes \rho_1)(A^1_n) \leq e^{\beta} e^{-n\beta}$, then Claim 8.6 applies to A^1_n and A^2_n, which gives that at time $n+1$, that $1 - e^{-99/100\alpha}$ proportion of the points in A^1_n and A^2_n will recover at time $n+1$. Thus by Claim 8.7 and our choice of $\alpha > 7\beta$, we see that

$$(\mu \otimes \rho_i)(A^i_{n+1}) \leq e^{-\frac{99}{300}\alpha}(\mu \otimes \rho_i)(A^i_n) < e^{-2\beta}(\mu \otimes \rho_i)(A^i_n), \tag{8.19}$$

and for the next iterate we are back in the first case, $(\mu \otimes \rho_1)(A^1_{n+1}) \leq e^{-(n+1)\beta}$.

In order to conclude, we apply the two options above inductively to obtain equation (8.16) for all n. In fact, we will show something slightly stronger: there are never two consecutive indices $n, n+1$ such that

$$e^{-n\beta} < (\mu \otimes \rho_1)(A^1_n) \leq e^{-n\beta} e^{\beta}$$

holds for both n and $n+1$.

Throughout the induction either we have

$$(\mu \otimes \rho_i)(A^i_n) < e^{-\beta n} \text{ or } e^{-n\beta} \leq (\mu \otimes \rho_i)(A^i_n). \tag{8.20}$$

In the former case, we may apply item (1) in the list just mentioned.

Suppose we are in the latter case, that at time $n-1$ that $(\mu \otimes \rho_i)(A^i_n) < e^{-\beta(n-1)}$ and at time n that $e^{-\beta n} \leq (\mu \otimes \rho_i)(A^i_n) \leq e^{-\beta(n-1)}$, and that for all prior iterates equation (8.16) holds. Then we may apply (2) above to find that

$$(\mu \otimes \rho_i)(A^i_{n+1}) < e^{-2\beta}(\mu \otimes \rho_i)(A^i_n) \leq e^{-2\beta} e^{-(n-1)\beta} = e^{-(n+1)\beta}. \tag{8.21}$$

Thus for the iteration $n+1$ we have $(\mu \otimes \rho_i)(A^i_{n+1}) < e^{-(n+1)\beta}$. Note that this means that the second case in (8.20) cannot occur twice in a row. Hence we may proceed inductively to verify that (8.16) holds for every n. This concludes the proof of the lemma. □

9. Precoupling

In this section, we prove the finite time mixing proposition, Proposition 7.11, which prepares curves for the application of the local coupling lemma.

9.1. **Fibrewise mixing.** In this subsection we study fiber-wise mixing properties of the skew product $F\colon \Sigma \times M \to \Sigma \times M$. A skew product being mixing does not imply that it has any mixing properties fiberwise. For example, the system could be isometric on the fibers. For this reason we will leverage the mixing of $F_k\colon \Sigma \times M^k \to \Sigma \times M^k$. We will obtain a sort of coarse fiberwise mixing by using a concentration of measure argument. The basic idea of the argument is that if A is a subset of M, and $B \subset \Sigma \times M$ is a set giving equal measure to each fiber, then if B does not mix with A fiberwise, then it implies that on many fibers $A \cap F^n(B)$ is quite concentrated. As a consequence of this concentration we show that F_k cannot be mixing as there are too many points that stay in the set $A^k \subset M^k$.

Proposition 9.1. *Suppose that the skew product $F_k\colon \Sigma \times M^k \to \Sigma \times M^k$ from (6.1) is mixing for $\mu \otimes \mathrm{vol}^k$ for all $k \in \mathbb{N}$. Let $A \subseteq M$ be a positive measure set. Then for all $\epsilon_1, \epsilon_2 > 0$ if $U \subseteq \hat{\Sigma} \times M$ is a set giving exactly mass $\alpha_0 > 0$ to $(1-\epsilon_2)$ of the fibers of $\hat{\Sigma}$ and 0 to the rest, then there exists $N \in \mathbb{N}$, such that for all $n \geq N$, there exist $(1-2\epsilon_2)$ proportion of words $\omega \in \hat{\Sigma}$, such that*

$$\mathrm{vol}(A)\alpha_0(1-\epsilon_1) \leq \mathrm{vol}(f^n_\omega(U_\omega) \cap A) \leq \mathrm{vol}(A)\alpha_0(1+\epsilon_1), \tag{9.1}$$

where we write $U_\omega \subseteq M$ for the portion of U in the fibre over ω.

Proof. We will prove the lower bound; the upper bound then follows by taking the complement of A. For the sake of contradiction, suppose that the lower bound in (9.1) is false. Then there exist $\epsilon_1, \epsilon_2 > 0$ such that for arbitrarily large n, there exist measure $2\epsilon_2$ words ω such that

$$\operatorname{vol}(U_\omega) = \alpha_0 \text{ and } \operatorname{vol}(f^n_\omega(U_\omega) \cap A) < \operatorname{vol}(A)\alpha_0(1-\epsilon_1). \tag{9.2}$$

For these words ω

$$\operatorname{vol}(f^n_\omega(U_\omega) \cap (M \setminus A)) \geq \alpha_0(\operatorname{vol}(M \setminus A) + \epsilon_1 \operatorname{vol}(A)). \tag{9.3}$$

We now consider what this implies on $\hat{\Sigma} \times M^k$. Write U^k for the union of the sets $\{\omega\} \times U^k_\omega$. Then for the words ω satisfying (9.2), we obtain

$$(\operatorname{vol}^k)(F^n_{k,\omega}(U^k_\omega) \cap \{\sigma^k(\omega)\} \times (M \setminus A)^k) \geq \alpha_0^k(\operatorname{vol}(M \setminus A) + \epsilon_1 \operatorname{vol}(A))^k, \tag{9.4}$$

because fiberwise this intersection is equal to the product $(f^n_\omega(U_\omega) \cap (M \setminus A))^k$. Thus integrating over this set of ω of measure $2\epsilon_2$, we find that

$$(\hat{\mu} \otimes \operatorname{vol}^k)(F^n_k(U^k) \cap \hat{\Sigma} \times (M \setminus A)^k) \geq 2\epsilon_2\alpha_0^k(\operatorname{vol}(M \setminus A) + \epsilon_1 \operatorname{vol}(A))^k. \tag{9.5}$$

Note that $(\hat{\mu} \otimes \operatorname{vol}^k)(U^k) \leq (1-\epsilon_2)\alpha_0^k$ by the definition of U. Since $(\hat{\mu} \otimes \operatorname{vol}^k)(\hat{\Sigma} \times (M \setminus A)^k) = \operatorname{vol}(M \setminus A)^k$, mixing of F_k implies that for sufficiently large n,

$$(\hat{\mu} \otimes \operatorname{vol}^k)(F^n_k(U^k) \cap \hat{\Sigma} \times (M \setminus A)^k) \leq (1-\epsilon_2/2) \operatorname{vol}(M \setminus A)^k \alpha_0^k. \tag{9.6}$$

For large k the bounds (9.5) and (9.6) are incompatible, so we obtain a contradiction. □

9.2. **Proof of the finite time mixing proposition.** In this subsection we prove the finite time mixing Proposition 7.11. The idea is straightforward. We can saturate the curve $\hat{\gamma}$ with stable manifolds to embed $\hat{\gamma}$ in a positive measure set that will contract onto the image of $\hat{\gamma}$ forward in time. As the skew product $F\colon \hat{\Sigma} \times M \to \hat{\Sigma} \times M$ is fibrewise mixing (Proposition 9.1), this positive measure thickening of $\hat{\gamma}$ must equidistribute for most words. Simultaneously, we know that most images of $\hat{\gamma}$ will be relatively smooth. This allows us to conclude.

In the proof we will need some intermediate claims.

Definition 9.2. An *ϵ-thickening* of a curve γ for a word $\omega \in \Sigma$ consists of two pieces of information. The first piece is a subset $\gamma_0 \subset \gamma$ that will be thickened. The second piece is a set of the form

$$\bigcup_{x \in \gamma_0} W^s_{\epsilon(x)}(\omega, x),$$

and $W^s_{\epsilon(x)}(\omega, x)$ is the local stable leaf of radius $0 < \epsilon(x) < \epsilon$ through x. We will often denote such sets by $\kappa_\omega(\gamma)$.

Note that although the thickening can in principle be defined over all of γ, we will usually only use it on a special subset γ_0 that has better properties.

The following lemma shows that we may choose thickenings of γ so that the pushforward of the volume along the thickening to γ by the stable holonomy is proportional to ρ on γ_0.

Lemma 9.3. *(Local Thickening Lemma) Fix $\epsilon_1 > 0$ and $C_0 > 0$, a level of goodness of standard pairs. For any $\epsilon_2 > 0$, there exist $\epsilon_3, c_1, C_2, \varrho > 0$ such that for $1-\epsilon_2$ of words $\omega \in \Sigma$, and any C_0-good standard pair $\hat{\gamma} = (\gamma, \rho)$ of unit mass, we can form an ϵ_1-thickening of γ, $\kappa_\omega(\gamma)$, in the sense of Definition 9.2, such that:*

(1) Let π^s be the projection to γ along the stable leaves. Then $\pi^s_*(\operatorname{vol}|_{\kappa_\omega(\gamma)}) = c_1\rho|_{\pi^s(\kappa_\omega(\gamma))}$ and $\rho(\pi^s(\kappa_\omega(\gamma))) > \varrho$.

(2) Every stable leaf in $\kappa_\omega(\gamma)$ is uniformly $(C_2, \lambda, \epsilon_3)$-tempered under forward iterations.
(3) The choice of thickening $\kappa_\omega(\gamma)$ depends measurably on ω.

Proof. We know that for every point x and almost every word ω, that x is in the Pesin block $\Lambda^\omega_\infty(C)$ for some sufficiently large C, and on a measure one subset, E^s is not tangent to γ. Thus we can saturate a positive measure subset of γ with stable manifolds with uniformly controlled geometry by increasing C. By taking a shorter subset of the saturating stable curves in such a Pesin block, we can ensure that the volume measure of the saturation projected along the stable leaves to γ gives a measure that is proportional to ρ restricted to the images of π^s. □

The following lemma says that if we start with C-good curve, then we can ensure that a large proportion of the images of the curve are C_0-good at any time in the future.

Lemma 9.4. *For any $\epsilon > 0$, there exists C_0, such that for any $C > 0$, a level of goodness, there exists $N_0 \geq 0$, such that for any C-good standard pair $\hat{\gamma}$ and all $n \geq N_0$, there exists a set $\Sigma_0^n \subseteq \Sigma$ of measure at least $1 - \epsilon$, such that for $\omega \in \Sigma_0^n$,*

$$\rho(x : f^n_\omega(\hat{\gamma}) \text{ has a } C_0\text{-good neighborhood of } f^n_\omega(x)) \geq (1-\epsilon)\rho(\hat{\gamma}). \tag{9.7}$$

The same holds for a C-good standard family.

Proof. This is immediate from Proposition 8.4, which says that for large enough Δ, we may ensure that $1 - \epsilon$ of the pairs (ω, x) will have a tempered time between times $n + \Delta$ and $n + 2\Delta$ for any n. We choose N_0 large enough that such a tempered time recovers a C-good curve to being D-good for some uniform D. Then we wait until to the end of the block, which gives a further, bounded loss of goodness. As in other places in the paper, a Fubini argument gives the fiberwise estimate stated here. Finally, note that this argument is independent of $n \geq N_0$. □

We are now ready to prove the finite time mixing proposition.

Proof of Proposition 7.11. The outline of the proof is as follows. We first find a collection of balls in M that a thickened version of γ_1 and γ_2 will mix onto due to the fibered mixing lemma. Then once mixing is accomplished most subcurves of $f^n_\omega(\gamma_1)$ and $f^n_\omega(\gamma_2)$ will still be long. Consequently, if there are subcurves intersecting a small ball $B_\upsilon(x)$ then those subcurves will form a (C_1, δ, υ) configuration for some C_1. To achieve this setup, we will construct subsets $\Sigma_0, \ldots, \Sigma_4$ of Σ. Each of these sets will consist of words ω that have some particular finite time mixing properties, so that their intersection has all the properties we need to conclude along the lines just described. We will also have some additional parameters m_i that are chosen below.

The input to this proposition requires some constants. First, let $0 < \tau < 1$ be the constant from the Local Coupling Lemma, Lemma 7.10, which says that the conclusions of that lemma hold for $(C, \delta, \tau\delta)$-configurations for any C as long as δ is sufficiently small relative to C. We then obtain the following claim—note that this holds for all sufficiently small δ with a uniform lower bound in the last term.

Claim 9.5. *There exists m_0 such that for all sufficiently small $\delta > 0$, we can find a family of disjoint balls $B_i = B_{\delta_i}(x_i)$ in M such that:*
(1) Each B_i has equal volume between $(1/10)\delta^2$ and $10\delta^2$;
(2) Each B_i contains a ball B'_i of diameter at most $\tau\delta/2$ so that $d(\partial B'_i, \partial B_i) > \delta/2$;
(3) Each B'_i contains a ball B''_i with the same center and radius between $\tau\delta/9$ and $\tau\delta/10$, and the balls B''_i all have equal volume;

(4) $\operatorname{vol}(\bigcup_i B_i'') \geq 10^{-m_0}$.

We now pick Σ_0, which are words where γ_1 and γ_2 have good thickenings. Both γ_1 and γ_2 are C_0-good by assumption. Then for any $m_1 \in \mathbb{N}$, which we will pick later, we see that there exists c_1, which is distinct from C_1, and ϱ such that there is a set $\Sigma_0 \subseteq \Sigma$, such that $\mu(\Sigma_0) > 1 - 10^{-m_1}$ such that for $\omega \in \Sigma_0$, there exists a thickening $\kappa_\omega(\gamma_i)$, $i \in \{1, 2\}$ satisfying the properties of Lemma 9.3. By possibly shrinking the thickening we may make the thickenings each have the same identical mass ϱ. For the words in Σ_0, we form a set $U^1 \subseteq \Sigma \times M$ by taking the union of the sets $\{\omega\} \times \kappa_\omega(\gamma_1)$, similarly we define U^2. We denote by U^1_ω the part of U^1 above ω and use a similar notation for U^2_ω.

We now choose C_1, the regularity of the pairs that will be in $(C_1, \delta, \tau\delta)$-configurations in the conclusion of the proposition, as well as Σ_1^n and Σ_2^n, words where most images of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are C_1-good curves. Choose $C_1 > 0$ such that the conclusion of Lemma 9.4 holds for a set Σ_1^n of words of measure $(1 - 10^{-m_2})$, for some m_2 that we will choose later, so that for $\omega \in \Sigma_1^n$,

$$\rho_1(x : f^n_\omega(\hat{\gamma}_1) \text{ has a } C_1\text{-good neighborhood of } f^n_\omega(x)) \geq (1 - 10^{-m_2})\rho_1(\hat{\gamma}_1). \tag{9.8}$$

For all $\epsilon > 0$, there exists $D(\epsilon) > 0$ such that for a C_0-good standard pair $\hat{\gamma} = (\gamma, \rho)$, the measure of the points $x \in \hat{\gamma}$ such that

$$\rho(x \in \gamma : d(x, \partial\hat{\gamma}) < D) < \epsilon\rho(\gamma). \tag{9.9}$$

Recalling Definition 7.8, the previous equation implies that there exists $D_1 > 0$ such that we may strengthen the conclusion in equation (9.8) above:

$$\rho_1(x : f^n_\omega(x) \text{ is } (C_1, D_1)\text{-well positioned in } f^n_\omega(\hat{\gamma}_1)) \geq (1 - 10^{-(m_2-1)})\rho_1(\hat{\gamma}_1). \tag{9.10}$$

Call this set of (C_1, D_1)-well positioned points $G^1_{n,\omega}$. Similarly, for $\hat{\gamma}_2$ there exists a set Σ_2^n and a set $G^2_{n,\omega}$ with this same property.

Take a covering B_i as in Claim 9.5 applied with the parameter δ small enough that the local coupling lemma holds for $(C_1, \delta, \tau\delta)$-configurations. Let $m_1 = m_2 = m_3 = m_4 = 20$.

Next we choose Σ_3^n and Σ_4^n, which are sets that mix the thickenings of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ onto the balls B_i''. Let $\epsilon_2 = 10^{-m_3}$ from above, and let $0 < \epsilon_1 < 10^{-m_3}$. Then by the fibrewise mixing proposition (Proposition 9.1), there exists N_1 such that for $n \geq N_1$, there is a set Σ_3^n of $\omega \subseteq \Sigma$ of μ-measure $1 - 2 \cdot 10^{-m_3}$ such that for $\omega \in \Sigma_3^n$, U^1_ω mixes onto the B_i'' for each B_i'' in the covering, i.e. for $\omega \in \Sigma_3^n$,

$$(1 - 10^{-m_3})\varrho \operatorname{vol}(B_i'') \leq \operatorname{vol}(f^n_\omega(U^1_\omega) \cap \{\omega\} \times B_i'') \leq \operatorname{vol}(B_i'')\varrho(1 + 10^{-m_3}). \tag{9.11}$$

Similarly we have a cutoff N_2, and sets Σ_4^n such that the same holds for U_2. We will strengthen this estimate even further, we will let B_i''' be a ball with the same center as B_i'' but with slightly larger radius so that the ratio of the volumes $\operatorname{vol}(B_i''')/\operatorname{vol}(B_i'') = 1 + 10^{-m_4}$. Then by possibly enlarging the numbers N_1 and N_2, we can arrange that the same estimate holds simultaneously for the sets B_i''' as well.

Now consider what happens for $\omega \in \Sigma_0 \cap \Sigma_1^n \cap \Sigma_3^n$. These are words ω where γ_1 has a good thickening by stable manifolds, and many of the points in the image of $\hat{\gamma}_1$ are good standard pairs and there is equidistribution. For any m_4 as long as n is sufficiently large, the diameter of the image of any $W^s_{\epsilon(x)}(\omega, x)$ leaf in the thickening of $\hat{\gamma}_1$ is at most $\tau\delta/10^{2m_4}$. Thus from the measure preservation of the projection π^s of $f^n_\omega(U_\omega)$ onto $f^n_\omega(\hat{\gamma}_1)$, we see that if some point $x \in f^n_\omega(U^1_\omega)$ is in B_i'', then as B_i''' contains a neighborhood of B_i'' of radius $\tau\delta/10^{m_4}$, all points on $f^n_\omega(W^s_{\epsilon(x)}(\omega,x))$ and, in particular, the points of $\hat{\gamma}$ lie in this set. Hence, writing $\rho_1^{n,\omega}$ for the density on $f^n_\omega(\hat{\gamma}_1)$,

$$\rho_1^{n,\omega}(f^n_\omega(\hat{\gamma}_1)\cap B_i''') \geq \operatorname{vol}(B_i'')(1-10^{-m_3})\rho_1(\hat{\gamma}_1). \tag{9.12}$$

We claim that for such $\omega\in\Sigma_0\cap\Sigma_1^n\cap\Sigma_3^n$ that there exists a subfamily of the B_i'' containing at least 90% of the B_i'', and such that for each of these B_i'',

$$\rho_1^{n,\omega}(G^1_{n,\omega}\cap B_i''') \geq \operatorname{vol}(B_i''')\rho_1(\hat{\gamma})/2. \tag{9.13}$$

Suppose that this were not the case, then for such an ω there is a set of 10% of the balls B_i''' such that for these balls we have $\rho_1^{n,\omega}(G^1_{n,\omega}\cap B_i''')<\operatorname{vol}(B_i''')\rho_1(\hat{\gamma})/2$. Then, from (9.10) and the fibrewise mixing estimate (9.11),

$$\operatorname{vol}(f^n_\omega(U_\omega)\cap\bigcup_i B_i'') \leq \frac{\rho_1^{n,\omega}(G^1_{n,\omega}\cap\bigcup_i B_i''')\varrho}{\rho_1(\hat{\gamma})}+10^{-(m_2-1)}\varrho$$

$$\leq .1\sum_i \operatorname{vol}(B_i''')\varrho\frac{1}{2}+.9\sum_i\operatorname{vol}(B_i''')\varrho(1+10^{-m_3})+10^{-(m_2-1)}\varrho$$

$$\leq .96\sum_i\operatorname{vol}(B_i''')\varrho \leq .96(1+\frac{1}{10^{m_4-1}})\varrho\sum_i\operatorname{vol}(B_i'')$$

which contradicts fiberwise mixing of U^1 with the set $\bigcup_i B_i'''$.

Now consider $\omega\in\Sigma_0\cap\Sigma_1^n\cap\Sigma_2^n\cap\Sigma_3^n\cap\Sigma_4^n$. We have that for 90% of the balls B_i''', that B_i''' has radius at most $\tau\delta/8$, and this ball contains points of $f^n_\omega(\hat{\gamma}_1)$ that are (C_1,D_1)-well centered of measure at least $\rho_1(\hat{\gamma}_1)\tau^2\delta^2/200$. The same holds for $\hat{\gamma}_2$ for a possibly different 90% of balls. Thus for 80% of the balls B_i''' each of $f^n_\omega(\hat{\gamma}_1)$ and $f^n_\omega(\hat{\gamma}_1)$ contains measure $\rho_1(\hat{\gamma}_1)\tau^2\delta^2/200$ points that are (C_1,D_1)-well centered. As these points are in a ball of radius $\tau\delta/8$. From our choice of δ, it follows that any pair of such images is $(C_1,\delta,\tau\delta)$-configured. Thus the needed conclusion follows by possibly subdividing the standard pairs we have identified so that they may be coupled in a measure preserving way. We may now conclude because

$$\mu(\Sigma_0\cap\Sigma_1^n\cap\Sigma_2^n\cap\Sigma_3^n\cap\Sigma_4^n)\geq 1-10^{-m_1}-2\cdot10^{-m_2}-4\cdot10^{-m_3}\geq 1-10^{-19}. \qquad \square$$

10. Proof of the Local Coupling Lemma

10.1. **Inductive local coupling procedure.** To prove the Local Coupling Lemma 7.10, we would like two positive measure sets to be intertwined under the true stable holonomy. However, at any finite time, we do not yet know what the true limiting stable manifold is. To compensate, at finite times we approximate the limiting holonomy by using the fake stable manifolds. In the proof of the local coupling lemma, we will consider the differences between different standard families as discussed in §7.1.

To begin this section, we introduce a notion of a "fake coupling" of two standard pairs $\hat{\gamma}_1$ and $\hat{\gamma}_2$. We use fake couplings because in our setting we cannot use the stable manifold as is done in the deterministic setting. In the deterministic setting, if $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are near each other, then we can immediately determine which points in $\hat{\gamma}_1$ attract to which in $\hat{\gamma}_2$ by using the stable holonomy. We work in an opposite manner: at each time n we discard points that cannot couple yet. For example, if $y\in\hat{\gamma}_2$ and none of the time n fake stable manifolds come near y, then y can't couple because the true stable manifold is near the fake one. Consequently, we stop trying to couple y at time n. After we see the dynamics for all time, the points that remain in $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are those that can be coupled with each other using the stable manifold. Hence after the fact, we see that they were coupled. The fake coupling is not a coupling. A time n fake-coupling is a pair of subfamilies $P_n^1\subseteq\hat{\gamma}_1$ and $P_n^2\subseteq\hat{\gamma}_2$ that could *potentially* be coupled by the true stable manifolds. For a time n fake coupling, we insist that the holonomies of the time n fake stable manifolds carry P_n^1 to P_n^2. Another way to describe this is that P_n^1 and P_n^2 seem coupled until time n.

The definition of a fake coupling that follows that is adapted to the neighborhood $B_{\delta_0}(x)$ from Proposition 10.12 and relies on the constants obtained in that proposition. Fake stable manifolds W_n^s and their properties are discussed in detail in Appendix B.

Definition 10.1. Suppose that $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are two standard pairs that we are attempting to couple that are (C, δ', υ)-configured where C, δ, υ are parameters as in Proposition 10.12. Fix some x and neighborhood $B_{\delta_0}(x)$, $\mathcal{C}_\theta$ as in part 4 of that Proposition. We will use the other constants from that proposition as well without reintroducing them.

For $n \geq N$, we say that $P_n^1 \subseteq \hat{\gamma}_1$ and $P_n^2 \subseteq \hat{\gamma}_2$ are a $(b_0, \hat{\eta})$*-fake coupled pair at time* $n \geq N$ for some word ω on $B_{\delta_0}(x)$ if the following statements hold. Write ρ_n^1 and ρ_n^2 for the densities of P_n^1 and P_n^2 on γ_1 and γ_2. Let $\mathcal{I}_n^1$ and $\mathcal{I}_n^2$ be the underlying curves of P_n^1 and P_n^2.

(1) P_n^1 and P_n^2 have equal mass and $(H_{n-1}^s)_*(\rho_n^1) = \rho_n^2$.
(2) H_{n-1}^s carries $\mathcal{I}_n^1$ to $\mathcal{I}_n^2$.
(3) If $x \in \gamma_1$ is $(C, \lambda, \epsilon, \mathcal{C}_\theta)$-tempered for times $N \leq i \leq n$, then $x \in \mathcal{I}_n^1$.
(4) At each point x in the curve underlying P_n^1, we have that

$$\rho_n^1(x) \geq b_0 \prod_{N \leq i \leq n} (1 - e^{-i\hat{\eta}})\rho^1(x).$$

We will see below that if for a given word ω we are able to arrange that the statements above hold for each n, then in the limit, for each point $x \in \gamma_1$ that is (C, λ, ϵ)-tempered and in each P_n^1 that at least ϵ_0 of the mass above x in $\hat{\gamma}_1$ couples. Thus as typically a positive measure set of x have this property, a positive proportion of the mass of P_n^1 couples.

The structure of the rest of this section is as follows. In §10.2 and §10.3 we show that if a trajectory has a tempered splitting then nearby trajectories also have tempered splittings. In §10.4 we prove Proposition 10.12 which shows how small a scale we need to work at in order to run a coupling procedure. Then in §10.5 we prove the local coupling lemma in two steps. First, we prove Lemma 10.13, which describes a deterministic local coupling procedure that can be applied to a fixed word ω under the choice of constants provided by Proposition 10.12. We then finish the proof of Lemma 7.10 by using that the hypotheses of this deterministic local coupling procedure are satisfied with high probability.

10.2. **Nearby points inherit tempered splitting.** In this subsubsection we prove Proposition 10.3, which says that nearby trajectories inherit splittings from each other. This will be used later to show that the set of points on a curve that have a tempered splitting after n iterations is quite fat. The idea that points close to hyperbolic orbits inherit hyperbolicity is useful in many problems in dynamics. For example, a classical Collet–Eckmann condition is used in one dimensional dynamics to show that near critical orbits recover hyperbolicity if the critical orbit is hyperbolic (see [CE80]). Analogous results for two dimensional strongly dissipative maps appear in [BC91, WY01]. In this paper we present a version for general two dimensional maps based on Pesin theory.

We begin with a fact showing how far attracting and repelling directions of a linear map of $\mathbb{R}\mathrm{P}^1$ move under perturbation.

Lemma 10.2. *Fix some $\lambda > 1$, then there exists $C, \epsilon_0, \delta_0, N_0 > 0$, such that if $L\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a linear map of the form*

$$\begin{bmatrix} \sigma_1 & 0 \\ 0 & \sigma_2 \end{bmatrix} \tag{10.1}$$

with $|\sigma_1|, |\sigma|_2^{-1} \geq |\lambda| > 1$, $g_0\colon \mathbb{R}\mathrm{P}^1 \to \mathbb{R}\mathrm{P}^1$ is the induced map, and g_ϵ is a perturbation with $d_{C^1}(g_0, g_\epsilon) = \epsilon < \epsilon_0$, then:

(1) g_ϵ has a unique repelling fixed point r_ϵ and a unique attracting fixed point a_ϵ, and these satisfy $d(r_\epsilon, (0,1)) \leq C\epsilon$, $d(a_\epsilon, (1,0)) \leq C\epsilon$.

(2) On the neighborhood $B_{\delta_0}((0,1))$, $\|Dg_\epsilon\| \geq \lambda - C\epsilon$ and on the neighborhood $B_{\delta_0}((1,0))$, $\|Dg_\epsilon\| \leq \lambda^{-1} + C\epsilon$. These neighborhoods are overflowing and under-flowing, respectively.

(3) If $y \notin B_{\delta_0}((0,1))$, then $g_\epsilon^{N_0}(y) \in B_{\delta_0}((1,0))$.

We omit the proof of the above lemma as these are standard facts about the dynamics in a neighborhood of a hyperbolic fixed point. The proof of the next result is long and relies on a number of intermediate lemmas.

Proposition 10.3. *(Nearby points inherit temperedness) Fix $C_0, C_1, \lambda, \alpha, \epsilon_0, D_0, \sigma > 0$ and $0 < \lambda' < \lambda$. Then for sufficiently small $\epsilon > 0$ there exist $\nu, k, D_1, N > 0$ such that $k\epsilon < \epsilon_0$ and if we have a sequence of matrices of length $n \geq N$ $(A_i)_{1\leq i\leq n} \in \mathrm{SL}(2,\mathbb{R})$ that are uniformly bounded in norm by D_0 and are (C_0, λ, ϵ)-tempered, and $(B_i)_{1\leq i \leq n}$ is another sequence of matrices such that $\|A_i - B_i\| \leq C_1 e^{-\alpha(n-i)}$ then:*

1. *B_i has a $(D_1C_0, \lambda', k\epsilon)$-subtempered splitting with the stable direction equal to the contracting singular direction of B^n, and*
2. *The angle between $(B_i)_{i=1}^n$ and $(A_i)_{i=1}^n$'s stable directions is at most $e^{-\nu n}$.*
3. *$\|B^n\| \geq \|A^n\|^{(1-\sigma)}$.*

Proof. Before we begin, observe that due to the presence of the factor D_1 in the conclusion, it suffices to show that the needed claim holds for n sufficiently large as we may always deal with small n by adjusting D_1. Let $\hat{\lambda} = \frac{\lambda+\lambda'}{2}$.

As long as $\epsilon_0 < (\lambda - \lambda')/2$, we may view the sequence of matrices A_i in the finite time Lyapunov charts from Lemma A.1, where we view the sequence as being $(C_0, \hat{\lambda}, \epsilon + \frac{\lambda-\lambda'}{2})$-tempered. In these charts, we have: $A_i = \begin{bmatrix} \sigma_{1,i} & 0 \\ 0 & \sigma_{2,i} \end{bmatrix}$, where $\min\{\sigma_{1,i}, \sigma_{2,i}^{-1}\} \geq e^{\hat{\lambda}}$. From Lemma A.1, the ratio of the reference norm and the Lyapunov norm at step i is $O_{C_1,\alpha,\lambda,\lambda'}(e^{4\epsilon i})$.

As B_i is a perturbation of size $e^{-\alpha(n-i)}$ by viewing B_i in the same Lyapunov coordinates as A_i, we have that

$$B_i = \begin{bmatrix} \sigma_{1,i} & 0 \\ 0 & \sigma_{2,i} \end{bmatrix} + O_{C_1,\alpha,\lambda,\lambda'}(e^{-\alpha(n-i)}e^{4\epsilon i}), \tag{10.2}$$

where $\min\{\sigma_{1,i}, \sigma_{2,i}^{-1}\} \geq e^{\hat{\lambda}}$.

Using this representation, we will now study B_i as a perturbation of the matrix product involving the A_i. For most i, the two are quite close and consequently B_i will inherit temperedness of its norm. The remaining i will be negligible. To show this, we first identify where the stable direction of B^n lies. Then using this we show that the norm of B^i is subtempered up to a particular time. Then we do a little bookkeeping to show that if we relax the subtemperedness condition, then norm will remain subtempered up to time n.

First we study how temperedness changes as we continue appending matrices to a sequence.

Lemma 10.4. *Fix some bound $e^{\Delta} > 1$. Suppose that $A_1, \ldots, A_n$ is a sequence of matrices whose splitting into singular directions is (C, λ, ϵ)-tempered. Then for any k and sequence $B_1, \ldots, B_m$ with $\|B_i\| \le e^{\Delta}$ and $m < \Delta^{-1}(nk\epsilon - C)$, the sequence $A_1, \ldots, A_n, B_1, \ldots, B_m$ is $(C, \lambda - k\epsilon, k\epsilon)$-tempered.*

Proof. A straightforward generalization of Lemma 4.15 gives that if we have a sequence of matrices $A_1, \ldots, A_n$ with (C, λ, ϵ)-tempered norm and we append a sequence $B_1, \ldots, B_m$ that is $(-\Delta m, \lambda - \epsilon, \epsilon)$-tempered, then the concatenation is a $(\tilde{C}, \lambda - k\epsilon, k\epsilon)$ tempered sequence with

$$\tilde{C} = \min\{C, C - m\Delta + nk\epsilon/2, -m\Delta + nk\epsilon\}. \tag{10.3}$$

Thus the needed conclusion holds as long as $m \le \dfrac{nk\epsilon - C}{\Delta}$. □

The following lemma gives tight control on where v^s, the most contracted vector for the sequence $(B_i)_{i=1}^n$ lies. Below we will write $g_{i,\epsilon}$ for the map on $\mathbb{R}\mathrm{P}^1$ induced by B_i, viewed in the Lyapunov coordinates above. We write g_ϵ^i for the composition $g_{i,\epsilon} \circ \cdots \circ g_{1,\epsilon}$.

Lemma 10.5. *For all $C_0, C_1, \alpha, \lambda, \lambda', D_0 > 0$ as above and all sufficiently small $\epsilon > 0$, there exists $\nu > 0$ and $N_s \in \mathbb{N}$ such that if $n \ge N_s$ and $(B_i)_{i=1}^n$ is a sequence of matrices as above, a perturbation of $(A_i)_{i=1}^n$, a sequence of matrices with a (C_0, λ, ϵ)-subtempered splitting, then the most contracted direction of B^n, v_B^s, lies within a neighborhood of size $e^{-n\nu}$ of the most contracted direction of A^n.*

Proof. We will use the perturbed dynamics $g_{i,\epsilon}$ on $\mathbb{R}\mathrm{P}^1$ from above and prove this result by studying how fast a vector near the vector $(0,1)$ escapes and goes to $(1,0)$. We will use the estimates of Lemma 10.2 freely and not restate them here. Given $\delta_0 > 0$ in the conclusion of that lemma, we see that as long as the size of the perturbation is at most some ϵ_{δ_0}, then on the neighborhoods of size δ_0 of $(0,1)$ the expansion is by a factor of at least $e^{.9\hat{\lambda}}$ and similarly in the δ_0-neighborhood of a_ϵ, the contraction of distance is by a factor of $e^{-.9\hat{\lambda}}$. As long as ϵ is sufficiently small relative to α and $i \le \frac{99}{100}n$, then $g_{i,\epsilon}$ is a perturbation of size less than ϵ_{δ_0} and the estimate for the norm of $g_{i,\epsilon}$ on $B_{\delta_0}((1,0))$ and $B_{\delta_0}((0,1))$ holds.

Next, we study the norm growth of v over its entire trajectory. Define $\Phi_\epsilon^i \colon \mathbb{R}\mathrm{P}^1 \to \mathbb{R}^+$ by

$$\Phi_\epsilon^1(v) = \ln \frac{\|B_i v\|}{\|v\|}. \tag{10.4}$$

Then $\|B^i v\|$ is the sum of Φ_ϵ^i along the trajectory of v. We divide the trajectory of v into three segments. The first segment is when v is does not yet lie in $B_{\delta_0}((1,0))$. The middle segment is when it lies in $B_{\delta_0}((1,0))$ and B_i remains a small enough perturbation of A_i that we may use the approximations of Lemma 10.2. Finally, during the last part of the trajectory i is so big that these estimates no longer hold. We will let $1 < n_1 < n_2 < n$ denote the indices where $g_\epsilon^i(v)$ first enters $B_{\delta_0}((1,0))$ and n_2 the index where the approximations of Lemma 10.2 first cease to hold. We now proceed to estimate how large n_1 and n_2 are. Then using this information we will calculate $\|B^n v\|$.

By estimating in this manner, we will see that any vector that starts at distance more than $e^{-n\nu}$ from $(0,1)$ cannot be a stable vector as its norm grows. Below, we will track the estimates for B_i, the same apply to A_i. Consequently, we see that the stable vector for both A_i and B_i must lie within distance $e^{-n\nu}$ of $(1,0)$ for some sufficiently large ν.

We now estimate n_1, i.e. we study how long it takes a vector v near $(0,1)$ to leave $B_{\delta_0}((0,1))$. We claim that if ν is sufficiently small then for sufficiently large n, any vector v that starts

$e^{-n\nu}$ away from $(0,1)$ will exit $B_{\delta_0}((0,1))$ after at most $(2\nu/\lambda)n$ iterates. To this end consider

$$\begin{aligned}d(g^i_\epsilon(v),(0,1)) &\geq d(g^i_\epsilon(v), g_{i,\epsilon}(0,1)) - d(g_{i,\epsilon}((0,1)),(0,1))\\ &\geq e^{.9\hat{\lambda}}d(g^{i-1}_\epsilon(v),(0,1))\left(1-\frac{C_1e^{-n\alpha}e^{i(\alpha+4\epsilon)}}{e^{.9\hat{\lambda}}d(g^{i-1}_\epsilon(v),(0,1))}\right)\end{aligned}$$

As long as $i\leq (1/3)n$, $\epsilon<\alpha/100$, and n is sufficiently large,

$$C_1e^{-n\alpha}e^{i(\alpha+4\epsilon)}\leq e^{-\frac{\alpha}{2}n}. \tag{10.5}$$

Thus if $\nu<\alpha/2$, then for sufficiently large n, if $d(g^{i-1}_\epsilon(v),(0,1))\geq e^{-n\nu}$, then

$$\left(1-\frac{C_1e^{-n\alpha}e^{i(\alpha+4\epsilon)}}{e^{.9\hat{\lambda}}d(g^{i-1}_\epsilon(v),(0,1))}\right)\geq e^{-.1\hat{\lambda}}. \tag{10.6}$$

From the above, we see that as long as n is sufficiently large, $i\leq n/3$, and the trajectory of v has not left the $B_{\delta_0}((0,1))$ after i, iterates, then

$$d(g^i_\epsilon(v),(0,1))\geq e^{.8\hat{\lambda}}d(g^{i-1}_\epsilon((0,1)),(0,1)). \tag{10.7}$$

Proceeding iteratively, we see that after i iterations, assuming $i\leq n/3$ and that the trajectory of v has not left $B_{\delta_0}((0,1))$,

$$d(g^i_\epsilon(v),(0,1))\geq e^{.8\hat{\lambda}i}d(v,(0,1)). \tag{10.8}$$

In particular, if $g^i_\epsilon(v)$ has not left $B_{\delta_0}((0,1))$ after $(2\nu/.8\hat{\lambda})n$, iterates then we would have that $d(g^i_\epsilon(v),(0,1))\geq e^{\nu n}$, which is absurd.

Thus as long as $\epsilon<\alpha/10$ it follows for sufficiently large n that $g^i_\epsilon(v)$ exits $B_{\delta_0}((0,1))$ after at most $\frac{2\nu}{.8\hat{\lambda}}n$ steps. Moreover, by Lemma 10.2, it enters the neighborhood $B_{\delta_0}((1,0))$ after an additional N_0 iterates. Thus for sufficiently large n, $n_1\leq \frac{2\nu}{.79\hat{\lambda}}n$.

We now estimate n_2. In the Lyapunov charts, B_i is a perturbation of A_i of size $e^{-\alpha(n-i)}e^{4\epsilon i}$. Lemma 10.2 ceases to hold when the size of the perturbation is size $O_{\epsilon_0}(1)$. This will occur when $e^{-\alpha(n-i)}e^{4\epsilon i}$ is order 1, which happens when $i\approx\frac{\alpha}{\alpha+4\epsilon}n$. If ϵ is sufficiently small relative to α, then $\alpha/(\alpha+4\epsilon)\geq 1-8\epsilon/\alpha$. Hence by picking some N_2' depending only on ϵ_0,δ_0 and C_1, we see that n_2 may be chosen to be the smallest number satisfying $n_2\geq(1-8\frac{\epsilon}{\alpha})n-N_2'$. Hence for sufficiently large n we can take the bound $n_2\geq(1-9\frac{\epsilon}{\alpha})n$.

Thus between times n_1 and n_2 there are at least $(1-9\frac{\epsilon}{\alpha}-\frac{2\nu}{.79\hat{\lambda}})n$ iterates. As long as n is sufficiently large and

$$\left(1-9\frac{\epsilon}{\alpha}-\frac{2\nu}{.79\hat{\lambda}}\right)>\frac{1}{2} \tag{10.9}$$

which we can certainly arrange if we take ϵ,ν sufficiently small, we see that there are at least $n/2$ iterates between n_1 and n_2.

We now estimate $\|B^nv\|$. Let us first consider the norm $\|B^{n_2}v\|$ by estimating in the Lyapunov metric. Let v^i equal $g^i_\epsilon(v)$. Then, for $i\leq n_1$ and n sufficiently large, using (10.2) and (10.5) and the inequality $e^X+Y\leq e^{X+Y}$, valid for $X,Y\geq 0$, we obtain
$\|B_i\|'\leq e^{\Delta}+e^{-\frac{\alpha}{2}n}\leq e^{\Delta+e^{-\frac{\alpha}{2}n}}$. Taking logarithms we get $\ln\|B_i\|'\leq\Delta+e^{-\frac{\alpha}{2}n}$. Thus,

$$\ln\|B^{n_2}v\|'\geq\sum_{i=n_1}^{n_2}\Phi^i_\epsilon(v^i)+\sum_{i=0}^{n_1}\Phi^i_\epsilon(v^i)\geq(n_2-n_1).8\hat{\lambda}-\left(n\frac{2\nu}{.79\hat{\lambda}}\right)(\Delta+e^{-\frac{\alpha}{2}n}).$$

This is the amount of growth in the Lyapunov coordinates. For the original metric, by Lemma A.1(3) this implies from our bounds on n_1 and n_2, that

$$\ln \|B^{n_2} v\| \geq (n_2 - n_1).8\hat{\lambda} - n\frac{2\nu}{.79\hat{\lambda}}(\Delta + e^{-\frac{\alpha}{2}n}) - 4n_2\epsilon. \tag{10.10}$$

Since $\ln\|B_i\| \leq \Delta$, and because $n - n_2 \leq (9\epsilon/\alpha)n$, and $n_2 - n_1 > n/2$, we see that

$$\ln\|B^n\| \geq .4\hat{\lambda}n - n\frac{2\nu}{.79\hat{\lambda}}(\Delta + e^{-\frac{\alpha}{2}n}) - n_2 4\epsilon - \frac{9\Delta\epsilon n}{\alpha}. \tag{10.11}$$

So, we may conclude if

$$.4\hat{\lambda} - \frac{2\nu}{.79\hat{\lambda}} - 9\Delta\frac{\epsilon}{\alpha} > 0, \tag{10.12}$$

which is certainly true as long as ϵ and ν are sufficiently small relative to α, λ', and Δ. □

Remark 10.6. *Note that the proof of the previous claim shows something more precise: letting v_A^s, v_A^u be the most contracted and expanded direction of A^n, in the Lyapunov charts both v_A^s and v_B^s lie within the neighborhood $B_{\delta_0}((0,1))$ and v_A^u and v_B^u both lie within the neighborhood $B_{\delta_0}((1,0))$ of where the conclusions of Lemma 10.2 hold.*

Now that we have located where v^s, and hence v^u lies, we check that the norm of B^i is subtempered.

Lemma 10.7. *For any $\epsilon_0 > 0$, suppose that we have a sequence of matrices as above. Then there exists $k(C, \lambda, \epsilon, \alpha, \Delta)$ such that $k\epsilon < \epsilon_0$ and the norm $\|B^i\|$ is $(C, \lambda - k\epsilon, 4k\epsilon)$ subtempered.*

Proof. From Lemma 10.2, we see that if $v^u \in (E_0^s)^\perp$, then v^u lies in $B_{\delta_0}((1,0))$. Given any $\beta_0 > 0$ and n sufficiently large, any vector v in this neighborhood satisfies that for $i < n_2$,

$$\Phi_\epsilon^i(v) \geq (1 - \beta_0)\lambda. \tag{10.13}$$

Thus we see that along the trajectory from time 1 to n_2 that every vector that begins in $B_{\delta_0}(0,1)$ is $(C, (1-\beta_0)\lambda, 0)$-subtempered for the sequence of matrices B_i viewed in Lyapunov charts. Take β_0 such that $(1-\beta_0)\lambda > (\lambda + \lambda')/2$.

With respect to the reference metric, such a sequence is $(C, (1-\beta_0)\lambda, 4\epsilon)$-tempered due to Lemma A.1(3). This gives temperedness up to time n_2.

Recall that Lemma 10.4 says that if we extend the sequence by m matrices where

$$m < \Delta^{-1}(nk\epsilon - C),$$

then the result will be $(C, (1-\beta_0)\lambda - k\epsilon, 4k\epsilon)$-tempered. In our case because $n_2 \geq (1 - \frac{9\epsilon}{\alpha})n$, we would like to append $\frac{9\epsilon}{\alpha}$ matrices of norm at most e^Δ and have the resulting sequence still be tempered. So, we need that

$$\frac{9\epsilon}{\alpha}n < \Delta^{-1}(nk\epsilon - C) \tag{10.14}$$

For sufficiently large n, this holds as long as $k\epsilon\Delta^{-1} > 9\epsilon/\alpha$, that is, $k > 9\Delta/\alpha$. Taking ϵ sufficiently small we can arrange that $k\epsilon < \epsilon_0$. In particular choosing β_0 sufficiently small, we can have that $(1-\beta_0)\lambda - k\epsilon \geq \lambda'$, so the needed conclusion holds. □

The first and second conclusions of Proposition 10.3 for sufficiently large n are now immediate from the two lemmas once we apply Proposition 4.6, which constructs a splitting for a norm subtempered sequence.

We now turn to the proof of the third conclusion of the proposition. We need additional estimates.

We let n_2 be as above; it is the point past which the estimate in Lemma 10.2 ceases to hold. Note that there exists β_1 such that $\|A_i - B_i\|' \le e^{-\beta_1(n_2-i)}$ where $\|\cdot\|'$, denotes the Lyapunov metric. Also, recall that from our choice of n_2, that on a neighborhood of $(1,0)$ of size δ_0 that B_i contracts distances by a factor of $e^{-.9\hat{\lambda}}$.

Claim 10.8. *There exists $\beta_2 > 0$ such that if v^u is the unstable vector for the A_i, then*

$$d'(A^i v^u, B^i v^u) \le K e^{-\beta_2(n_2-i)},$$

where $d'(u_1,u_2) = \left\| \frac{u_1}{\|u_1\|'} - \frac{u_2}{\|u_2\|'} \right\|'$ is the metric on $\mathbb{R}\mathrm{P}^1$ with corresponding to the Lyapunov metric.

Proof. Recall that in the Lyapunov coordinates, we have $A^i((1,0)) = (1,0)$. Further, from the previous Lemma, v_A is within distance $e^{-n\nu}$ of $(1,0)$. Consequently, we begin by suppose that v is a vector with $d'(v,(1,0)) < e^{-n\nu}$ and then seeing how this vector shadows the trajectory of $(1,0)$. Then as both v_A and v_B are vectors satisfying this property, the needed conclusion follows by the triangle inequality.

This can be seen inductively because, by that lemma[1],

$$\begin{aligned} d'(B^i(v),(1,0)) &\le d'(B^i(v), B_i(1,0)) + d'(B_i(1,0),(1,0)) \\ &\le e^{-.9\hat{\lambda}} d'(B^{i-1}(v),(0,1)) + C\|A_i - B_i\|'. \end{aligned}$$

We may continue inductively as long as $B^i v$ still lies in the neighborhood B_{δ_0}. For such i before this point, the form of the estimate that we obtain is:

$$d'(B^i(v),(1,0)) \le e^{-n\nu} e^{-.9i\hat{\lambda}} + C\sum_{j=1}^{i} e^{-.9\hat{\lambda}(i-j)} \|A_j - B_j\|' \le C' e^{-\beta_2(n_2-i)}.$$

Note that as this estimate is growing exponentially quickly that the difference between the index i where it first exceeds δ_0 and n_2 is of size at most $\ln(C')/\beta_2$, which is constant. Hence by possibly adjusting the constant, the needed result follows.

To conclude we apply apply the triangle inequality to the corresponding estimates on $d'(B^i(v),(1,0)$ and $d'(A^i(v),(1,0))$ $\square$

Before proceeding further, we record an additional quantitative estimate about the norms of the maps considered in Lemma 10.2.

Claim 10.9. *For a matrix A as in Lemma 10.2, for all $\sigma > 0$, there exists $\epsilon_1 > 0$ such that if $E\colon \mathbb{R}^2 \to \mathbb{R}^2$ is a matrix of norm $\epsilon \le \epsilon_1$, then if $v \in \mathbb{R}\mathrm{P}^1$ with $d((1,0),v) \le \epsilon_1$:*

(1) $|\Phi_A(v) - \Phi_{A+E}(v)| \le \|E\|$.
(2) $|\Phi_A(v) - \Phi_A((1,0))| \le (\sigma/2)\ln\|A\|$.

Proof. This claims follows easily because we are restricting to a neighborhood in $\mathbb{R}\mathrm{P}^1$ where A has large norm. Note that if v is a unit vector v and ϵ_1 is sufficiently small then $\|Av\|$ and $\|(A+E)v\|$ are both greater than 1, hence as ln is 1-Lipschitz on $[1,\infty)$, so the first claim follows. The second claim is straightforward because by assumption $A = \mathrm{diag}(\sigma_1,\sigma_2)$. $\square$

[1] Note that Lemma 10.2 applies to the Lyapunov metric since the eigenvalues of the matrices A_i are uniformly bounded in both in both original and Lyapunov coordinates.

Similar to before we have the map $\Phi'_A(v) = \ln(\|Av\|'/\|v\|')$ on $\mathbb{R}\,\mathrm{P}^1$; note that this measures the expansion of vectors with respect to the Lyapunov metric. By possibly decreasing the constants in the statement of Lemma 10.2, we can arrange that the conclusions of Claim 10.9 hold as well for all $i \le n_2$. (Both statements hold with respect to the Lyapunov metric, see footnote 1). We record two facts that follow from Claim 10.8 along with the estimate $\|A_i - B_i\|' \le e^{-\beta_1(n_2-i)}$:

$$\left|\Phi'_{A_i}(A^{i-1}v^u) - \Phi'_{A_i}(B^{i-1}v^u)\right| \le (\sigma/2)\ln\|A_i\|'$$

$$\left|\Phi'_{A_i}(B^{i-1}v^u) - \Phi'_{B_i}(B^{i-1}v^u)\right| \le \|B_i\|' \le e^{-\beta_1(n_2-i)}.$$

Using these claims, we now estimate $\|B^{n_2}v^u\|'$:

$$\left|\|B^{n_2}v^u\|' - \|A^{n_2}v^u\|'\right| = \left|\sum_{i=1}^{n_2} \Phi'_{A_i}(A^{i-1}v^u) - \Phi'_{B_i}(B^{i-1}v^u)\right|$$

$$\le \sum_{i=1}^{n_2} \left|\Phi'_{A_i}(A^{i-1}v^u) - \Phi'_{A_i}(B^{i-1}v^u)\right| + \left|\Phi'_{A_i}(B^{i-1}v^u) - \Phi'_{B_i}(B^{i-1}v^u)\right|$$

$$\le \sum_{i=1}^{n_2} \left((\sigma/2)\ln\|A_i\|' + e^{-\beta_1(n_2-i)}\right)$$

Thus we see that $\ln\|B^{n_2}\|' \ge (1-\sigma/2)\ln\|A^{n_2}\|' - C_1$. Using this we now estimate the norm of $\|B^n\|$. As the norm of all B_i and A_i are uniformly bounded by e^{Δ} by assumption, it follows that $|\ln\|A^n\| - \ln\|A^{n_2}\|| \le (n-n_2)\Delta$. Thus,

$$\begin{aligned}\ln\|B^n\| &\ge \ln\|B^{n_2}\| - (n-n_2)\Delta\\ &\ge (1-\sigma/2)\ln\|A^{n_2}\| - 4n_2\epsilon - C_1 - (n-n_2)\Delta\\ &\ge (1-\sigma/2)\ln\|A^n\| - (\sigma/2)(n-n_2)\Delta - 4n_2\epsilon - C_1 - (n-n_2)\Delta.\end{aligned}$$

By subtemperedness $\ln\|A^n\| \ge \lambda n - C_2$ for some C_2, and hence if ε is small enough compared to λ and σ, then as $n - n_2 = O(\epsilon)$ the estimate of part (3) of Lemma 10.2 holds. □

Proposition 10.3 implies that nearby points have close splittings so that the blocks where a tempered splitting fails to exist are not too small.

10.3. **Cushion of nearby points.** In this subsection, we prove a refinement of the estimate from the previous subsection. Recall Definition 4.9. We show that points with very close trajectories have cushion that differs by $O(1)$. This will be used later because it shows that if a short curve has a single point with bad cushion, then all of these points have bad cushion.

Proposition 10.10. *Fix (C_0,λ), $\Lambda > 0$, $\sigma > 0$, $\varpi > 0$, then for all sufficiently small $\epsilon > 0$ there exists N and D such that the following holds. Suppose that $(A_i)_{1\le i\le n}$ and $(B_i)_{1\le i\le n}$ are sequences of matrices in $\mathrm{SL}(2,\mathbb{R})$ with norm at most e^{Λ} that are (C_0,λ,ϵ)-tempered such that $\|A_i - B_i\| \le C_1 e^{-\sigma(n-i)}e^{-n\varpi}$. Let $U(A)$ and $U(B)$ denote the cushion of A and B. Then*

$$|U(A) - U(B)| \le D.$$

Proof. In view of the definition of the cushion, it suffices to prove that there exists $D > 0$ such that for two such sequences and $1 \le k \le n$, $\left|\ln\|A^k\| - \ln\|B^k\|\right| \le D$. This will follow from the claim below, which gives an exponential shadowing for the most expanded directions of A^k and B^k

Claim 10.11. *There exists N_1, D_1 such that as long as $n \geq N$, There exists $\beta_2(\lambda, \epsilon, \sigma)$ and $K(C_0, \lambda, \Lambda, \sigma, \varpi, \epsilon)$ such that for any $N \leq k \leq n$ the following holds. If v_k is the unstable vector for the A^k, then for $i \leq k$,*

$$d(A^i v_k, B^i v_k) \leq K(e^{-\beta_2(k+i)} + e^{-n\varpi/2}).$$

Proof. This essentially follows due to an enhancement of the argument surrounding Claim 10.8, which we can improve due to the stronger assumptions of the present claim.

As before, we work in Lyapunov charts, and estimate the distance that a vector near $(1,0)$ can drift away from it. Comparing with (10.2), when we look in the Lyapunov charts adapted to the sequence $A_1, \ldots, A_k$, we now have that

$$A_i = \begin{bmatrix} \sigma_{i,1} & 0 \\ 0 & \sigma_{2,i} \end{bmatrix} = B_i + O_{C,\lambda,\Lambda}(e^{-\sigma(n-i)} e^{-\varpi n} e^{4\epsilon i}).$$

Hence Lemma 10.2 holds for all $1 \leq i \leq n$, i.e. for the entire sequence, as long as ϵ is sufficiently small relative to ϖ. Note that this implies that there exists some C' such that $\|A_j - B_j\|' \leq C' e^{-\sigma(n-i)} e^{-\varpi n} e^{4\epsilon i}$.

We now do an induction similar to that in Claim 10.8. Denote

$$d_{n,k,i} = e^{-k\nu} e^{-.9\hat{\lambda} i} + \bar{C} e^{-n(2/3)\varpi} e^{-\sigma(n-k)},$$

where $\bar{C}$ is a large constant that will be chosen below. From Lemma 10.2, we can take δ_0 so small that any vector making angle less than δ_0 with $(1,0)$ is contracted by at least $e^{-.9\hat{\lambda}}$. Take N so large that for all $N \leq k \leq n$ we have that $d_{n,k,0} \leq \delta_0$ and hence also for all $i \leq N$, $d_{n,k,i} \leq \delta_0$. We now verify by induction on i that if we start with a vector v such that $d'(v, (1,0)) \leq e^{-k\nu}$, then for all $i \leq k$, $d'(B^i(v), (1,0)) \leq d_{n,k,i}$. Indeed

$$\begin{aligned} d'(B^i(v), (1,0)) &\leq e^{-.9\hat{\lambda}} d_{n,k,i-1} + \|A_i - B_i\|' \\ &\leq e^{-k\nu} e^{-.9 i\hat{\lambda}} + e^{-.9\hat{\lambda}} \bar{C} e^{-n(2/3)\varpi} e^{-\sigma(n-k)} + C' e^{-\sigma(n-i)} e^{-\varpi n} e^{4\epsilon i}. \end{aligned}$$

As long as $\bar{C}$ is sufficiently large and ϵ is sufficiently small relative to ϖ, it then follows that:

$$d'(B^i(v), (1,0)) \leq e^{-k\nu} e^{-.9\hat{\lambda} i} + \bar{C} e^{-n(2/3)\varpi} e^{-\sigma(n-k)} \leq d_{n,k,i}.$$

Thus for $1 \leq i \leq k$,

$$d'(B^i v, A^i(1,0)) \leq C_1(e^{-k\nu} e^{-.9\hat{\lambda} i} + \bar{C} e^{-n(2/3)\varpi} e^{-\sigma(n-k)}).$$

Lemma A.8, which compares distance on S^1 for different metrics, implies that as long as ϵ is sufficiently small relative to λ and σ, then respect to the reference metric on $\mathbb{R}\mathrm{P}^1$ that there exists C_2 such that

$$d(B^i v, A^i((1,0))) \leq C_2(e^{-k\nu} e^{-.45\hat{\lambda} i} + \bar{C} e^{-n\varpi/2} e^{-\sigma(n-k)}).$$

The above estimate holds for any vector v at distance $e^{-k\nu}$ from $(1,0)$.

In particular, from Lemma 10.5 whose weaker hypotheses $(A_i)_{1 \leq i \leq k}$ and $(B_i)_{1 \leq i \leq k}$ satisfy, we see that v_A^k and v_B^k are both within $e^{-k\nu}$ distance of $(1,0)$ in the Lyapunov charts as long as $k \geq N_2$ for some N_2. Thus by specializing to these vectors and applying the triangle inequality, we find that $d(B^i v_A^k, A^i v_A^k) \leq C_3(e^{-k\nu} e^{-.45\hat{\lambda} i} + e^{-n\varpi/2})$, which is the desired claim. ∎

Because the norm of all the matrices we are considering is uniformly bounded by e^{Λ}, the estimate in Claim 10.11 gives that for $k > N$,

$$\left|\ln\|A^k\| - \ln\|B^k\|\right| \le \sum_{i=1}^{k} Ke^{-\beta_2(k+i)} + e^{-n\varpi/2} \le K'$$

for some fixed K'. Note that this gives the conclusion of the lemma about cushioning for all indices greater than N. For those less than N, since there are only finitely many such words and the norms of matrices are bounded, we can accommodate them by increasing the constant in the conclusion of the theorem. □

10.4. **Scale selection proposition.** Given two nearby standard pairs, we can attempt to "couple" them using the fake stable manifolds. For this we need more quantitative estimates on how close and smooth standard pairs need to be so that we can couple a significant proportion of them. For example, if they are too far apart then a fake stable leaf may not reach from one to the next. Proposition 10.12 below are mostly a summary of results appearing elsewhere in the paper. Note that the first parts of the proposition are statements about temperedness and splittings on uniformly large balls in M. Part (4) shows that for fixed C_0 if we consider sufficiently small (C_0, δ, υ)-configurations that on balls of radius $O(\delta)$ that transversality to the contracting direction and temperedness of the splitting imply that the holonomies between the curves in a configuration exist and converge exponentially fast.

Below we say that a curve and a cone field are θ_0-transverse if the smallest angle they make is at least θ_0. Also, see Definition B.11 in the appendix for the definition of $(C, \lambda, \epsilon, \mathcal{C})$-tempered, which means (C, λ, ϵ)-tempered plus the additional condition that the stable direction lies in the cone $\mathcal{C}$.

Proposition 10.12. *Suppose that $(f_1, \ldots, f_m)$ is an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$ with M a closed surface. There exists $\lambda > 0$ such that for any $0 < \lambda' < \lambda$, $0 < \sigma$ there exists $0 < \epsilon_0, \tau < 1$ such that for any $0 < \epsilon < \epsilon' < \epsilon_0$ there exist $\delta_0, \delta_1, \theta, b_0, C, C', C'', \theta_0, \eta > 0$ and $N \in \mathbb{N}$ such that: for any $x \in M$, $i \in \{1, 2, 3\}$, there are three nested cone fields $\mathcal{C}^i_\theta \subset \mathcal{C}^i_{2\theta} \subset \mathcal{C}^i_{3\theta}$ of angles θ, 2θ, 3θ, respectively defined on $B_{\delta_0}(x)$ by parallel transport from a cone at x. Further, the $\mathcal{C}^i_{3\theta}$ are uniformly transverse on $B_{\delta_0}(x)$. These conefields satisfy the following properties for words ω, where probabilities below are with respect to the Bernoulli measure μ on Σ.*

1. *(Positive probability of tangency to $\mathcal{C}^i_\theta$) For any point $y \in B_{\delta_0}(x)$ and any $i \in \{1, 2, 3\}$, the probability that $D_x f^n_\omega$ is $(C, \lambda, \epsilon, \mathcal{C}^i_\theta)$-tempered for all $n \ge N$ is at least $b_0 > 0$.*
2. *(Nearby points are also tempered) For any curve γ, if $x \in \gamma$ is $(C, \lambda, \epsilon, \mathcal{C}^i_\theta)$-tempered at time n and $y \in \gamma$ is a point with $d_\gamma(x, y) \le \|D_x f^n_\omega\|^{-(1+\sigma)}$, then y is $(C', \lambda', \epsilon', \mathcal{C}^i_{2\theta})$-tempered at time n and*

$$\|D_y f^n_\omega\| \ge \|D_x f^n_\omega\|^{1-\sigma}. \tag{10.15}$$

3. *(Existence of fake stable manifolds) For any $(C', \lambda', \epsilon', \mathcal{C}^i_{2\theta})$-tempered point $y \in B_{\delta_0}(x)$ at time $n \ge N$, the fake stable curve $W^s_{n,\delta_1}(\omega, y)$ exists, has length at least δ_1, has C^2 norm at most C'', and is tangent to $\mathcal{C}^i_{3\theta}$ on $B_{\delta_0}(x)$.*
4. *(There exists a well configured neighborhood) For any C_0, there exists $\delta \in (0, 1)$, $a_0, a_1, D_1, D_2 > 0$ and $N_1 \in \mathbb{N}$ such that for all $0 < \delta' < \delta$, and any $\upsilon \le \delta'\tau$, the following holds for any (C_0, δ', υ)-configuration $(\hat\gamma_1, \hat\gamma_2)$. There exists $x \in M$ and $i \in \{1, 2, 3\}$ such that γ_1 and γ_2 are uniformly θ_0-transverse to $\mathcal{C}^i_{3\theta}$ on $B_{\delta_0}(x)$. We let $B_{2\nu}(x)$ be a ball that demonstrates that $\hat\gamma_1$, and $\hat\gamma_2$ are in a (C_0, δ', υ)-configuration, i.e. it contains points of γ_1 and γ_2 that*

are distance at least υ from the boundary of those curves. We maintain this choice of x and i in the following lettered items:

(a) (Fake stable manifolds tangent to $\mathcal{C}^i_{3\theta}$ are transverse to pairs) If $y \in B_{2\upsilon}(x)$ is as in item (3) above, then $W^s_{n,\delta_1}(\omega, y)$ intersects both γ_1 and γ_2 and the points of intersection are both θ_0-transverse, i.e. both γ_1 and γ_2 make an angle at least θ_0 with $W^s_{n,\delta_1}(\omega, y)$.

(b) (Lower bound on derivative of the holonomy) For $n \geq N_1$, if $B \subseteq \gamma_1 \cap B_{2\nu}(x)$ is a subset of γ_1 consisting of $(C', \lambda', \epsilon', \mathcal{C}^i_{2\theta})$-tempered points at time n, then $H^s_n(B) \subseteq \gamma_2$ has length at least $D_1 \operatorname{len}(B)$. Further, as long as $\hat{\gamma}_1$ and $\hat{\gamma}_2$ have equal mass and are at most $4\delta'$-long, there are a pair of connected components of $\hat{\gamma}_1 \cap B_{2\upsilon}(x)$ and $\hat{\gamma}_2 \cap B_{2\upsilon}(x)$ each containing at least a_1 proportion of the mass of $\hat{\gamma}_1$ and $\hat{\gamma}_2$ such that if B is as above and lies in this set, then

$$a_0 (H^s_n)_* \rho^1|_B \leq \rho^2|_{H^s_n(B)}.$$

(c) (Fluctuations in the holonomies) For any (C_0, δ', υ)-configured pair (γ_1, γ_2), if $z \in B_{2\upsilon}(x)$ is a $(C', \lambda', \epsilon', \mathcal{C}^i_{2\theta})$ tempered point at times $n, n-1 \geq N_1$ and y is any point with $d_{\gamma_1}(x, y) \leq \|D_x f^n_\omega\|^{-(1+\sigma)}$, then

$$d_{\gamma_2}(H^s_n(y), H^s_{n-1}(y)) \leq e^{-1.99 \ln \|D_x f^n_\omega\|}. \tag{10.16}$$

Further, for $n \geq N_1$ the rate of convergence of the Jacobians is exponentially fast

$$\left|\operatorname{Jac} H^s_n - \operatorname{Jac} H^s_{n-1}\right| \leq e^{-\eta n}. \tag{10.17}$$

(d) (Log-α-Hölder control of Jacobian) If $B \subseteq \gamma_1 \cap B_{2\upsilon}(x)$ is an open set comprised of $(C', \lambda', \epsilon', \mathcal{C}^i_\theta)$-tempered points at time n, then

$$|\log \operatorname{Jac} H^s_n(x) - \log \operatorname{Jac} H^s_n(y)| \leq D_2 d_{\gamma_1}(x, y)^\alpha. \tag{10.18}$$

Proof. The main non-trivial input to this proposition is the definition of the cones. After they are chosen correctly, the remaining statements follow in a straightforward manner from facts about the fake stable manifolds proven elsewhere.

For any point $x \in M$, we let ν_x denote the distribution of the true stable directions E^s at the point x, which is a measure on $\mathbb{R}\mathrm{P}^1_x$, the projectivization of $T_x M$. As ν_x is non-atomic, we can find three disjoint intervals I_1, I_2, I_3 of width θ that are each separated by angle at least 4θ for some angle $\theta > 0$ and such that $\nu_x(I_1), \nu_x(I_2), \nu_x(I_3)$ are each positive. We then use these intervals to define nested cones $\mathcal{C}^i_{\theta/2}(x) \subset \mathcal{C}^i_\theta(x) \subset \mathcal{C}^i_{2\theta}(x) \subset \mathcal{C}^i_{3\theta}(x)$ at x for $i \in \{1, 2, 3\}$. Due to the continuity of ν_x from Proposition B.4, we see that if we parallel translate I_1, I_2, I_3 to form cone fields $\mathcal{C}_1, \mathcal{C}_2, \mathcal{C}_3$ over a ball $B(x)$ around x, then we similarly have that $\nu_y(\mathcal{C}^i_{\theta/2})$ is uniformly positive for all $y \in B(x)$. All these properties are uniform, so we can do this for any $x \in M$ and obtain a neighborhood of uniform size, with uniform lower bound on $\nu_y(\mathcal{C}^i_{\theta/2})$ over all these neighborhoods.

We now verify item (1). There exist $\lambda, \epsilon > 0$ such that for any $y \in M$ and almost every ω, $D_y f^n_\omega$ is $(C(\omega), \lambda, \epsilon)$-tempered for some $C(\omega)$. Further, by Proposition 4.7 we have a uniform estimate on the tail on $C(\omega)$ independent of the point y. Thus by choosing C_1 sufficiently large for any $y \in B(x)$ and $1 \leq i \leq 3$, with probability at least b_0, $D_y f^n_\omega$ is (C_1, λ, ϵ)-subtempered and $E^s_n(\omega, y) \in \mathcal{C}^i_{\theta/2}(y)$ for all $n \geq N$. By Proposition 4.6, there exists $N_0 \in N$ such that for any (C_1, λ, ϵ)-subtempered trajectory of length $n \geq N_0$, then for all $n \geq N_0$, $\angle(E^s_n(\omega, y), E^s) < \theta/4$ and so $E^s_n \in \mathcal{C}^i_\theta$. This gives us the uniformly positive probability of at least $b_0 > 0$.

Item (2) is immediate from Proposition 10.3.

Item (3), which states the existence of the fake stable manifolds for $(C', \lambda', \epsilon', \mathcal{C}^i_{2\theta})$-tempered points, follows from Proposition B.10 (possibly after decreasing δ).

We now verify item (4), which has many subparts. The statement in the initial part follows by making a judicious choice of x as well as the particular cone $\mathcal{C}^i_{3\theta}$ on $B_{\delta_0}(x)$ that the fake stable manifolds will be tangent to. Because $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are a (C_0, δ', υ)-configuration then there exists a pair of points $x \in \gamma_1$ and $y \in \gamma_2$ with $d(x, y) < \upsilon$. We choose to work on the neighborhood $B_{\delta_0}(x)$. We then must show that we can pick one of the cones $\mathcal{C}^i_{3\theta}$ that is uniformly transverse to γ_1 and γ_2 on $B_{\delta_0}(x)$. Let $\mathcal{K}_1$ be a small cone around $\gamma_1'(x)$ and $\mathcal{K}_2$ be a small cone around $\gamma_2'(y)$. We can extend both cones to the whole of $B_{\delta_0}(x)$ by parallel transport. Since there are three cones, we let $i \in \{1, 2, 3\}$ be an index such that the cone $\mathcal{C}^i_{3\theta}$ is transverse to both $\mathcal{K}_1$ and $\mathcal{K}_2$. We let θ_0 be a lower bound on the angle that γ_1, γ_2 make with $\mathcal{C}^i_{3\theta}$ and note that, as before, that θ_0 is uniform as it only relies on knowing C_0, δ'. We now proceed to checking the lettered items that follow.

Item (4a) says that the fake stable manifolds of $(C', \lambda', \epsilon', \mathcal{C}^i_{2\theta})$-tempered points are θ_0-transverse to γ_1, γ_2 and intersect them. This follows from Proposition B.10 because by choice of our constants, for such a tempered point y, it follows that $W^s_n(\omega, y)$ is tangent to $\mathcal{C}^i_{3\theta}$, and the uniform transversality follows from our control on the C^2 norm of $W^s_n(\omega, y)$ and the Hölder continuity of the most contracting subspace E^s_n. Further, the fact that we only need the curves to be at most $\upsilon = \tau\delta'$ apart from each other, with τ depending only on $\theta_0, \lambda', \lambda$ is clear from the uniform C^2 bound on the norm of the fake stable manifolds $W^s_n(\omega, y)$ from item (3). Item (4a) follows because as long as δ is sufficiently small compared with C_0, the tangent direction to γ_i is close to constant on a segment of length δ.

The first part of item (4b) saying that there is a lower bound on the derivative of the holonomies follows from Proposition B.13.

The next claim is that restricted to a segment in B_{δ_0}, γ_1 and γ_2 have a positive proportion of their mass there. This follows due to the log-Hölder regularity of ρ^1 and ρ^2 as long as δ is sufficiently small. Due to the boundedness of the Jacobian, the log-Hölderness of the densities and them both having a positive amount of their mass on $B_{\delta_0}(x)$, it additionally follows that there exists such a uniform constant a_0 as stated in item (4b).

Item (4c) is immediate from the statement of Proposition B.12.

Finally, item (4d), which concerns the fluctuations in the Jacobian of H^s_n, follows from Proposition B.13. □

10.5. **Proof of Inductive Local Coupling Lemma.** We are now ready to prove the inductive local coupling lemma.

First we prove a result that does not make any assertions about the quantity of points on the curve γ_1 that have a tempered splitting. It just shows that given an infinite trajectory $\omega \in \Sigma$, we may use this trajectory to define a fake coupling in the sense of Definition 10.1 at all future times.

Lemma 10.13. *(Inductive Coupling Lemma.) Let $(f_1, \ldots, f_m)$ be an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$ for M a closed surface.*

For any $C_0 > 0$ let $\lambda, \lambda', \epsilon_0, \tau, \epsilon, \epsilon'$, etc. be a valid choice of constants in the first paragraph of Proposition 10.12 and $\delta, \delta', \upsilon$, etc., be a valid choice of constants in part (4) of that proposition. Then there exist $b_1, \hat{\eta}, \Lambda > 0$ such that for any (C_0, δ', υ)-configuration $(\hat{\gamma}_1, \hat{\gamma}_2)$ the conclusions of Proposition 10.12 apply and the following holds. If $x \in M$ and $B_{\delta_0}(x)$ is the neighborhood where the statements from Proposition 10.12(4) hold, then we can construct a $(b_1, \hat{\eta})$-fake couplings out of $(\hat{\gamma}_1, \hat{\gamma}_2)$: For each $\omega \in \Sigma$ there exists a decreasing sequence of pairs of standard *subfamilies $P_n^1 \subseteq \hat{\gamma}_1$ and $P_n^2 \subseteq \hat{\gamma}_2$ that are $(b_1, \hat{\eta})$-fake coupled at each time $n \geq N_1$. Further, for $n \geq N_1$ and $i \in \{1,2\}$ $P_n^i \setminus P_{n+1}^i$ are $n\Lambda$-good standard families.*

These sequences of standard families are decreasing and converge to measures P_∞^1 and P_∞^2. Further, for such a fake coupling we also have the true stable holonomies H_∞^s and these satisfy $(H_\infty^s)_* P_\infty^1 = P_\infty^2$.

Proof of Lemma 10.13. We divide the proof into several steps. In Step 0, we introduce the constants that will be used later in the proof; naturally we will also make use of many constants from Proposition 10.12, which is essentially the setup for this lemma. Then in the following steps we give an iterative procedure showing how one may construct a new fake coupled pair out of an old one. By iterating that procedure, we then obtain the result.

Step 0: Introduction of constants. At this step we introduce some of constants that will be used in the proof. Most of these constants will be chosen when they appear in the proof.

(1) First, we let $\lambda, \lambda', \epsilon', D_1$, etc., be the constants from the statement of Proposition 10.12. For the given $\hat{\gamma}_1$ and $\hat{\gamma}_2$ we let $B_{\delta_0}(x)$ be a neighborhood so that the conclusions of part 4 of that proposition apply. We will simply write $\mathcal{C}_\theta$ rather than $\mathcal{C}_\theta^i$ below for the cones defined on $B_{\delta_0}(x)$ such that $\hat{\gamma}_1$ and $\hat{\gamma}_2$ have segments that are both uniformly θ_0-transverse to $\mathcal{C}_{3\theta}^i$ on $B_{\delta_0}(x)$. We let $\Lambda_{\max}$ be sufficiently large so that $\|D_x f_i\| \leq e^{\Lambda_{\max}}$ for all $x \in M$ and $1 \leq i \leq m$.
(2) Further, in the application of Proposition 10.12 we will insist that δ is so small that for any C_0-good curve with density ρ on a ball of size δ, the log-Hölder condition on ρ implies that $1/2 < \rho(y)/\rho(x) < 2$ on this ball.
(3) Below, we have certain estimates that will only hold as long as n is sufficiently large. We will have some cutoffs N_1, N_2 that we define in the course of the proof at the ends of steps 2 and 6, respectively. The cutoffs N_1 and N_2 only depend on the fixed constants from (1) and (2) above. We then set $N_0 = \max\{N, N_1, N_2\}$ in the conclusion of the theorem where N is the cutoff for Proposition 10.12 to hold.

Step 1: Definition of $\mathcal{I}_n^1$. Let Γ_1 be a connected component of $B_{\delta_0/2}(x) \cap \gamma_1$ within distance υ of γ_2. Let G_ω^n be the $(C', \lambda', \epsilon', \mathcal{C}_{2\theta})$-tempered points at time n lying in Γ_1 (See Definition B.11). Note that $G_\omega^n \subseteq G_\omega^{n-1}$. We set

$$\eta_n(x) = \frac{1}{4(\max\limits_{1 \leq m \leq n}\{\|D_x f_\omega^m\| e^{(n-m)\lambda'/2}\})}, \tag{10.19}$$

and

$$\delta_n(x) = \eta_n^{(1+\sigma)}(x). \tag{10.20}$$

We now construct $\mathcal{I}_n^1$. For each $x \in G_\omega^n$, we say that x is *padded* if $B_{\delta_n(x)}^{\gamma_1}(x) \subseteq G_\omega^n$, where the $B_{\delta_n}^{\gamma_1}(x)$ denotes a ball of radius δ_n about x in γ_1 with respect to the arclength on γ_1. We let H_ω^n denote the set of all such padded points. Let $\hat{\mathcal{I}}_n^1 \subset \gamma_1$ be the set

$$\hat{\mathcal{I}}_n^1 = \bigcup_{x \in H_\omega^n} B_{\delta_n(x)}^{\gamma_1}(x). \tag{10.21}$$

Note that $\hat{\mathcal{I}}_n^1$ is a finite union of intervals. Delete intervals of length $K_1 e^{-4\Lambda_{\max} n}$ from the edges of each component where $K_1 > 0$ is a fixed small constant that we choose below. Call this trimmed collection of intervals $\mathcal{I}_n^1$.

We next check that $\mathcal{I}_n^1 \subseteq \mathcal{I}_{n-1}^1$. By the definition of δ_n, $\delta_n(x) \le e^{-\lambda'/2}\delta_{n-1}(x)$, thus as long as K_1 is sufficiently small,

$$\delta_n \le e^{-\lambda'/4}\delta_{n-1} - K_1 e^{-4(n-1)\Lambda_{\max}}. \tag{10.22}$$

Thus from the definition of $\mathcal{I}_n^1$, it is immediate that $\mathcal{I}_n^1 \subseteq \mathcal{I}_{n-1}^1$.

Step 2: Definition of $\mathcal{I}_n^2$. From the previous step, we know that any point in $\mathcal{I}_n^1$ satisfies the hypotheses of Proposition 10.12. Since $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are uniformly θ_0-transverse to $\mathcal{C}_{3\theta}$, it follows from Proposition 10.12(4a) that the fake stable manifold $W^s_{n,\delta_1}(y)$ of each point $y \in \mathcal{I}_n^1$ intersects γ_2. Hence, there is a well defined holonomy $H_n^s\colon \mathcal{I}_n^1 \to \gamma_2$ which satisfies all the conclusions of Proposition 10.12. We define

$$\mathcal{I}_n^2 = H_n^s(\mathcal{I}_n^1). \tag{10.23}$$

Next we check that $\mathcal{I}_n^2 \subseteq \mathcal{I}_{n-1}^2$. For this we will use the control on the fluctuations in the size of the holonomies from Claim 10.14 below. As we vary n, the fluctuations in $H_n^s(y)$ are smaller than the width of the neighborhoods δ_n in (10.20), and the result will follow.

Suppose that $x \in \mathcal{I}_n^1$. We must show that $H_n^s(x) \in \mathcal{I}_{n-1}^2$. Note that while x might not be in H_ω^n it is in G_ω^n. So, there exists some point y such that $x \in B^{\gamma_1}_{\delta_n(y)}(y)$ and hence also in $B^{\gamma_1}_{\delta_{n-1}(y)}(y)$.

To show that $H_n^s(x) \in \mathcal{I}_{n-1}^2$, we estimate how far $H_n^s(x)$ is from $H_{n-1}^s(x)$ and then estimate how far $H_n^s(x)$ is from the boundary of $H_{n-1}^s(B^{\gamma_1}_{\delta_{n-1}(y)}(y))$. For the former, we use the following claim.

Claim 10.14. *There exists $C_1 > 0$ such that if $x \in B^{\gamma_1}_{\delta_n(y)}(y)$ for some $y \in \mathcal{I}_n^1$ then*

$$d_{\gamma_2}(H_n^s(x), H_{n-1}^s(x)) \le C_1 \eta_n^{1.99(1-\sigma)^2}(y).$$

Proof. First we show that there exists C_a such that

$$\|D_y f_\omega^n\| \ge C_a \eta_n^{-(1-\sigma)}(y). \tag{10.24}$$

Let $N \le m \le n$, be the number achieving the maximum in the definition of η_n, (10.19). From $(C', \lambda', \epsilon')$ temperedness,

$$\|D_y f_\omega^n\| = \|D_y f_\omega^{m+(n-m)}\| \ge e^{-C'} e^{\lambda'(n-m)} e^{-\epsilon' m} \|D_y f_\omega^m\|. \tag{10.25}$$

By the definition of η_n,

$$\eta_n^{-(1-\sigma)} \le 4^{(1-\sigma)} e^{(n-m)(1-\sigma)\lambda'/2} \|D_y f_\omega^m\|^{(1-\sigma)}. \tag{10.26}$$

But from $(C', \lambda', \epsilon')$-temperedness, $\|D_y f_\omega^m\| \ge e^{-C'} e^{\lambda' m}$. Hence as long as $\lambda'\sigma > 2\epsilon'$, it follows that there exists C_b such that for all $n \ge N$ we have $C_b \|D_y f_\omega^m\|^{(1-\sigma)} \le \|D_y f_\omega^m\| e^{-2\epsilon' m}$. Hence there is some C_c such that

$$C_c \eta_n^{-(1-\sigma)} \le e^{(n-m)\lambda'/2} \|D_y f_\omega^m\| e^{-2\epsilon' m}.$$

Comparing the above equation with (10.25) yields equation (10.24).

Next, as explained in Step 1 above, all points in $\mathcal{I}_n^1$ satisfy the conclusions of Proposition 10.12(2). Thus,

$$\|D_x f_\omega^n\| \ge \|D_y f_\omega^n\|^{(1-\sigma)}.$$

Combining this with (10.24) gives:

$$\|D_x f_\omega^n\| \ge C_a^{(1-\sigma)} \eta_n^{-(1-\sigma)^2}(y).$$

Then applying Proposition 10.12(4c) gives the conclusion. □

We now continue with the proof that $\mathcal{I}_n^2 \subseteq \mathcal{I}_{n-1}^2$. First, note that by the triangle inequality,

$$d_{\gamma_1}(x, \partial B^{\gamma_1}_{\delta_{n-1}(y)-K_1e^{-\Lambda_{\max}(n-1)}}(y)) \geq \delta_{n-1}(y) - K_1 e^{-4(n-1)\Lambda_{\max}} - \delta_n(y).$$

We then apply H^s_{n-1}. By Proposition 10.12(4b) it follows that

$$\begin{aligned} d_{\gamma_2}(H^s_{n-1}(x), \partial H^s_{n-1}(B^{\gamma_1}_{\delta_{n-1}(y)-K_1e^{-4\Lambda_{\max}(n-1)}}(y))) \\ \geq D_1(\delta_{n-1}(y) - K_1 e^{-4(n-1)\Lambda_{\max}} - \delta_n(y)). \end{aligned} \tag{10.27}$$

But by Claim 10.14, $d_{\gamma_2}(H^s_{n-1}(x), H^s_n(x)) \leq C_2\eta_n^{1.99(1-\sigma)^2}(y)$. Hence by the triangle inequality

$$\begin{aligned} d_{\gamma_2}(H^s_n(x), \partial H^s_{n-1}(B^{\gamma_1}_{\delta_{n-1}(y)-K_1e^{-4\Lambda_{\max}(n-1)}}(y))) \\ \geq D_1(\delta_{n-1}(y) - K_1 e^{-4(n-1)\Lambda_{\max}} - \delta_n(y)) - C_2\eta_n^{-1.99(1-\sigma)^2}(y). \end{aligned}$$

By (10.22) $\delta_{n-1} - K_1 e^{-(n-1)\Lambda_{\max}} - \delta_n \geq \left(1 - e^{-\lambda/4}\right)\delta_{n-1}$. Hence as $\eta_n^{1.99(1-\sigma)^2}$ is of a higher order than δ_n, there exists some N_1 such that for $n \geq N_1$,

$$d_{\gamma_2}(H^s_n(x), \partial \mathcal{I}^2_{n-1}) \geq 2^{-1} D_1 \left(1 - e^{-\lambda/4}\right)\delta_{n-1}(y) > 0. \tag{10.28}$$

This shows that $H^s_n(\mathcal{I}^1_n) \subset \mathcal{I}^2_{n-1}$ as desired.

Step 3. Lengths of curves in $\mathcal{I}^i_n \setminus \mathcal{I}^i_{n-1}$. This is needed to estimate the regularity of $P^i_n \setminus P^i_{n+1}$.

First we consider the size of the trimmed segments when we pass from $\hat{\mathcal{I}}^1_n$ to $\mathcal{I}^1_n$. Any connected component of $\hat{\mathcal{I}}^1_n$ has length at least $\delta_n(x)$ for some x. Note that this is bounded below by an exponential $e^{-\Lambda_{\max}n}$. Then as we trim a remaining $K_1e^{-4\Lambda_{\max}n}$ length off these intervals when we pass from $\hat{\mathcal{I}}^1_n$ to $\mathcal{I}^1_n$, we see that each interval we trim has length at least $K_1e^{-4\Lambda_{\max}n}$.

There are two ways that $x \in \mathcal{I}^1_{n-1}$ may fail to be in $\mathcal{I}^1_n$. Write $\mathcal{I}^1_{n-1}(x)$ for the connected component of $\mathcal{I}^1_{n-1}$ containing x. Then either $\mathcal{I}_{n-1}(x)$ contains a point y that is in $\mathcal{I}^1_n$ or the entire component containing x is deleted. In the first case the connected component of $\mathcal{I}^1_{n-1} \setminus \mathcal{I}^1_n$ containing x has length at least $k_1e^{-4\Lambda_{\max}n}$ by the previous paragraph. In the second case, the removed segment is at least $e^{-\Lambda_{\max}n}$ long. Thus we have obtained an exponential lower bound on the lengths of curves in $\mathcal{I}^1_{n-1} \setminus \mathcal{I}^1_n$.

As $H^s_n(\mathcal{I}^1_n) = \mathcal{I}^2_n$ we can use the size of the gaps in $\mathcal{I}^1_n$ to estimates the size of those in $\mathcal{I}^2_n$. Note that from estimate (10.27), each segment in $\mathcal{I}^2_{n-1} \setminus \mathcal{I}^2_n$ has width at least $D_1K_1e^{-4\Lambda_{\max}n}$.

Step 4. Definition of the densities. So far we have defined the underlying curves $\mathcal{I}^1_n, \mathcal{I}^2_n$ that the standard families P^1_n, P^2_n will be defined on. We now define the densities on $\mathcal{I}^1_n$ and $\mathcal{I}^2_n$. To begin, we will define ρ^1_N and ρ^2_N where N is the first time we attempt to fake-couple. From Proposition 10.12(4b) , there exists $a_0 > 0$ such that for $B \subseteq \mathcal{I}^1_n$,

$$a_0(H^s_n)_*\rho^1|_B \leq \rho^2|_{H^s_{n-1}(B)}. \tag{10.29}$$

We then take as our initial definition:

$$\rho^1_N = a_0\rho^1|_{\mathcal{I}^1_N} \text{ and } \rho^2_N = (H^s_N)_*\rho^1_N. \tag{10.30}$$

This gives us ρ^1_N and ρ^2_N.

We now define ρ^1_n and ρ^2_n for $n \geq N$. We set:

$$\rho^1_n = \rho^1_{n-1}(1 - e^{-n\hat{\eta}})|_{\mathcal{I}^1_n} \tag{10.31}$$

where $\hat{\eta}$ is chosen in equation (10.41) below. We then define

$$\rho_n^2 = (H_n^s)_*(\rho_n^1). \tag{10.32}$$

As we push forward ρ_n^1 by the holonomy H_n^s, which carries $\mathcal{I}_n^1$ to $\mathcal{I}_n^2$, ρ_n^2 is a measure on $\mathcal{I}_n^2$. This defines completely P_n^1 and P_n^2.

The rest of the proof will be checking that the standard families P_n^1 and P_n^2 have the required properties to be a fake coupling. Some are evident from the definition above, but it remains to check:

(1) the regularity of ρ_n^1 and ρ_n^2,
(2) that ρ_n^2 is a decreasing sequence of measures, and
(3) the goodness of the standard families $P_n^i \setminus P_{n-1}^i$ for $i \in \{1, 2\}$.

Step 5: Regularity of ρ_n^1 and ρ_n^2. In this step we study the log-Hölder constants of ρ_n^1 and ρ_n^2 for $n \geq N$. Note that ρ_n^1 is ρ_N^1 scaled by a constant that it has the same log-Hölder constant as ρ_N^1.

Before proceeding to study the regularity of ρ_n^2, we introduce some notation related to the Jacobian of the holonomies. Typically the Jacobian of an invertible, absolutely continuous map $\phi\colon (X, \nu) \to (Y, \mu)$ is the Radon-Nikodym derivative $d\phi^*\mu/d\nu$. In our case, as we are pushing forward the density ρ_n^1 by H_n^s, the result is the same thing as pulling back ρ_n^1 by $(H_n^s)^{-1}$. To simplify notation, we will simply write J_n for the Jacobian of $(H_n^s)^{-1}$, which is a function $J_n\colon \mathcal{I}_n^2 \to \mathbb{R}_{>0}$. Returning to ρ_n^2, this function satisfies for $y \in \mathcal{I}_n^2$ that

$$\rho_n^2(y) = J_n(y)\rho_n^1((H_n^s)^{-1}(y)). \tag{10.33}$$

As the assumptions on the holonomies are symmetric in γ_1 and γ_2, we know from Proposition 10.12(4b) that H_n^s is D_1-bilipschitz. Thus by Proposition 10.12(4d), there exists D_2 such that J_n is log-α-Hölder with constant D_2 for all $n \geq N$. Next, since ρ_n^1 is log-α-Hölder with constant C_0, $\rho_n^1 \circ (H_n^s)^{-1}$ is log-α-Hölder with constant $D_1^\alpha C_0$. As mentioned before, J_n is log-α-Hölder with constant D_2. The product of log-α-Hölder functions is log-α-Hölder with constant equal to the sum of the constants. Thus by (10.33), we see that ρ_n^2 is $D_1^\alpha C_0 + D_2$ log-α-Hölder. Thus we have obtained uniform log-α-Hölder control for ρ_n^1 and ρ_n^2.

We need one more estimate before we continue: an actual Hölder, rather than log-Hölder, bound on ρ_n^1 and ρ_n^2; we need this as at a certain point we will compare the difference of these functions rather than their ratio. We obtain this bound by rescaling the functions by a constant; however we need to be sure the constant is not too big.

From (10.29), it follows from the C_0 log-α-Hölder constant of the density that there exists $D \geq 1$ such for any $x \in \hat{\gamma}_1$ and $y \in \hat{\gamma}_2$,

$$D^{-1} \leq \frac{\rho^1(x)}{\rho^2(y)} \leq D. \tag{10.34}$$

Note that for a log-α-Hölder function $\rho\colon K \to (0, \infty)$ on a set K of diameter at most 1 that there exists D depending only on the log-Hölder constant of ρ such that

$$D^{-1} \leq \rho / \max \rho \leq 1.$$

If we let M denote the larger of the maximum of ρ_n^1 and the maximum of ρ_n^2, then we may define for $i \in \{1, 2\}$, $\widetilde{\rho}_n^i = \rho^i / M$. Then as the maximums of ρ^1 and ρ^2 are uniformly comparable, note that there exists $D > 0$ depending only on C_0 such that for $i \in \{1, 2\}$,

$$D^{-1} \leq \widetilde{\rho}^i \leq 1.$$

In particular, as as exp is 1-Lipschitz on $(-\infty, 0]$, it follows that $\widetilde{\rho}_n^1, \widetilde{\rho}_n^2$ are both uniformly α-Hölder with the same constant as their log-Hölder constant. Below we will work with these rescaled functions that have maximum 1 and just write ρ_n^1 instead of $\widetilde{\rho}_n^1$. Note that we have not gained any extra regularity for free: to get the lower bound D depending only on the log-Hölder constant on both at the same time used substantial input from our setup.

Step 6. Sign and regularity of $\rho_{n-1}^2 - \rho_n^2$. We now analyze $\rho_{n-1}^2 - \rho_n^2$. In particular, we show that ρ_n^2 is a decreasing sequence of densities. To begin, we will obtain a lower bound on $\rho_{n-1}^2 - \rho_n^2$. Then we will use the various lemmas relating Hölder and log-Hölder functions to conclude a bound on the regularity of $\rho_{n-1}^2 - \rho_n^2$. By definition:

$$\begin{aligned}\rho_{n-1}^2 - \rho_n^2 =& \rho_{n-1}^1((H_{n-1}^s)^{-1}y)J_{n-1}(y) - (1 - e^{-n\hat{\eta}})\rho_{n-1}^1((H_n^s)^{-1}y)J_n(y)\\ =& [J_{n-1}(y)(\rho_{n-1}^1((H_{n-1}^s)^{-1}(y)) - \rho_{n-1}^1((H_n^s)^{-1}(y)))]+\\ & [\rho_{n-1}^1((H_n^s)^{-1}(y))(J_{n-1}(y) - J_n(y))] + [e^{-n\hat{\eta}}\rho_{n-1}^1((H_n^s)^{-1}y)J_n(y)]\\ =& A + B + C.\end{aligned}$$

We next estimate A, B, and C.

Term A. To estimate term A, we first pull the function back to γ_1 by composing with H_n^s. Let $Q_n = (H_{n-1}^s)^{-1} \circ H_n^s$. For $y \in \mathcal{I}_n^1$, there exists $y' \in G_\omega^m$ satisfying the hypotheses of Claim 10.14 such that $d_{\gamma_2}(H_n^s(y), H_{n-1}^s(y)) < C_1\eta_n(y')^{-1.99(1-\sigma)^2}$. By Lipschitzness of the holonomies from Proposition 10.12(4b), this implies that

$$d_{\gamma_1}(Q(y), y) < D_1C_1\eta_n(y')^{1.99(1-\sigma)^2}. \tag{10.35}$$

Precomposing again with $(H_n^s)^{-1}$ gives that for $y \in \mathcal{I}_n^2$,

$$d_{\gamma_1}((H_{n-1}^s)^{-1}(y), (H_n^s)^{-1}(y)) \leq D_1^2C_1\eta_n(y')^{1.99(1-\sigma)^2}. \tag{10.36}$$

But this implies, using Lemma A.11 and (10.36) in the second line, that:

$$|A| = \left|J_{n-1}(y)(\rho_{n-1}^1((H_{n-1}^s)^{-1}(y)) - \rho_{n-1}^1((H_n^s)^{-1}(y)))\right| \tag{10.37}$$

$$\leq |J_{n-1}(y)|\,(D_1^2C_2\eta_n(y')^{1.99(1-\sigma)^2})^\alpha\rho_{n-1}^1((H_n^s)^{-1}y) \tag{10.38}$$

$$\leq |J_{n-1}(y)|\,C_3\eta_n^{1.99(1-\sigma)^2\alpha}\rho_{n-1}^1(H_n^s(y)) \tag{10.39}$$

$$\leq C_Ae^{-1.99\lambda'(1-\sigma)^2\alpha n}\rho_{n-1}^1((H_n^s)^{-1}y). \tag{10.40}$$

where we have used temperedness to pass to the last line. We now turn to the next term.

Term B. This term is simpler. We use (10.17) in the third step below:

$$\begin{aligned}|B| \leq \left|\rho_{n-1}^1((H_n^s)^{-1}(y))(J_{n-1}(y) - J_n(y))\right| \leq \left|\rho_{n-1}^1\right|\,|J_{n-1}(y) - J_n(y)|\\ \leq \left|\rho_{n-1}^1\right|e^{-n\eta} \leq e^{-n\eta}\rho_{n-1}^1((H_n^s)^{-1}(y)).\end{aligned}$$

Term C. The final term is straightforward

$$C = e^{-\hat{\eta}n}\rho_{n-1}^1((H_n^s)^{-1}y)J_n(y) \leq D_2e^{-\hat{\eta}n}\rho_{n-1}^1((H_n^s)^{-1}(y)).$$

We can now conclude. Combining the estimates on A, B, C, we see that

$$\rho_{n-1}^2(y) - \rho_n^2(y) \geq [D_2e^{-\hat{\eta}n} - e^{-\eta n} - C_Ae^{-1.99\lambda'(1-\sigma)^2\alpha n}]\rho_{n-1}^1((H_n^s)^{-1}y).$$

In particular, as long as

$$0 < \hat{\eta} < \min\{\eta/2, -1.99\lambda'(1-\sigma)^2\alpha/2\}, \tag{10.41}$$

it follows that there exists N_2 such that for $n \geq N_2$,

$$\rho^2_{n-1} - \rho^2_n \geq e^{-2\hat{\eta}n}\rho^2_{n-1}.$$

Also because $J_n, \rho^2_n, \rho^2_{n-1}$ are uniformly bounded, there exists D_3 such that

$$D_3 \geq \rho^2_n - \rho^2_{n-1} \geq e^{-2\hat{\eta}n}D_3^{-1}.$$

Thus we can apply Claim A.10 to the function $(\rho^2_n - \rho^2_{n-1}) \geq D_3^{-1}e^{-2\hat{\eta}n}$. As ρ^2_n and ρ^2_{n-1} are uniformly α-Hölder from Step 5, we obtain that there exists D_4 such that $\rho^2_{n-1} - \rho^2_n$ is uniformly $D_4 e^{2\hat{\eta}n}$ log-α-Hölder. This concludes the analysis of the Hölder regularity of $\rho^2_{n-1} - \rho^2_n$.

Step 7: Bookkeeping. In this step we verify that for each point $y \in \mathcal{I}^1_n$ that a positive proportion of the mass over y is retained during the fake coupling procedure. This is straightforward to see because at each step, we discard $e^{-n\hat{\eta}/2}$ proportion of the remaining mass in $\rho^1_n(y)$. Thus from the definition 10.30 of ρ^1_N the amount of mass is bounded below by

$$\rho^1_n(y) \geq a_0\rho^1(y)\prod_{n\geq N}(1 - e^{-n\hat{\eta}}) > 0.$$

Thus we keep a positive proportion of the mass above each $y \in \mathcal{I}^n_\omega$ for all $n \geq N$.

Step 8: $n = \infty$ behavior As the sequences ρ^1_n and ρ^2_n are decreasing they converge to some limiting measures ρ^1_∞ and ρ^2_∞. Further, by Proposition B.13, the true stable holonomies H^s_∞ satisfy $(H^s_\infty)_*\rho^1_n = \rho^2_n$ as required.

Step 9: $(C, \lambda, \epsilon, \mathcal{C}_\theta)$-tempered points are never dropped. Finally, we must show that we actually keep the $(C, \lambda, \epsilon, \mathcal{C}_\theta)$ tempered points throughout the entire procedure, so that part (3) of the requirements for a fake coupling are satisfied. Suppose that (ω, x) is such a $(C, \lambda, \epsilon, \mathcal{C}_\theta)$-tempered trajectory. It suffices to show that for each n that all points in $B_{\delta_n(x)}(x)$ are $(C', \lambda', \epsilon', \mathcal{C}_{2\theta})$-tempered, as from the procedure above this ensures that $x \in \mathcal{I}^1_n$ for all n. By Part (2) of Proposition 10.12, this follows as long as $\delta_n(x) \leq \|D_x f^n_\omega\|^{-(1+\sigma)}$. This inequality holds because by the definition of η_n, (10.19), $\eta_n(x) \leq \|D_x f^n_\omega\|^{-1}$, and $\delta_n(x) = \eta_n^{(1+\sigma)}$.

Thus we have verified all of the required claims in the definition of fake coupling as well as the additional required claim about the goodness of the families $P^i_{n-1} \setminus P^i_n$, we conclude the proof. □

We now have everything ready to prove the local coupling lemma, Lemma 7.10.

Proof of Lemma 7.10. Almost everything in the statement of Lemma 7.10 is contained in the statement of Lemma 10.13. We explain them in order.

Item 1 follows because the points we stop trying to couple at time n are precisely the points in $\hat{\gamma}_i$ that are in $P^i_n \setminus P^i_{n-1}$. As the standard family $P^i_n \setminus P^i_{n-1}$ is $n\Lambda$-good, the claim follows with $L = \Lambda$.

Item 2 is the statement in the final paragraph of Lemma 10.13.

Item 3 is more complicated. There are two ways that a point $x \in \mathcal{I}^1_{n-1}$ fails to appear in $\mathcal{I}^1_n$. The first is that x is not in any interval $B^{\gamma_1}_{\delta_n(y)}(y)$ for any $y \in H^n_\omega$. The second is if x is in an interval that gets trimmed off of $\hat{\mathcal{I}}^1_n$.

First we consider the former case. This means that some y such that $x \in B^{\gamma_1}_{\delta_{n-1}(y)}(y)$ failed to be tempered at time n. In $\Sigma \times \gamma_1$, we consider the union of these intervals:

$$U_n = \bigcup\{\{\omega\} \times B^{\gamma_1}_{\delta_{n-1}(y)}(y) : y \in \mathcal{I}^1_{n-1} \setminus \mathcal{I}^1_n \text{ for the word } \omega\}.$$

Note that as each of these sets $B^{\gamma_1}_{\delta_{n-1}(y)}(y)$ contains a point z that fails to be tempered at time n that (ω, z) has cushion that is within $\Lambda_{\max}$ of C', the cutoff for tempering to fail. By Proposition 10.10, as all the points in $B^{\gamma_1}_{\delta_{n-1}(y)}(y)$ satisfy the hypotheses of that proposition due to the size of $\delta_{n-1}(y) \le \|D_y f^n_\omega\|^{-(1+\sigma)}$ and the tempering, this implies that all points in $B^{\gamma_1}_{\delta_{n-1}(y)}(y)$ have cushion at most $C' + \Lambda_{\max} + D$. But by Proposition 4.10, the number of points having cushion of this size is exponentially small. Thus $\mu \otimes \rho(U_n) \le D_1 e^{-n\eta}$ for some $D_1, \eta > 0$, and we have an exponential tail for points experiencing the first type of failure.

In the case that a point fails to be included because it was trimmed off, it was observed in Step 3 of the coupling construction, that every curve being trimmed has length at least $e^{-(1+\sigma)\Lambda_{\max} n}$ and the amount we cut off has length $2K_1 e^{-4\lambda_{\max} n}$. Thus as $1/2 \le \rho(x)/\rho(y) \le 2$ for two points x, y along the curve we are coupling, the amount we trim has mass at most $4e^{-2\Lambda_{\max} n}$ times the mass of the curve. Thus summing over all curves we stop on at most $4e^{-2\Lambda_{\max} n}$ mass, which is exponentially small.

The last way that mass is lost during the local coupling procedure is when we rescale the density by $(1 - e^{-n\hat{\eta}})$ in Step 4, which also gives at most an exponentially small amount of mass is stopping at time n. This concludes the proof of the tail bound.

Item 4 follows from Proposition 10.12(1). □

11. Mixing theorems

11.1. **Overview of the section.** In this section we prove our main result, Theorem 1.1. The proof will rely on coupling and expansion following the standard argument, see e.g. [CM06].

First, we show that coupling implies equidistribution of standard families by coupling a given family to a family representing volume and using that volume is invariant by the dynamics. See Proposition 11.9 for details.

Next, we use the expansion and exponential equidistribution to obtain exponential mixing using the following reasoning. Consider an R-good standard family $\hat{\gamma}$ and let $f^n_\omega(\hat{\gamma})$ be its image after n iterations. We shall show that for almost all ω that $f^n_\omega(\hat{\gamma})$ contains a subfamily P_n with the following properties:

(1) P_n consists of ϵn-good standard pairs
(2) standard pairs in P_n contract backwards in time
(3) the forward image of pairs from P_n equidistribute at an exponential rate
(4) the complement of P_n has exponentially small measure.

Now given Hölder functions ϕ and ψ we obtain exponential decorrelation between $\phi \circ f^N_\omega$ with $N = cn$ and ψ using that ψ is constant on the elements of P_n (up to exponentially small error), $\phi \circ f^N_\omega$ is equidistributed on the elements of P_n (up to exponentially small error), and the complement of P_n is exponentially small.

The purpose of this section is to execute this argument precisely, using the results of Sections 4, 8, and the appendices.

11.2. **Preparatory lemmas.** Below we will use Definition A.14 from §A.6 in the appendix. Briefly, this definition concerns a $(C, \lambda, \epsilon, \theta)$-*forward tempered point* at time n for a vector $v \in T_x M$, which is a (C, λ, ϵ)-forward tempered time n such that E^s_n makes angle at least θ with v.

Proposition 11.1. *Suppose that M is a closed surface and $(f_1, \dots, f_m)$ is an expanding on average tuple of diffeomorphisms in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. There exists $\lambda > 0$ such that for all sufficiently small $\epsilon > 0$ there exist C_0, $N \in \mathbb{N}$, and $\alpha > 0$ such that for all $n \ge N$, and any direction*

$v \in T^1_x M$,

$$\mu(\omega : (\omega, x) \text{ is not } (\epsilon n + C_0, \lambda, \epsilon, C_0 e^{-\epsilon n})\text{-forward tempered at time } n \text{ relative to } v) \leq e^{-n\alpha}.$$

Proof. Proposition 4.8 says that there exist $\lambda > 0$ such that for arbitrarily small $\epsilon > 0$, there exists $\alpha > 0$ such that the measure of the words ω that are not (C, λ, ϵ)-subtempered for all $n \geq 0$ is at most $e^{-\alpha C}$. From Proposition 4.14 there exists some $C_2, c, \theta > 0$ such that for all sufficiently small ϵ' as long as $n \geq c\,|\ln(\epsilon')| = N_0$, then for all $n \geq N_0$, the probability that $E^s_n \in B_{\epsilon'}(v)$ is at most $C_2(\epsilon')^\theta$. Taking $\epsilon' = e^{-\epsilon n}$, this gives that the probability that
$E^s_n \in B_{e^{-\epsilon n}}(v)$ for $n \geq N_0$ is at most $C_2 e^{-\theta\epsilon n}$ as long as ϵ is sufficiently small relative to c, $n \geq c\epsilon n$. Combining these two estimates, we obtain the result. □

Below, we will typically assume that the standard family or standard pair we are considering has unit mass. The statements below can be adapted to any amount of mass by multiplying the right hand side of the bound by the mass of the family.

Definition 11.2. Given a standard pair $\hat{\gamma} = (\gamma, \rho)$, for $x \in \gamma$ we say that (ω, x) is (n, λ, ϵ)-*backwards good* if

(1) $f^n_\omega(x)$ is contained in a standard pair $B(\omega, x) \subseteq f^n_\omega(\hat{\gamma})$ that is ϵn-good, and
(2) $Df^{-n}_\omega B^\gamma_{e^{-14\epsilon n}}(x)$ has diameter at most $e^{-(\lambda/2)n}$.

We define analogously the same notion for a standard family.

Proposition 11.3. *(Annealed goodness) Suppose that M is a closed surface and $(f_1, \ldots, f_m)$ is an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. Then there exists $\lambda > 0$ such that for all sufficiently small $\epsilon > 0$, if we fix $R > 0$ there exists $\alpha, C > 0$ such that for any R-good, unit mass standard family $\hat{\gamma}$ with associated measure ρ:*

$$(\mu \otimes \rho)(\{(x, \omega) : (x, \omega) \text{ is not } (n, \lambda, \epsilon)\text{-backwards good}\}) \leq Ce^{-\alpha n}. \tag{11.1}$$

Proof. This is immediate from Propositions A.15 and 11.1. □

From Proposition 11.3, we can deduce a related quenched statement for almost every ω.

Lemma 11.4. *(Quenched goodness) Under the hypotheses of Proposition 11.3, there exist $\lambda, \alpha, D > 0$ such that for all sufficiently small $\epsilon > 0$ and a unit mass R-good standard family $\hat{\gamma}$, then for almost every ω, there exists C_ω such that $1 - C_\omega e^{-\alpha n}$ proportion of points in $\hat{\gamma}$ are (n, λ, ϵ)-backwards good for ω. Further,*

$$\mu(\omega : C_\omega > C) \leq DC^{-1}.$$

Proof. Let A^ω_n be the set of points in $\hat{\gamma}$ that are not (n, λ, ϵ)-backwards good for ω. Then

$$\mu(\omega : \exists n\ \rho(A^\omega_n) > Ce^{-(\alpha/2)n}) \leq \sum_{n \geq 0} \mu(\omega : \rho(A^\omega_n) > Ce^{-(\alpha/2)n}) \leq \sum_{n \geq 0} C^{-1} C_1 e^{-(\alpha/2)n} \leq C^{-1} D,$$

where the second inequality follows from (11.1) and the Markov inequality. The result follows. □

We also need another proposition, that says that on the ϵn-good neighborhoods at time n that we have rapid coupling, which will then imply that these neighborhoods rapidly equidistribute. The following estimate is immediate from Proposition 7.7.

Proposition 11.5. *Suppose* $(f_1, \dots, f_m)$ *is as in Proposition 11.3. Then there exists* $\lambda > 0$ *such that for any sufficiently small* $\epsilon > 0$ *there exist* $C, \alpha > 0$ *such that the following holds. For any* $n \in \mathbb{N}$*, suppose* P^1 *and* P^2 *are two unit mass standard families of* ϵn*-good curves. Then there exists a coupling function* Υ *and stopping times* $\hat{T}^1, \hat{T}^2$ *as in Proposition 7.7 such that for* $i \in \{1, 2\}$*:*

$$(\mu \otimes \rho^i)(\{(x, \omega) : \hat{T}^i(x, \omega) > j\}) \le Ce^{\epsilon n} e^{-\alpha j}.$$

Remark 11.6. *In the applications of Proposition 11.5 below we will assume unless it is explicitly stated otherwise that* P^2 *is the family representing the volume from Proposition 7.5. We couple with a family representing volume because it implies that the statistics of an arbitrary standard family* P_1 *approach those of volume.*

In what follows for a word ω *at time* i*, we have subfamilies* $P^1_{i,\omega}$ *and* $P^2_{i,\omega}$ *of* $f^i_\omega(P^1)$*. We then apply Proposition 11.5 above, to find a pair of stopping times* $\hat{T}^1_i$ *and* $\hat{T}^2_i$ *defined on* $f^i_\omega(P^1_{i,\omega})$ *and* $f^i_\omega(P^2_{i,\omega})$ *respectively. Note that the the* $\hat{T}^i$ *are not defined on all of* $f^i_\omega(\hat{\gamma})$ *because not all points in this pair need be* ϵn*-good.*

Then from Proposition 11.5 we obtain the following.

Proposition 11.7. *Let* $(f_1, \dots, f_m)$*,* $\hat{\gamma}$*,* ρ*, and* $\lambda, \epsilon, \alpha > 0$ *as be as in Proposition 11.3, then there exists* C *such that if we let the* $\hat{T}^1_n$ *be the stopping time defined as in Remark 11.6, for all* $i, n \ge 0$ *we have the bound:*

$$(\mu \otimes \rho)((x, \omega) : x \in P^1_{i,\omega} \text{ and } \hat{T}^1_i(x, \omega) > i + n) \le Ce^{\epsilon i} e^{-n\alpha}. \tag{11.2}$$

From this, we easily deduce a statement about each ω.

Proposition 11.8. *Let* $(f_1, \dots, f_m)$*,* $\lambda, \epsilon > 0$ *and* $\hat{\gamma}, \rho$ *be as in the setting of Proposition 11.3 and Remark 11.6, then there exists* $\alpha, D_1 > 0$ *such that*

$$\mu(\omega : \text{there exists } i \text{ such that } \rho(x : (x, \omega) \text{ is } (i, \lambda, \epsilon)\text{-backwards good}) < 1 - Ce^{-i\alpha} \text{ or} \tag{11.3}$$

$$\text{there exist } (i, n) \text{ such that } \rho(x \in P^1_{i,\omega} : \hat{T}^1_i(x, \omega) \ge i + n) \ge C^2 e^{\epsilon i} e^{-n\alpha}) \le D_1 C^{-1}. \tag{11.4}$$

Proof. To control the event in (11.4) let $B^\omega_{i,n} = \{x \in P^1_{i,\omega} : \hat{T}^1_i(x, \omega) > i + n\}$. By (11.2) and the Markov inequality, there is $C_1 > 0$ such that

$$\mu(\{\omega : \rho(B^\omega_{i,n}) > Ce^{2\epsilon i} e^{-(\alpha/2)n}\}) \le C_1 C^{-1} e^{-\epsilon i} e^{-(\alpha/2)n}. \tag{11.5}$$

Then using (11.5), we find that

$$\mu(\omega : \text{for some } i, n\ \rho(\{x : \hat{T}^1_i(x, \omega) \ge i + n\}) \ge C^2 e^{\epsilon i} e^{-n\alpha/2})$$

$$\le \sum_{i \ge 0} \sum_{n \ge 0} \mu(\{\omega : \rho(B^\omega_{i,n}) \ge C^2 e^{\epsilon i} e^{-(\alpha/2)n}\}) \le \sum_{i \ge 0} \sum_{n \ge 0} C_1 C^{-1} e^{-\epsilon i} e^{-(\alpha/2)n} \le C^{-1} C_2$$

for some C_2 provided that ϵ is small enough. Combining this estimate with Proposition 11.3 to control the event in (11.3) allows us to conclude. □

11.3. **Quenched equidistribution.** Using the quenched coupling lemmas above, it is straightforward to deduce quenched equidistribution and correlation decay theorems. The ideas in the proofs below are essentially standard, compare with [CM06, Ch. 7], however some modifications are necessary because the quenched random dynamics is not stationary.

We start with quenched equidistribution.

Proposition 11.9. *(Quenched exponential equidistribution on subfamilies) Let* $(f_1, \dots, f_m)$ *be an expanding on average tuple in* $\mathrm{Diff}^2_{\mathrm{vol}}(M)$*, where* M *is a closed surface. There exists* $\lambda > 0$ *such that for all sufficiently small* $\epsilon > 0$*, fixed* $\beta \in (0,1)$ *and* R*, there exists* D_1 *such that for any* R*-good, unit mass standard family* $\hat{\gamma}$*, there exists* $\alpha, \nu > 0$ *such that for almost every* ω*, there exists* $C_\omega \geq 1$ *such that, such that coupling as in Remark 11.6:*

(1) There exists a subfamily $P_{i,\omega}$ *of* $(f^i_\omega)_*\hat{\gamma}$ *of* $e^{\epsilon i}$*-good standard pairs having total* ρ*–measure* $(1 - C_\omega e^{-\alpha i})$

(2) The atoms of $(f^i_\omega)^{-1}(P_{i,\omega})$ *have diameter at most* $e^{-\lambda/2 i}$*.*

(3) The atoms $A_{i,\omega} \in P_{i,\omega}$ *exponentially equidistribute, i.e., letting* $\overline{A}_{i,\omega}$ *be the normalized measure on* $A_{i,\omega}$*,*

$$\left| \int \phi \circ f^n_{\sigma^i\omega}\, d\overline{A}_{i,\omega} - \int \phi\, d\,\mathrm{vol} \right| \leq C_\omega e^{\epsilon i} e^{-\alpha n} \|\phi\|_{C^\beta}. \tag{11.6}$$

(4) We have a tail bound $\mu(\{\omega : C_\omega > C\}) \leq D_1 C^{-1}$*.*

Proof. From Lemma 11.4, the only thing that remains to be checked is that the individual atoms of $A_{i,\omega}$ are exponentially equidistributing.

Let $P_{i,\omega}$ be the subfamily of $f^i_\omega(\hat{\gamma})$ of curves that are $i\epsilon$-good. Let P^2 be a standard family representing volume as in Remark 11.6. Then coupling with P^2, we have the stopping time $\hat{T}_i$ on $P_{i,\omega}$ as discussed in Proposition 11.7 and uniform $\alpha, C_\omega > 0$ such that for all $i, n \in \mathbb{N}$,

$$\rho(x \in P_{i,\omega} : \hat{T}_i(x,\omega) > n + i) \leq C_\omega e^{\epsilon i} e^{-n\alpha}. \tag{11.7}$$

We would like to know that most of the curves in $P_{i,\omega}$ have all but an exponentially small amount of their points coupling quickly.

We claim that for a.e. ω there exists a subfamily $G_{i,\omega}$ of ϵi-good curves in $P_{i,\omega}$ of measure at least $1 - C_\omega e^{-\alpha i/3}$ such that for each $A \in G_{i,\omega}$ all but $e^{i\epsilon n} e^{-\alpha/3n}$ of the mass of the subfamily has coupled to volume by time $i + n$, i.e. $\hat{T}_i(x,\omega) \leq i + n$. Suppose that ω satisfies (11.7) and for the sake of contradiction, suppose that there is a subfamily B_i (of bad pairs) of $\hat{P}_{i,\omega}$ having measure more than than $e^{-\alpha i/3}$ so that for some n all pairs in B_i have more than $e^{i\epsilon} e^{-n\alpha/3}$ proportion of points not coupled at time $n + i$, i.e. $\hat{T}_i > i + n$. This implies that $\rho(x : \hat{T}_i(x,\omega) > n+i) \geq C_\omega e^{-2\alpha n/3} e^{i\epsilon}$, contradicting (11.7). Thus the claim about $G_{i,\omega}$ holds.

Suppose now that $A_{i,\omega} \in G_{i,\omega} \subseteq P_{i,\omega}$ is such a good atom where at time $n+i$ all but at most $e^{i\epsilon} e^{-(\alpha/3)n}$ proportion of the mass of $A_{i,\omega}$ has coupled to volume. Let $A^n_{i,\omega} \subseteq A_{i,\omega}$ be the set of points that have coupled by time $i+n$. Let Υ be the measure preserving coupling function and let $V^n = \Upsilon(A^n_{i,\omega})$ be the corresponding set of points in the standard family representing volume that have $\hat{T}_i(x,\omega) \leq i + n$. Then we may write the integral in question as

$$\left| \int \phi \circ f^n_{\sigma^i\omega}\, d\overline{A}_{i,\omega} - \int \phi\, dP_{\mathrm{vol}} \right|$$

$$\leq \left| \int_{A^{n/2}_{i,\omega}} \phi \circ f^n_{\sigma^i\omega}\, d\overline{A}_{i,\omega} - \int_{V^{n/2}} \phi\, dP_{\mathrm{vol}} \right| + \left| \int_{A_{i,\omega} \setminus A^{n/2}_{i,\omega}} \phi \circ f^n_{\sigma^i\omega}\, d\overline{A}_{i,\omega} \right| + \left| \int_{(V^{n/2})^c} \phi\, dP_{\mathrm{vol}} \right|$$

$$\leq \left| \int_{\Upsilon^{-1}(V^{n/2})} \phi \circ f^n_{\sigma^i\omega}(\Upsilon(x)) - \phi(x)\, dP_{\mathrm{vol}} \right| + 2C_\omega e^{i\epsilon} e^{-n\alpha/6} \|\phi\|_{C^\beta}.$$

As the points $\Upsilon(x)$ and x both lie in a common (C_0, λ, ϵ)-tempered local stable leaf of uniformly bounded length at time $i + n/2$, then we see that at time $i + n$, that

$$d(f^n_{\sigma^i\omega}\Upsilon(x), f^n_{\sigma^i\omega}(x)) \leq C_0^{-1} e^{-\lambda/2n}.$$

Now the Hölder regularity of ϕ implies that

$$\left|\int \phi \circ f^n_{\sigma^i\omega}\, d\overline{A}_{i,\omega} - \int \phi\, dP_{\mathrm{vol}}\right| \leq C_0^{-\beta} e^{-\lambda\beta/2n}\|\phi\|_{C^\beta} + 2C_\omega e^{i\epsilon} e^{-n\alpha/6}\|\phi\|_{C^\beta}, \tag{11.8}$$

which is what what we wanted for the pair $A_{i,\omega}$. The required tail bound on C_ω follows from Proposition 11.8 and (11.7) by taking D_1 sufficiently large because the first term involving C_0^β is uniformly bounded independent of $C_\omega \geq 1$. □

Theorem 11.10. *(Quenched, tempered equidistribution) Suppose that M is a closed surface, $(f_1, \ldots, f_m)$ is an expanding on average tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$, and $\beta \in (0,1)$ is a Hölder regularity. For any $\epsilon > 0$ there exists $\eta > 0$ such that for any R-good standard family $\hat{\gamma}$ with associated measure ρ, this family satisfies quenched, tempered equidistribution. Namely, for a.e. $\omega \in \Sigma$, there exists C_ω such that for any $\phi \in C^\beta(M)$, for all natural numbers k and n,*

$$\left|\int \phi \circ f^n_{\sigma^k(\omega)}\, d\rho - \int \phi\, d\,\mathrm{vol}\right| \leq C_\omega e^{k\epsilon} e^{-\eta n}\|\phi\|_{C^\beta}.$$

The above theorem is an immediate consequence of Proposition 11.9, so we do not write a separate proof of it. Next we turn to exponential mixing.

11.4. **Exponential mixing.** We are now ready to prove exponential mixing. In a subsequent paper we plan to show that several classical statistical limit theorems are valid in our setting.

Proof of Theorem 1.1. As before, let P_{vol} be an R-good standard family representing volume. We then apply Proposition 11.9 with $\hat{\gamma} = P_{\mathrm{vol}}$, and obtain $\lambda, \epsilon, \alpha > 0$ such that the conclusions of that proposition hold for these constants. Pick some $\omega \in \Sigma$ such that the conclusion of Proposition 11.9 holds for ω, and let C_ω be the associated constant. We will now show that f^n_ω is exponentially mixing. Let $\delta \in (0,1)$ be some fixed number small enough that $\epsilon\delta - (1-\delta)\alpha < 0$.

Below, we will be implicitly rounding to nearest integers so that everything makes sense. In particular, we will denote by $P_{\delta n}$ the standard family $P_{\lfloor \delta n \rfloor, \omega}$ from Proposition 11.9; as ω is fixed we will omit it below.

We now record some useful properties of $P_{\delta n}$. First, $P_{\delta n}$ comprises all but $C_\omega e^{-\delta\alpha n}$ of the mass of $f^{\delta n}_\omega(P_{\mathrm{vol}})$. Thus, by volume preservation:

$$\int \phi \cdot \psi \circ f^n_\omega\, dP_{\mathrm{vol}} = \int \phi \circ (f^{\delta n}_\omega)^{-1} \cdot \psi \circ f^{(1-\delta)n}_{\sigma^{\delta n}(\omega)}\, d(f^{\delta n}_\omega)_*(P_{\mathrm{vol}}) \tag{11.9}$$

$$= \sum_{A \in P_{\delta n}} \int \phi \circ (f^{\delta n}_\omega)^{-1} \cdot \psi \circ f^{(1-\delta)n}_{\sigma^{\delta n}(\omega)}\, dA \pm C_\omega e^{-\delta\alpha n}\|\phi\|_{C^\beta}\|\psi\|_{C^\beta}. \tag{11.10}$$

Now, by Proposition 11.9, the preimage of each curve $A \in P_{\delta n}$ has length at most $e^{-\delta\lambda n/2}$. By Hölder continuity of ϕ

$$\left|\max \phi \circ (f^{\delta n}_\omega)^{-1}|_A - \min \phi \circ (f^{\delta n}_\omega)^{-1}|_A\right| < e^{-\beta\delta\lambda n/2}\|\phi\|_{C^\beta}. \tag{11.11}$$

In particular, applying this observation to each summand in (11.10), we see that

$$\sum_{A \in P_{\delta n}} \int \phi \circ (f^{\delta n}_\omega)^{-1} \cdot \psi \circ f^{(1-\delta)n}_{\sigma^{\delta n}(\omega)}\, dA = \sum_{A \in P_{\delta n}} \int \phi \circ (f^{\delta n}_\omega)^{-1}\, d\overline{A} \int \psi \circ f^{(1-\delta)n}_{\sigma^{\delta n}(\omega)}\, dA$$

$$\pm\, e^{-n\beta\delta\lambda/2}\|\phi\|_{C^\beta}\|\psi\|_{C^\beta}, \tag{11.12}$$

where $\overline{A}$ denotes the unit mass version of A. By the exponential equidistribution estimate from Proposition 11.9,

$$\int \psi \circ f_{\sigma^{\delta n}(\omega)}^{(1-\delta)n}\, dA = \rho(A)\left(\int \psi \, d\,\mathrm{vol} \pm C_\omega e^{-((1-\delta)\alpha-\delta\epsilon)n}\|\psi\|_{C^\beta})\right), \tag{11.13}$$

where $\rho(A)$ is the mass of the pair A. Note by our choice of δ that the exponent appearing in the above equation is negative.

Combining (11.10), (11.13), and (11.12), we find that

$$\begin{aligned}\int \phi \cdot \psi \circ f_\omega^n \, dP_{\mathrm{vol}} = &\sum_{A\in P_{\delta n}} \left(\int \phi \circ (f_\omega^{\delta n})^{-1}\, dA\right)\left(\int \psi \, d\,\mathrm{vol}\right) \\ &\pm C_\omega(e^{-\delta\alpha n} + e^{-\beta\delta\lambda n/2} + e^{-((1-\delta)\alpha-\delta\epsilon)n})\|\phi\|_{C^\beta}\|\psi\|_{C^\beta}\end{aligned}$$

But as $P_{\delta n}$ comprises all but at most $C_\omega e^{-\delta\alpha n}$ of the mass of $f_\omega^n(P_{\mathrm{vol}})$, it follows that:

$$\begin{aligned}\int \phi \cdot \psi \circ f_\omega^n \, dP_{\mathrm{vol}} &= \left(\int \phi \, dP_{\mathrm{vol}} \pm C_\omega\|\phi\|_{C^\beta} e^{-\delta\alpha n}\right)\left(\int \psi \, d\,\mathrm{vol}\right) \\ &\quad \pm C_\omega(e^{-\delta\alpha n} + e^{-\beta\delta\lambda n/2} + e^{-((1-\delta)\alpha-\delta\epsilon)n})\|\phi\|_{C^\beta}\|\psi\|_{C^\beta} \\ &= \int \phi \, d\,\mathrm{vol} \int \psi \, d\,\mathrm{vol} \pm 4C_\omega(e^{-\eta n}\|\phi\|_{C^\beta}\|\psi\|_{C^\beta}),\end{aligned}$$

where $\eta = \min\{\delta\alpha, \beta\delta\lambda/2, (1-\delta)\alpha-\delta\epsilon\}$. Since the tail bound on C_ω is part of Proposition 11.9, the proof is complete. □

We now give the proof of annealed exponential mixing, i.e. exponential mixing of the skew product.

Proof of Corollary 1.2. Let $\bar{\Phi}(\omega) = \int_M \Phi(\omega, x)\, d\,\mathrm{vol}$, $\bar{\Psi}(\omega) = \int_M \Psi(\omega, x)\, d\,\mathrm{vol}$. Note that

$$\iint \Phi(\Psi \circ F^n)\, d\mu d\,\mathrm{vol} = \mathbb{E}_\omega\left(\Phi(\omega, x)\Psi(\sigma^n\omega, f_\omega^n x)\, d\,\mathrm{vol}\right).$$

Splitting the right hand side into the regions where $C_\omega \le e^{\eta n/2}$ and $C_\omega > e^{\eta n/2}$ and using (1.2) in the first region and (1.3) in the second region we obtain

$$\iint \Phi(\Psi \circ F^n)\, d\mu\, d\,\mathrm{vol} = \int \bar{\Phi}(\bar{\Psi} \circ \sigma^n) d\mu + O\left(e^{-\eta n/2}\|\Phi\|_{C^\beta}\|\Psi\|_{C^\beta}\right).$$

Now the result follows from the exponential mixing for the shift, see [PP90, Chapter 2]. □

Appendix A. Finite time smoothing estimates

In the following two appendices we present finite time estimates for nonuniformly hyperbolic systems. While such estimates should be familiar to experts in Pesin theory, it is difficult to find precise references in the literature since most works concentrate on infinite orbits. The finite time estimates play an important role in the paper because in the main coupling algorithm we want to use the independence of the dynamics, hence we decide to stop at time n based only on the dynamics on the time interval from zero to n.

A.1. **Finite time Lyapunov metrics.** Typically one defines Lyapunov metrics for an infinite sequence of diffeomorphisms. In our case have only a finite sequence, so we show that these also have Lyapunov metrics. The most important point in Lemma A.1 below is item (3), which tells us that at a reverse tempered point the Lyapunov metric will not be distorted.

The appearance of λ' in Lemma A.1 reflects that we need to make a small sacrifice in the rate of growth to obtain the uniform estimates. If we consider sequences that are (C,λ,ϵ)-tempered, and construct the Lyapunov metrics that guarantee a growth rate of exactly e^{λ} up to a factor of ϵ, then as we let ϵ go to zero, the Lyapunov metrics get very distorted with respect to the reference metrics. With the lemma below, as ϵ goes to zero the metrics do not get any more distorted, however, they guarantee only expansion at some rate $\lambda' \le \lambda$.

Lemma A.1. *(Lyapunov Metric Estimates) Fix (C,λ). Then for any $0<\lambda'\le\lambda$, and any sequence of linear maps $A_1,\dots,A_n\in \mathrm{SL}(2,\mathbb{R})$ that have a (C,λ,ϵ)-subtempered splitting, $E^s_i\oplus E^u_i$ with respect to a sequence of uniformly bounded reference metrics $\|\cdot\|_i$, there exists a sequence of metrics $\|\cdot\|'_i$ such that*

(1) $\|A_i|_{E^s}\|'_i \le e^{-\lambda'}$
(2) $\|A_i|_{E^u}\|'_i \ge e^{\lambda'}$
(3) $\frac{1}{\sqrt{2}}\|\xi\|_i \le \|\xi\|'_i \le 4e^{2C+2\epsilon i}\left(1-e^{2(\lambda'-\lambda)}\right)^{-1/2}\|\xi\|_i$, for $\xi\in\mathbb{R}^2$.

The same holds for reverse tempered sequences of maps, mutatis mutandis.

The estimates below are similar to [LQ95, Lem. III.1.3]. The reverse version follows by just taking inverses. This result holds because dropping terms from the definition of the Lyapunov metric doesn't stop them from satisfying the required estimates.

Proof. We begin by defining the new Lyapunov metric. Then we check the desired properties. For $\xi\in E^s_i$, let $\|\xi\|'_i = \left(\sum_{l=0}^{n-i}\|A^l_i\xi\|_i^2 e^{2\lambda' l}\right)^{1/2}$ and for $\xi\in E^u_i$, let $\|\xi\|'_i=\left(\sum_{l=0}^{i} e^{2\lambda' l}\|[A^l_{i-l}]^{-1}\xi\|^2_{i-l}\right)^{1/2}$.

We then define $\|\cdot\|'_i$ on all of $\mathbb{R}^2$ by declaring E^s_i and E^u_i to be orthogonal.

We now check the required estimate for the stable norm. Let $\xi\in E^s_i$, then

$$(\|A_i\xi\|'_{i+1})^2=\sum_{l=0}^{n-i-1}\|A^l_{i+1}A_i\xi\|^2e^{2\lambda' l}=\sum_{l=0}^{n-i-1}\|A^{l+1}_i\xi\|^2e^{2\lambda' l}$$
$$=e^{-2\lambda'}\sum_{l=0}^{n-i-1}\|A^{l+1}_i\xi\|^2e^{2\lambda'(l+1)}\le e^{-2\lambda'}(\|\xi\|'_i)^2.$$

Note that the last inequality follows because the penultimate expression is missing the first term in the sum that defines $\|\xi\|'_i$.

We now check the estimate on E^u_i. Suppose $\xi\in E^u_i$, $i<n$, then

$$(\|A_i\xi\|'_{i+1})^2=\sum_{l=0}^{i+1}e^{2\lambda' l}\|[A^l_{i+1-l}]^{-1}A_i\xi\|^2_{i+1-l}$$
$$=\|A_i\xi\|^2_{i+1}+e^{2\lambda'}\sum_{l=1}^{i+1}e^{2\lambda'(l-1)}\|[A^{l-1}_{i-(l-1)}]^{-1}\xi\|^2_{i-(l-1)}$$
$$=\|A_i\xi\|^2_{i+1}+e^{2\lambda'}\sum_{l=0}^{i}e^{2\lambda' l}\|[A^l_{i-l}]^{-1}\xi\|^2_{i-l}\ge e^{2\lambda'}(\|\xi\|'_i)^2.$$

This verifies the first two estimates in the lemma. Note that neither of the above required any control on the angle between E^s and E^u.

We now compare the two norms on E_i^s and E_i^u. For $\xi \in E_i^s$,

$$\|\xi\|_i'^2 = \sum_{l=0}^{n-i} \|A_i^l \xi\|_i^2 e^{2\lambda'} \le \sum_{l=0}^{n-i} e^{2C} e^{-2\lambda l} e^{2\epsilon i} \|\xi\|_i^2 e^{2\lambda' l} \le \frac{e^{2C} e^{2\epsilon i}}{1 - e^{2(\lambda' - \lambda)}} \|\xi\|_i^2.$$

Next for $\xi \in E_i^u$, we estimate

$$(\|\xi\|_i')^2 = \sum_{l=0}^{i} e^{2\lambda' l} \|[A_{i-l}^l]^{-1} \xi\|_{i-l}^2 = \sum_{l=0}^{i} e^{2\lambda' l} e^{2C} e^{2(i-l)\epsilon} e^{-2\lambda l} \|\xi\|_i^2$$

$$\le e^{2C} e^{2i\epsilon} \sum_{l=0}^{i} e^{2(\lambda' - \lambda) l} e^{-2\epsilon l} \|\xi\|_i^2 \le \frac{e^{2C} e^{i2\epsilon}}{1 - e^{2(\lambda' - \lambda)}} \|\xi\|_i^2.$$

We now check final estimate in the theorem. For the lower bound, note that by definition $\|\xi^s\|_i' \ge \|\xi^s\|_i$ and $\|\xi^u\|_i' \ge \|\xi^u\|_i$, thus

$$\|\xi\|_i^2 \le (\|\xi^s\|_i + \|\xi^u\|_i)^2 \le 2[(\|\xi^s\|_i')^2 + (\|\xi^u\|_i')^2] = 2(\|\xi\|_i')^2. \tag{A.1}$$

For the upper bound, we have that

$$\|\xi\|_i' \le \|\xi^s\|_i' + \|\xi^u\|_i' \le \frac{e^{C+\epsilon i}}{\sqrt{1 - e^{2(\lambda' - \lambda)}}} (\|\xi^s\|_i + \|\xi^u\|_i). \tag{A.2}$$

But we know from subtemperedness that the angle θ between E_i^s and E_i^u is at least $e^{-C} e^{-i\epsilon}$. So by the Law of Sines we have that for $* \in \{u, s\}$ that $\|\xi^*\|_i \le \|\xi\|_i / \sin\theta \le 2\|\xi\|_i/\theta$ because for $0 \le \theta \le \pi/2$, $\theta/2 \le \sin(\theta)$. Thus (A.2) gives $\|\xi\|_i' \le \dfrac{4e^{2C+2\epsilon i}}{\sqrt{1 - e^{2(\lambda' - \lambda)}}} \|\xi\|_i$, which completes the final estimate in the proof. □

A.2. **Basic calculus facts.** We now record some facts from calculus that will be needed when we study estimates for the graph transform. In the following statements, as elsewhere, we use $\|\phi\|_i$ to denote the supremum of norm of the ith partial derivatives of ϕ.

Lemma A.2. *(Norms of functions in twisted charts) Suppose that $\phi \colon \mathbb{R}^2 \to \mathbb{R}^2$ is a C^2 function. Then if we apply a linear change of coordinates L_1, L_2 to ϕ, then we see that*

$$\|L_2 \circ \phi \circ L_1\|_1 \le \|L_1\| \|L_2\| \|\phi\|_1.$$

Further, for the second derivatives of ϕ:

$$\|L_2 \circ \phi \circ L_1\|_2 \le \|L_2\| \|\phi\|_2 \|L_1\|^2.$$

The next lemma studies how the C^2 norm of a curve changes when we apply a linear map.

Lemma A.3. *Suppose that γ is a C^2 curve in $\mathbb{R}^2$ and that $L \colon \mathbb{R}^2 \to \mathbb{R}^2$ is an invertible linear map. Then $\|L \circ \gamma\|_{C^2} \le \dfrac{\|L\|}{(m(L))^2} \|\gamma\|_{C^2}$. Here $\|\gamma\|_{C^2}$ refers to the C^2 norm of γ as a curve in $\mathbb{R}^2$ and $m(L)$ is the conorm of the matrix, $m(L) = \min_{v \neq 0} \|Lv\|/\|v\|$.*

Proof. By definition, the C^2 norm of a curve is the supremum of the second derivative of its graph over each of its tangent spaces. So, without loss of generality suppose that γ passes through the origin and that at this point γ is the curve $t \mapsto (t, \lambda t^2)$ ($O(t^3)$ terms do not change the computation below). Then we apply $L = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$ to $(t, \lambda t^2)^T$ to get the curve $t\begin{pmatrix} a \\ c \end{pmatrix} + \lambda t^2 \begin{pmatrix} b \\ d \end{pmatrix}$.

To study the C^2 norm of $L \circ \gamma$ at 0, we must write it as a graph over its tangent space, i.e. in the form $tu + t^2\hat{\lambda}u^{\perp}$, where u is a unit vector and $\hat{\lambda}$ is to be determined. Let $v = (a, c)^T$, $u = v/\|v\|$ and $w = (b, d)^T$. Then we may reparametrize $vt+\lambda wt^2$ in the form $ut+\lambda(w/\|v\|^2)t^2$. Decomposing $w = pu + qu^{\perp}$ we obtain the parametrization $us + (\lambda q/\|v\|^2)s^2u^{\perp} + O(s^3)$ where $s = t + p\lambda/\|v\|^2t^2$. Thus $\hat{\lambda} = q\lambda/\|v\|^2$. Since $|q| \le \|w\|$, $\|w\| \le \|L\|$, and $1/\|v\| \le 1/m(L)$, the result follows. □

We now estimate the C^2 norm of a function in terms of its inverse.

Lemma A.4. *Suppose that $\phi\colon \mathbb{R} \to \mathbb{R}$ (or from one interval to another) is a C^2 diffeomorphism. If $|D\phi| > \lambda$, then $\left|D\phi^{-1}\right| \le \lambda^{-1}$ and $\|\phi^{-1}\|_2 \le \lambda^{-3}\|\phi\|_2$.*

Proof. At each point, we express the Taylor polynomial of ψ^{-1} in terms of the Taylor polynomial of ψ. Suppose that ψ has Taylor polynomial $\nu x + Ax^2$ at some point, with $|\nu| \le \lambda$. Then the Taylor polynomial of ψ^{-1} at the corresponding point is $\nu^{-1}x + Cx^2$, where $C = -\nu^{-3}A$. The conclusion follows. □

For the future reference, we record a bound on compositions. An overview of estimates like these is contained in [Hör76, App. A].

Lemma A.5. *Suppose we are composing three functions $f, g, h\colon \mathbb{R}^n \to \mathbb{R}^n$, then*

$$\|f \circ g\|_2 \le \|f\|_2\|g\|_1^2 + \|f\|_1\|g\|_2.$$

and

$$\|f \circ g \circ h\|_2 \le \|f\|_2\|g\|_1^2\|h\|_1^2 + \|f\|_1\|g\|_2\|h\|_1^2 + \|f\|_1\|g\|_1\|h\|_2.$$

When we study how fast the dynamics smooths curves, we will represent the curve as a graph and then apply the graph transform to it. The following relates the C^2 norm of an embedded curve with the C^2 norm of the curve represented as a graph. Recall that the C^2 norm of an embedded curve is the same thing as the norm of the curve as a graph over its tangent space at each point in an exponential chart.

Lemma A.6. *Suppose γ is a C^2 curve in $\mathbb{R}^2$ that is θ-transverse to the y-axis. Then if we represent γ as the graph over the x-axis of a function $\hat{\gamma}$, then*

$$\|\hat{\gamma}\|_1 \le \cot\theta, \text{ and } \|\hat{\gamma}\|_2 \le (\sin\theta)^{-3}\|\gamma\|_{C^2}.$$

Proof. The first estimate is essentially the definition of tangent, so we will show the second.

Locally we may represent γ as a graph:

$$p+(\sin\theta_p, \cos\theta_p)t+\phi_p(t)(-\cos\theta_p, \sin\theta_p)=p+(t\sin\theta_p-\phi_p(t)\cos\theta_p, 0)+(0, t\cos\theta_p+\phi_p(t)\sin\theta_p)$$

where $\phi_p'(t) = 0$. By definition of $\|\gamma\|_{C^2}$, $\left|\phi_p''(0)\right| \le \|\gamma\|_{C^2}$.

In order to estimate $\hat{\gamma}''(0)$, we must write the graph in the form $p + (t, \psi(t))$ for some ψ and estimate $\psi''(0)$. Accordingly, we make a change of variables $s = t/\sin\theta_p$ getting

$$p + \left(s - \phi_p\left(\frac{s}{\sin\theta_p}\right)\cos\theta_p, 0\right) + \left(0, \frac{\cos\theta_p}{\sin\theta_p}s + \phi_p\left(\frac{s}{\sin\theta_p}\right)\sin\theta_p\right). \tag{A.3}$$

To estimate the second derivative of the graph at 0, we need a representation of the form $(u + O(u^3), \psi(u) + O(u^3))$, so we make a further change of variables $u = s - \phi_p\left(\dfrac{s}{\sin\theta_p}\right)$.

Then $s = u + \dfrac{\phi_p''(0)u^2\cos\theta_p}{2\sin^2\theta_p} + o(u^2)$. Plugging this into (A.3) and using that $\cos^2\theta_p + \sin^2\theta_p = 1$ we obtain the parametrization

$$p + \left(u, \frac{u\cos\theta_p}{\sin\theta_p} + \frac{\phi''(0)u^2}{2\sin^3\theta_p} + o(u^2)\right)$$

and the result follows. □

The next lemma estimates how the density is distorted by diffeomorphisms.

Lemma A.7. *Suppose that M is a closed Riemannian manifold. There exists $C > 0$ such that if $f\colon M \to M$ is a C^2 diffeomorphism, γ is a C^2 curve in M and ρ is a log-α-Hölder density along γ, then the density $f_*\rho$ along $f(\gamma)$ satisfies*

$$\|\ln(f_*\rho)\|_{C^\alpha} \le (1/m(Df))^{1+\alpha}\left(\|\ln\rho\|_{C^\alpha} + C\|f\|_{C^2}(1 + \|\gamma\|_{C^2})\right). \tag{A.4}$$

The same estimate holds for local diffeomorphisms, mutatis mutandis.

We leave the proof of the lemma to the readers, since we provide a similar estimate below (see (A.25)).

Next we record an estimate comparing two inner products.

Lemma A.8. *Suppose that we have two inner products $\|\cdot\|_1$ and $\|\cdot\|_2$ on a vector space V and that*

$$A\|\cdot\|_1 \le \|\cdot\|_2 \le B\|\cdot\|_1.$$

Then for $v, w \in V\setminus\{0\}$

$$AB^{-1}\angle_1(v,w) \le \angle_2(v,w) \le A^{-1}B\angle_1(v,w),$$

where $\angle_i$ denotes the angle with respect to the metric $\|\cdot\|_i$.

Proof. We show the upper bound; the lower bound is a straightforward consequence. Let S_i^1 denote the unit sphere with respect to the inner product i and v and w be two unit vectors with respect to $\|\cdot\|_1$. Let I be a curve between v and w such that $\mathrm{len}_1(I) = \angle_1(v,w)$. Then $\mathrm{len}_2(I) \le B\,\mathrm{len}_1(I)$. Let $\pi_2\colon V\setminus\{0\} \to S_2^1$ denote the radial projection onto S_2^1. Then $\angle_2(v,w) \le \mathrm{len}_2(\pi_2(I))$. Note that the norm of $D\pi_2|_I$ is bounded above by $1/d_2(0,I)$. Since $d_2(0,I) \ge A$, we see that $\mathrm{len}_2(\pi_2(I)) \le A^{-1}B\,\mathrm{len}_1(I)$, so we are done. □

We now record an estimate on how fast a C^2 curve can get worse under the dynamics. Note that one iteration can instantaneously make a line into an $O(1)$ bad curve, hence the estimate has the form below.

Lemma A.9. *Fix $D > 0$ then there exists $\Lambda > 0$, such that if for $1 \le i \le n$, $f_i \in \mathrm{Diff}^2(M)$ is a sequence of diffeomorphisms of a closed Riemannian manifold M with $\|f\|_{C^2} < D$, γ is a C^2 curve in M, and $\gamma_n = f_1^n(\gamma)$, then*

$$\|\gamma_n\|_{C^2} \le \max\{e^{\Lambda n}\|\gamma\|_{C^2}, e^{\Lambda n}\}.$$

Proof. Recall that the C^2 norm of γ is bounded by the maximum over all $t \in \gamma$ of the second derivative of γ in an exponential chart at t where γ is viewed as a graph over its tangent plane. The result then follows because the second derivative of a sequence of maps with uniformly bounded C^2 norm grows at most exponentially fast. □

A.3. **Properties of Hölder functions.** In this subsection, we record some additional claims about Hölder and log-Hölder functions that will be used in the proof of the coupling lemma.

Claim A.10. *Suppose that* $\rho \colon M \to \mathbb{R}$ *is a* (C,α)*-Hölder function on a metric space* M *such that* $\rho \geq A^{-1}$*, for some* $A > 0$*. Then* $\ln \rho$ *is* AC*-log-*α*-Hölder.*

Proof. First, observe that on $[A^{-1}, \infty)$, that ln is A-Lipschitz because its derivative $1/x$ is at most A. Thus $|\ln(\rho(x)) - \ln(\rho(y))| \leq A\,|\rho(x) - \rho(y)| \leq AC\,|x-y|^{\alpha}$, as desired. □

The next lemma relates two different ways of dealing with log-Hölder functions.

Lemma A.11. *Suppose that* ρ *is an* (A,α)*-log Hölder function on a metric space of diameter at most* D*. Then there exists* $C_{A,D}$ *such that*

$$|\rho(x) - \rho(y)| \leq \rho(x) C_{A,D}\,|x-y|^{\alpha}\,. \tag{A.5}$$

Proof. Suppose that $\rho(y) \geq \rho(x)$. Then log-α-Hölder gives that

$$\ln(\rho(y)/\rho(x)) = |\ln(\rho(y)/\rho(x))| \leq A\,|x-y|^{\alpha}\,.$$

Thus taking e^x, by boundedness of the metric space and the constant A, there exists $C_{A,D}$ such that

$$\frac{\rho(y)}{\rho(x)} \leq e^{A|x-y|^{\alpha}} \leq 1 + C_{A,D}\,|x-y|^{\alpha}\,.$$

Thus

$$\rho(y) - \rho(x) \leq \rho(x) C_{A,D}\,|x-y|^{\alpha}\,.$$

The case when $\rho(y) < \rho(x)$ is similar, so we are done. □

A.4. **Graph transform with estimates on the second derivative.** We now study the graph transform and record how C^2 norms of curves are affected by it. If one constructs the stable manifolds by using the graph transform, then after one has checked that the stable manifold is C^1, one can check that the manifolds are C^r inductively by studying the action of the graph transform on the jet of the stable manifold which is C^1. See for instance the construction in [Shu87], which proceeds along these lines.

Proposition A.12. *(*C^2 *estimates for the graph transform) Suppose* $\lambda > 1$ *and* $F \colon \mathbb{R}^2 \to \mathbb{R}^2$ *is a* C^2 *diffeomorphism of the form*

$$F = (\sigma_1 x + f_1(x,y), \sigma_2 y + f_2(x,y)), \tag{A.6}$$

with $\min\{\sigma_1, \sigma_2^{-1}\} \geq \lambda$*. Suppose that* γ *is a* C^2 *curve given as the graph of a function* $\phi \colon I_1 \to \mathbb{R}$*. Assume that* $F(0,0) = (0,0)$ *and that we have the following estimates:*

$$\|f_1\|_{C^1} = \epsilon_1, \tag{A.7}$$

$$\|f_2\|_{C^1} = \epsilon_2 < \lambda^{-1}, \tag{A.8}$$

$$\lambda - \epsilon_1 - \epsilon_1\|\phi\|_1 > 0. \tag{A.9}$$

Then the following hold.

(1) *The curve* $F \circ \gamma$ *is given as the graph of a function* $\widetilde{\phi} \colon I_2 \to \mathbb{R}$ *and*

$$\operatorname{len}(I_2) \geq (\lambda - \epsilon_1 - \epsilon_1\|\phi\|_1)\operatorname{len}(I_1). \tag{A.10}$$

(2) *We have an estimate on how much* F *smooths* ϕ*,*

$$\|\widetilde{\phi}\|_{C^0} \leq \lambda^{-1}\|\phi\|_{C^0} + \epsilon_2, \tag{A.11}$$

$$\|\widetilde{\phi}\|_1 \leq (\lambda^{-1}\|\phi\|_1 + \epsilon_2 + \epsilon_2\|\phi\|_1)(\lambda - \epsilon_1 - \epsilon_1\|\phi\|_1)^{-1}. \tag{A.12}$$

(3) There is $\epsilon_0>0$ such that under the additional assumption that $\epsilon_1, \epsilon_2, \|\phi\|_1<\epsilon_0$

$$\|\widetilde{\phi}\|_2 \le \lambda^{-1.99}\|f\|_2 + \lambda^{-2.99}\|\phi\|_2. \tag{A.13}$$

(4) The graph transform smooths densities along curves. If $\rho(x,\phi(x))$ is a log α-Hölder density along γ with respect to the arclength, write $\widetilde{\rho}(x,\widetilde{\phi}(x))$ for the density of the pushforward of ρ along $F(\gamma)$. For $\epsilon_0 > 0$ as in (3), if $\epsilon_1, \epsilon_2, \|\phi\|_1 < \epsilon_0$, then

$$\|\ln\widetilde{\rho}\|_{C^\alpha} \le \lambda^{-.9\alpha}(\|\ln\rho\|_{C^\alpha} + \|f\|_2 + \|\phi\|_2). \tag{A.14}$$

Note that part (1) of the proposition implies that if I_1 contains a neighborhood of 0 of size δ, then then I_2 contains a neighborhood of size $(\lambda-\epsilon_1-\epsilon_1\|\phi\|_1)\delta$.

Proof. We write down explicitly a formula for $\widetilde{\phi}$ and then estimate each term that appears in the formula. It is tedious but straightforward. Throughout we will use π_1 and π_2 for the projections onto the two factors in $\mathbb{R}^2$.

We estimate the C^1 norm of ϕ as a graph over $\mathbb{R}\times\{0\}$. To this end we first study how much the graph of ϕ is stretched horizontally, which will verify (1) above. To do this we consider a natural map $\psi^{-1}\colon I_1\to\mathbb{R}$:

$$\psi^{-1}\colon x\mapsto (x,\phi(x))\mapsto \pi_1(F(x,\phi(x))) = \lambda x + f_1(x,\phi(x)). \tag{A.15}$$

From the definition of ψ^{-1},

$$\|D\psi^{-1}\| \ge \lambda - \epsilon_1 - \epsilon_1\|\phi\|_1, \tag{A.16}$$

thus by (A.9), $\|D\psi^{-1}\|$ is positive, so ψ^{-1} is monotone. Hence $F(\gamma)$ is the graph of a function $\widetilde{\phi}$, and we may write $\psi^{-1}\colon I_1\to I_2$. By (A.16), $(\lambda-\epsilon_1-\epsilon_1\|\phi\|_1)\operatorname{len}(I_1)\le\operatorname{len}(I_2)$. This completes the proof of item (1).

We now prove item (2). First we give the C^0 estimate and then the estimate on the first derivative. By the assumption on f_2, we see that the image of ϕ is at most $\lambda^{-1}\|\phi\|_{C^0}+\epsilon_2$ from the x-axis. Thus

$$\|\widetilde{\phi}\|_{C^0} \le \lambda^{-1}\|\phi\|_{C^0} + \epsilon_2. \tag{A.17}$$

Now we estimate $\|\widetilde{\phi}\|_1$. From equation (A.16), we obtain that:

$$\|D\psi\| \le (\lambda-\epsilon_1-\epsilon_1\|\phi\|_1)^{-1}. \tag{A.18}$$

This allows us to estimate the C^1 norm of $F\circ\gamma$ as a graph over $I_2\times\{0\}$. The curve $F(\gamma)$ is given by the graph of

$$x\mapsto \pi_2F(\psi(x),\phi(\psi(x))) = \lambda^{-1}\phi(\psi(x)) + f_2(\psi(x),\phi(\psi(x))) = \widetilde{\phi}. \tag{A.19}$$

Thus by the chain rule

$$\|\widetilde{\phi}\|_1 \le \lambda^{-1}\|\phi\|_1(\lambda-\epsilon_1-\epsilon_1\|\phi\|_1)^{-1} + \epsilon_2(\lambda-\epsilon_1-\epsilon_1\|\phi\|_1)^{-1} + \epsilon_2\|\phi\|_1(\lambda-\epsilon_1-\epsilon_1\|\phi\|_1)^{-1}.$$

Hence,

$$\|\widetilde{\phi}\|_1 \le (\lambda^{-1}\|\phi\|_1 + \epsilon_2 + \epsilon_2\|\phi\|_1)(\lambda-\epsilon_1-\epsilon_1\|\phi\|_{C^1})^{-1}, \tag{A.20}$$

which finishes the proof of item (2).

We now turn to the C^2 estimates and check item (3). To begin we need to obtain a C^2 estimate on the function ψ used above. By (A.15) and the chain rule,

$$\|\psi^{-1}\|_2 \le \|f\|_2 + 2\|f\|_2\|\phi\|_1 + \|f\|_2\|\phi\|_1^2 + \|f\|_1\|\phi\|_2.$$

Thus by Lemma A.4,

$$\|\psi\|_2 \le (\lambda-\epsilon_1-\epsilon_1\|\phi\|_1)^{-3}(\|f\|_2 + 2\|f\|_2\|\phi\|_1 + \|f\|_2\|\phi\|_1^2 + \|f\|_1\|\phi\|_2).$$

We can now plug everything in to estimate the C^2 norm of the image of ϕ. By definition $\widetilde{\phi}$ is equal to $\lambda^{-1}\phi(\psi(x)) + f_2(\psi(x), \phi(\psi(x)))$. For the first term, we have the estimate

$$\|\lambda^{-1}\phi \circ \psi\|_2 \le \lambda^{-1}(\|\phi\|_2\|\psi\|_1^2 + \|\phi\|_1\|\psi\|_2).$$

By the chain rule

$$\begin{aligned}\|f_2(\psi(x), \phi(\psi(x)))\|_2 \le \|f_2\|_2\|\psi\|_1^2 + \|f_2\|_2\|\phi\|_1\|\psi\|_1^2 + \|f_2\|_1\|\phi\|_1\|\psi\|_2 \\ + (\|f_2\|_1\|\phi\|_1\|\psi\|_2 + \|f_2\|_1\|\phi\|_2\|\psi\|_1^2 + \|f_2\|_2\|\phi\|_1^2\|\psi\|_1^2 + \|f_2\|_2\|\phi\|_1\|\psi\|_1^2).\end{aligned}$$

Hence if $\epsilon_1, \epsilon_2, \|\phi\|_1 < \epsilon_0$ and ϵ_0 sufficiently small, then

$$\|\psi\|_1 \le \lambda^{-.9999}, \tag{A.21}$$

$$\|\psi\|_2 \le \lambda^{-2.999}(\epsilon_0\|\phi\|_2 + \|f\|_2). \tag{A.22}$$

In particular, as long as $\epsilon_0 > 0$ is sufficiently small, under the assumptions just listed applying the estimates on $\|\phi\|_1$ and $\|\phi\|_2$ gives

$$\|\lambda^{-1}\phi \circ \psi\|_2 \le \lambda^{-2.999}(\epsilon_0\|f\|_2 + \|\phi\|_2), \tag{A.23}$$

$$\|f_2(\psi(x), \phi(\psi(x)))\|_2 \le \lambda^{-1.999}\|f\|_2 + \epsilon_0\lambda^{-1.8}\|\phi\|_2. \tag{A.24}$$

Combining these estimates, we see that as long as ϵ_0 is sufficiently small,

$$\|\widetilde{\phi}\|_2 \le \lambda^{-1.99}\|f\|_2 + \lambda^{-2.99}\|\phi\|_2.$$

We next study how the Hölder norm of the log of the density ρ along γ changes when we iterate the dynamics and prove item (4). From the change of variables formula, we must estimate the following:

$$\begin{aligned}&\ln[\rho(\psi(x), \phi(\psi(x)))\|DF|_{(1,d\phi/dx)}(\psi(x), \phi(\psi(x)))\|^{-1}] = \\ &\ln\rho(\psi(x), \phi(\psi(x))) + \ln\|DF|_{(1,d\phi/dx)}(\psi(x))\|^{-1} = I + II.\end{aligned} \tag{A.25}$$

Term I. The estimate of the term I is straightforward:

$$\|\ln\rho(\psi(x), \phi(\psi(x)))\|_{C^\alpha} \le \|\psi\|_1^\alpha\|\ln\rho\|_{C^\alpha} \le \lambda^{-.9\alpha}\|\ln\rho\|_{C^\alpha}$$

by equation (A.18) as we are assuming $\|f_1\|_1, \|f_2\|_1, \|\phi\|_1$ are all small.

Term II. The second term is more complicated to estimate. Note that this term does not actually involve ρ as it is just the Jacobian of the map between two curves. So, to control the log-α-Hölder norm of this function we can estimate the derivative of the logarithm, which is an upper bound on the log-α-Hölder constant for all $\alpha \le 1$. To begin, we write

$$\begin{aligned}D\ln\|DF|_{(1,d\phi/dx)}(\psi(x), \phi(\psi(x)))\|^{-1} &= 2^{-1}D\ln\|DF|_{(1,d\phi/dx)}\|^2 - 2^{-1}D\ln\|(1, d\phi/dx)\|^2 \\ &= III + IV,\end{aligned}$$

where III and IV are evaluated at the point $(\psi(x), \phi(\psi(x))$.

Term III. We now bound term III. Because $D\ln f\circ\psi = \psi' D\ln f$, we see that the required estimate will hold assuming that it holds without precomposing with ψ because $|\psi'| \le 1$ under these assumptions. Thus we suppress the ψ below. From before we have an expression for DF in terms of λ, f_1, f_2:

$$DF = \begin{bmatrix} \lambda + \frac{df_1}{dx} & \frac{df_1}{dy} \\ \frac{df_2}{dx} & \lambda^{-1} + \frac{df_2}{dy} \end{bmatrix}.$$

Thus we are reduced to evaluating

$$D\ln\left[\left(\lambda + \frac{df_1}{dx} + \frac{df_1}{dy}\frac{d\phi}{dx}\right)^2 + \left(\frac{df_2}{dx} + \lambda^{-1}\frac{d\phi}{dx} + \frac{df_2}{dx}\frac{d\phi}{dx}\right)^2\right], \tag{A.26}$$

where the df_i terms are evaluated at $(x, \phi(x))$. Then taking derivatives gives:

$$\frac{A+B}{\left(\lambda + \frac{df_1}{dx} + \frac{df_1}{dy}\frac{d\phi}{dx}\right)^2 + \left(\frac{df_2}{dx} + \lambda^{-1}\frac{d\phi}{dx} + \frac{df_2}{dx}\frac{d\phi}{dx}\right)^2} = \frac{A}{Q} + \frac{B}{Q}. \tag{A.27}$$

where A and B are the derivatives of the two parenthetical terms in equation (A.26) and Q is the denominator of the left hand side of equation (A.27). Note that Q can be made arbitrarily close to λ^2 as long as $df_1/dx, df_1/dy$ and $d\phi/dx$ are sufficiently small.

Keeping in mind that the f_i terms are evaluated at $(x, \phi(x))$, we find that:

$$A = 2\left(\lambda + \frac{df_1}{dx} + \frac{df_1}{dy}\frac{d\phi}{dx}\right)\left(\frac{d^2 f_1}{dx^2} + \frac{d^2 f_1}{dxdy}\frac{d\phi}{dx} + \left(\frac{d^2 f_1}{dxdy} + \frac{d^2 f_1}{dy^2}\frac{d\phi}{dx}\right)\frac{d\phi}{dx} + \frac{df_1}{dy}\frac{d^2\phi}{dx^2}\right)$$

and

$$B{=}2\left(\frac{df_2}{dx}{+}\lambda^{-1}\frac{d\phi}{dx}{+}\frac{df_2}{dx}\frac{d\phi}{dx}\right)\left(\frac{d^2 f_2}{dx^2} + \lambda^{-1}\frac{d^2\phi}{dx^2} + \left(\frac{d^2 f_2}{dxdy} + \frac{d^2 f_2}{dx^2}{+}\frac{d^2 f_2}{dxdy}\frac{d\phi}{dx}\right)\frac{d\phi}{dx}{+}\frac{df_2}{dx}\frac{d^2\phi}{dx^2}\right).$$

Pick a small number $\bar\epsilon$. Then as $\|f_1\|_1, \|f_2\|_1$ and $\|\phi\|_1$ are sufficiently small it is easy to see from the above expressions for A and B, that

$$|III| \le \frac{\lambda + \bar\epsilon}{\lambda^2 - \bar\epsilon}(\|f\|_2 + \|\phi\|_2). \tag{A.28}$$

Term IV. We now bound Term IV. For this term we have

$$2^{-1} D \ln \|(1, d\phi/dx)\|^2 = 2^{-1} D \ln\left(1 + \left(\frac{d\phi}{dx}\right)^2\right) = \frac{\frac{d\phi}{dx}\frac{d^2\phi}{dx^2}}{1 + \left(\frac{d\phi}{dx}\right)^2}.$$

Since we are assuming $\|\phi\|_1$ is small, we see that $|IV| \le \bar\epsilon \|\phi\|_2$.

Conclusion of estimates on $|D\ln\widetilde\rho|$. From the above discussion,

$$\|\ln\widetilde\rho(x, \widetilde\phi(x))\|_{C^\alpha} \le |I| + |III| + |IV|$$
$$\le \lambda^{-.9\alpha}\|\ln\rho\|_{C^\alpha} + \left[\frac{\lambda+\bar\epsilon}{\lambda^2 - \bar\epsilon} + \bar\epsilon\right](\|f\|_2 + \|\phi\|_2) \le \lambda^{-.9\alpha}(\|\ln\rho\|_{C^\alpha} + \|f\|_2 + \|\phi\|_2).$$

where the last inequality holds since the expression in square brackets is less than 1 provided that $\bar\epsilon$ is sufficiently small. This concludes the proof of the proposition. □

A.5. **Finite time smoothing estimate.** Now that we control the amount of smoothing due to a single iteration of the graph transform, we study a reverse subtempered point for a sequence of diffeomorphisms. An important feature of the estimate below is that it covers curves that are extremely close to the contracting direction. This complicates the estimates compared to the case that one only considers curves lying in a cone near the expanding direction.

Proposition A.13. *Fix constants $C, \lambda, \epsilon, D_1 > 0$ with $\epsilon < \lambda/30$. Suppose that $f_i \colon \mathbb{R}^2 \to \mathbb{R}^2$, $1 \le i \le n$, is a sequence of diffeomorphisms such that $f_i(0) = 0$, the sequence $D_0 f_i$ has a (C, λ, ϵ)-reverse tempered splitting E^s_i, E^u_i in the sense of Definition 4.2, $\|f_i\|_{C^2} < D_1$, and $\|Df_i^{-1}\| \ge D_1^{-1}$. Then there exist constants $\epsilon_0, \ell_{\max}, D_2, D_3, D_4, D_5, D_6, D_7, D_8, C_0$ depending only on (C, λ, ϵ) and D_1 such that the following holds. Let γ be a C^2 curve in $\mathbb{R}^2$ passing through 0 not tangent to E^s_0 at 0, containing an R-good neighborhood of 0. Let $f^n = f_n \circ \cdots \circ f_1$. Let $\theta = \angle(\dot\gamma(0), E^s_0)$, and γ_0 be a segment of γ containing 0 of length at least*

$$\operatorname{len}(\gamma_0) = D_2 \min\{e^{-R}\theta, e^{-.9\lambda n}\}. \tag{A.29}$$

There is an associated auxiliary quantity

$$l_0 = D_3\theta e^{-2\epsilon n}\min\{e^{-R}\theta, e^{-.9\lambda n}\}, \tag{A.30}$$

and a subcurve γ_n of $f^n(\gamma_0)$ containing 0 such that the following hold:

(1) The curve γ_n has length at least

$$\ell_{out} = \min\{l_0 e^{.9\lambda n}, \ell_{\max}\}.$$

(2) If the minimum in item (1) is realized by $\ell_{\max}$, then the preimage of γ_n in γ has length at most $D_4 e^{-.9\lambda n}$, and this occurs as long as

$$n \geq D_5 + \frac{\max\{R,0\} - 2\ln(\theta)}{.99\lambda}.$$

Further, in this case, the preimage of γ_n in $f^i(\gamma)$ has length at most $D_4 e^{-.9\lambda(n-i)}$. In fact if $I \subseteq f^i(\gamma)$ is a curve of length at least $D_4 e^{-.85\lambda(n-i)}$ containing a point $f^i(x)$, then $f^{n-i}(I)$ contains a C_0-good neighborhood of $f^n(x)$.

(3) On γ_n, we have the estimate:

$$\|\gamma_n\|_{C^2} < D_6 e^{-2.9\lambda n} e^{D_7 \ln\theta}\max\{\|\gamma\|_{C^2}, 1\} + D_8. \tag{A.31}$$

(4) Finally, for any arbitrarily large $D_9 > 0$ and fixed α, there exist D_{10}, D_{11} such that the following holds. Suppose that ρ is a density along γ that is log-α-Hölder. Then for the same collection of n, the density of $\rho_n = (f^n)_*(\rho)$ along γ_n with respect to arclength parametrization of γ_n satisfies the following estimate, as long as $\|\gamma_n\|_2 < D_9$,

$$\|\ln\rho_n|_{\gamma_n}\|_{C^\alpha} \leq D_{10} e^{-.9\alpha\lambda n} e^{D_7\ln\theta}(1 + \|\ln\rho\|_{C^\alpha} + \|\gamma\|_{C^2}) + D_{11}. \tag{A.32}$$

The analogous statement holds for sequences of local diffeomorphisms f_i defined on a sequence of neighborhoods of 0 in $\mathbb{R}^2$ or of a closed manifold.

Proof. We begin by fixing some notation and constants that we will use throughout the argument. Let $\lambda' = .999\lambda$. Then from Lemma A.1 we obtain finite time Lyapunov metrics $\|\cdot\|_i'$, $0 \leq i \leq n$, associated to this splitting that satisfy for all $\xi \in \mathbb{R}^2$:

$$\frac{1}{\sqrt{2}}\|\xi\|_i \leq \|\xi\|_i' \leq 4e^{2C+2\epsilon(n-i)}\left(1 - e^{2(\lambda'-\lambda)}\right)^{-1/2}\|\xi\|_i. \tag{A.33}$$

Note that because the sequence is reverse tempered $\|\cdot\|_n'$ is uniformly comparable to the original metric independent of n. As is standard, the metrics $\|\cdot\|_i'$ give new linear coordinates $L_i\colon \mathbb{R}^2 \to \mathbb{R}^2$ that satisfy that $(L_i)^*\|\cdot\|_i = \|\cdot\|_i'$. We let $\hat{f}_i = L_{i+1}\circ f_i \circ L_i^{-1}$. Thus from properties of the Lyapunov metric, $D_0\hat{f}_i$ is a uniformly hyperbolic sequence satisfying

$$D_0\hat{f}_i|_{E_i^u} \geq e^{.999\lambda}, \quad D_0\hat{f}_i|_{E_i^s} \leq e^{-.999\lambda}. \tag{A.34}$$

We write:

$$\hat{f}_i(x,y) = (\sigma_{1,i}x + \hat{f}_{i,1}(x,y), \sigma_{2,i}y + \hat{f}_{i,2}(x,y)), \tag{A.35}$$

where $D_0\hat{f}_i = \operatorname{diag}(\sigma_{1,i},\sigma_{2,i})$ and $\sigma_{i,1}, \sigma_{i,2}^{-1} \geq e^{.999\lambda}$.

We now record estimates on C^2 norms in these charts. By (A.33), there is C_1 such that:

$$\max\{\|L_i\|, \|L_i^{-1}\|\} \leq C_1 e^{2C+2\epsilon(n-i)}. \tag{A.36}$$

Thus by Lemma A.2, for $1 \leq i \leq n$,

$$\|\hat{f}_i\|_{C^2} \leq D_1 e^{6C} e^{6(n-i)\epsilon}. \tag{A.37}$$

For $0 \leq i \leq n$, let

$$r_i = C_2 \min\{\theta\|\gamma\|_2^{-1} e^{.9\lambda i}, e^{-.9\lambda(n-i)}\}, \tag{A.38}$$

where $0 < C_2 < 1$ is a small number that we will choose later. We then restrict to studying the segment of γ inside the cube B_i centered at $0 \in \mathbb{R}^2$ of side length r_i with respect to the $\|\cdot\|_i'$ metric. Let γ_i be the connected component of 0 in $f^i(\gamma) \cap B_i$. We write $\hat{\gamma}_i$ for the function giving γ_i as a graph over the x-axis and let l_i be the length of the projection of $\hat{\gamma}_i$ to the x-axis in $\mathbb{R}^2$ measured with respect to $\|\cdot\|_i'$.

We begin working with the ambient metric. By the mean value theorem, there exists C_3 such that for a C^2 curve γ in $\mathbb{R}^2$ in an arclength parametrization,

$$\angle(\dot{\gamma}(t), \dot{\gamma}(s)) \leq C_3\|\gamma\|_{C^2} |t-s|,$$

because $\dot{\gamma}(t)$ is orthogonal to $\ddot{\gamma}$. In particular, as our curve γ satisfies $\angle(\dot{\gamma}(0), E_0^s) > \theta$ restricted to a segment of γ of length $C_3^{-1}\|\gamma\|_{C^2}^{-1}\theta/2$ around $\gamma(0)$, that on this segment $\angle(E_0^s, \dot{\gamma}(t)) > \theta/2$. Then from Lemma A.8 in the Lyapunov chart we have that, letting $\angle'$ denote angle with respect to the Lyapunov metric, there exists C_4 such that:

$$C_4^{-1}\theta e^{-2\epsilon n} \leq \angle'(E_0^s, \dot{\gamma}(t)) \leq C_4\theta e^{2\epsilon n}. \tag{A.39}$$

From the construction of the Lyapunov metric, $\frac{1}{\sqrt{2}}\|\cdot\|_i \leq \|\cdot\|_i'$, thus the length of γ in the Lyapunov chart is at least $e^{-R}/2$. We now restrict to a segment of $\hat{\gamma}$, which we call $\hat{\gamma}_0$, with length with respect to the Lyapunov metric:

$$\operatorname{len}'(\hat{\gamma}_0) = \min\{C_3^{-1}e^{-R}\theta/2, r_0\}. \tag{A.40}$$

From (A.33), as the ratio of $\|\cdot\|_n$ to $\|\cdot\|_n'$ does not depend on n, we obtain the restriction (A.40) on the length of the initial segment $\hat{\gamma}_0$ gives the condition (A.29) appearing in the theorem.

Note that (A.40) implies that: $\angle'(E_0^s, \dot{\gamma}) \geq C_4^{-1}\theta e^{-2\epsilon n}/2$. So the length of the projection of $\hat{\gamma}_0$ to the E_0^u axis, which we call l_0, has length (with respect to the Lyapunov metric) of at least $\operatorname{len}'(\gamma_0)\sin(C_4^{-1}\theta e^{-2\epsilon n}/2)$. Thus

$$l_0 \geq \operatorname{len}'(\gamma_0)\sin(C_4^{-1}\theta e^{-2\epsilon n}/2) \geq C_5\theta e^{-2\epsilon n}\min\{e^{-R}\theta, e^{-.9\lambda n}\} \tag{A.41}$$

Also by Lemma A.6

$$\|\hat{\gamma}_0\|_1 \leq \cot(C_4^{-1}\theta e^{-2\epsilon n}) \leq 2C_4\theta^{-1}e^{2\epsilon n}.$$

We apply Proposition A.12(3), and get an $\epsilon_0 < 1/\sqrt{3}$, which is the cutoff for the one step C^2 smoothing estimate (A.13) to hold.

In keeping with the previous proposition, denote

$$\epsilon_{1,i} = \|\hat{f}_{i,1}|_{B_i}\|_1 \quad \text{and} \quad \epsilon_{2,i} = \|\hat{f}_{2,i}|_{B_i}\|_1.$$

Because $\hat{f}_i = D_0\hat{f}_i + (\hat{f}_1, \hat{f}_2)$, we see from the C^2 bound on $\hat{f}_i$ that on B_i,

$$\max\{\epsilon_{1,i}, \epsilon_{2,i}\} \leq r_i\|\hat{f}_i\|_2 \leq C_2e^{-.9\lambda(n-i)}e^{6(n-i)\epsilon} = C_2e^{-(.9\lambda-\epsilon)(n-i)}. \tag{A.42}$$

We now proceed to the main part of the proof.

Step 1. We begin by checking that if we inductively define: $\hat{f}_i\hat{\gamma}_i|_{B_i} = \hat{\gamma}_{i+1}$, and, as before, l_i is the length of the projection of $\hat{\gamma}_i$ to E_i^u measured with respect to $\|\cdot\|_i'$, then the sequence $\hat{\gamma}_i$ satisfies the following estimates:

$$l_i \geq \min\{r_i, e^{.99\lambda i}l_0\}, \tag{A.43}$$

$$\|\hat{\gamma}_i\|_1 \leq \max\{2\theta^{-1}e^{-i\lambda}, \epsilon_0\}. \tag{A.44}$$

(1) (l_i): By Proposition A.12(1)

$$l_{i+1} \geq \min\{(e^{.999\lambda} - \epsilon_{1,i} - \epsilon_{1,i}\|\hat{\gamma}_i\|_1)l_i, r_i\}.$$

Hence to verify (A.43), it suffices to show that

$$(e^{.999\lambda} - \epsilon_{1,i} - \epsilon_{1,i}\|\hat{\gamma}_i\|_1)l_i \geq e^{.99\lambda}l_i.$$

which follows by (A.42) and the inductive hypothesis (A.44) if C_2 is chosen sufficiently small.

(2) We now check the estimate on $\|\hat{\gamma}_{i+1}\|_1$ assuming it holds for i.

To begin, from Proposition A.12(2),

$$\|\hat{\gamma}_{i+1}\|_1 \leq (e^{-.999\lambda}\|\hat{\gamma}_i\|_1 + \epsilon_{2,i} + \epsilon_{2,i}\|\hat{\gamma}_i\|_1)(e^{.999\lambda} - \epsilon_{1,i} - \epsilon_{1,i}\|\hat{\gamma}_i\|_1)^{-1}. \tag{A.45}$$

There are two cases depending on whether $\|\hat{\gamma}_i\|_1 \geq \epsilon_0$ or not. If $\|\hat{\gamma}_i\|_1 \geq \epsilon_0$, then as long as C_2 is chosen sufficiently small, then the second parenthetical term in the above equation is at most $e^{-.9\lambda}$ by (A.42). Hence

$$\|\hat{\gamma}_{i+1}\|_1 \leq e^{-.9\lambda}(e^{-.999\lambda}\|\hat{\gamma}_i\|_1 + \epsilon_{2,i} + \epsilon_{2,i}\|\hat{\gamma}_i\|_1) \leq e^{-.9\lambda}\|\hat{\gamma}_i\|_1(e^{-.999\lambda} + \epsilon_{2,i}\epsilon_0^{-1} + \epsilon_{2,i}).$$

Because $\epsilon_0 > 0$ is independent of C_2, if C_2 is sufficiently small then (A.42) gives $\|\hat{\gamma}_{i+1}\|_1 \leq e^{-3\lambda/2}\|\hat{\gamma}_i\|_1$, which concludes the proof since $\|\hat{\gamma}_i\|_1 \leq 2\theta^{-1}e^{-i\lambda}$.

We now consider the case $\|\hat{\gamma}_i\|_1 \leq \epsilon_0$. In this case it suffices to show that $\|\hat{\gamma}_{i+1}\|_1 \leq \epsilon_0$. The argument in this case is similar and follows because, as in the previous case, we may ensure that $\epsilon_{1,i}, \epsilon_{2,i}$ are small relative to ϵ_0 through our initial choice of C_2.

Thus we have shown that both estimates hold inductively proving (A.43) and (A.44).

We now conclude item (1). Since the Lyapunov metric $\|\cdot\|_n'$ is uniformly comparable to the ambient metric $\|\cdot\|_n$ due to (A.33), it is enough to prove the lower bound on $\text{len}'(\gamma_n)$. Thus the length of γ_0 is at least $\min\{r_n, e^{.9\lambda n}l_0\}$. Note that

$$e^{.9\lambda n}l_0 \geq C_5 e^{-R}\theta^2 e^{-2\epsilon n}.$$

Hence if the minimum of $\min\{r_n, e^{.9\lambda n}l_0\}$ is realized by r_n, then $r_n = C_2$ because the first term in the definition of r_n (see (A.38)) is bigger than $e^{.9\lambda n}\ell_0$. This shows that $\text{len}'(\gamma_n) \geq \min(\ell_0 e^{.9\lambda n}, C_2)$, completing the proof of part (1).

We now check the claim about the length of the preimage of $\hat{f}^n\hat{\gamma}_0$ in part (2). This is immediate from our choice of $\text{len}'(\hat{\gamma}_0)$ in (A.40). Because the preimage of γ_n is contained in a segment of length at most $e^{-.9\lambda n}$ with respect to the Lyapunov metric, and because $\|\cdot\|_0 \leq \sqrt{2}\|\cdot\|_0'$, this implies that the length of the initial segment we consider with respect to the ambient metric is at most $\sqrt{2}e^{-.9\lambda n}$. Similar considerations give the claim about the length of the preimage of γ_n in $f^i(\gamma)$ at the end of item (2). Note that the final curve γ_n promised by the lemma is not unique: for instance, it need not be centered at $f^n(x)$. The final claim in item (2) follows because any such curve is long enough that it fills the entire segment of $f^i(\gamma)$ we are considering by our choice of r_i.

To finish the proof of item (2), we must see how large n must be in order too ensure that $r_n = \ell_{\max}$. For this to occur n must satisfy $e^{.99\lambda n}l_0 \geq \ell_{\max}$. That is, $n \geq \dfrac{\ln(\ell_{\max}) - \ln(l_0)}{.99\lambda}$. Now the definition of l_0 (see (A.41)) gives

$$n \geq \frac{\ln(\ell_{\max}) - \ln(C_5\theta e^{-2\epsilon n}\min\{e^{-R}\theta, e^{-.9\lambda n}\})}{.99\lambda}.$$

Now the needed conclusion in item (2) follows by considering the two cases depending on which term realizes the minimum and using that $\epsilon < \lambda/30$.

Step 2. We now obtain item (3), the C^2 estimate on $\hat{\gamma}_i$. Should it happen that there is an index i such that $\|\hat{\gamma}_i\|_1 \le \epsilon_0$, we call this index N_0. We proceed under the assumption that there is some such N_0. After concluding in this case, we explain how the same estimate holds otherwise. Observe that if $\|\hat{\gamma}_i\|_1 \le \epsilon_0$, then for all $j \ge i$, $\|\hat{\gamma}_j\|_1 \le \epsilon_0$ as well. Keeping in mind the strength of hyperbolicity from (A.34), for all indices $i \ge N_0$, we have from (A.13), that

$$\|\hat{\gamma}_{i+1}\|_2 \le e^{-1.99\lambda\cdot .999}\|\hat{f}_i\|_2 + e^{-2.99\lambda\cdot .999}\|\hat{\gamma}_i\|_2. \tag{A.46}$$

By applying the above equation iteratively, we can obtain an estimate on $\|\hat{\gamma}_n\|_2$ in terms of $\|\hat{\gamma}_{N_0}\|_2$. This gives the required estimate because the homogeneous part of (A.46) has multipliers smaller than 1.

By (A.37), $\|\hat{f}_i\|_2 \le D_1 e^{6C} e^{6(n-i)\epsilon}$. Let $M = n - N_0$. Applying iteratively (A.46), we get

$$\|\hat{\gamma}_n\|_2 \le \|\hat{\gamma}_{N_0}\|_2 e^{-2.99\lambda\cdot .999M} + \sum_{i=1}^{M} D_1 e^{6C} e^{6(n-i)\epsilon} e^{-1.99\lambda\cdot .999} e^{-2.99\lambda\cdot .999(M-i-1)}. \tag{A.47}$$

Note that the second term is bounded by a constant C_6 depending only on C, λ and ϵ.

To conclude, we also need a bound for $\|\hat{\gamma}_{N_0}\|_2$. By Lemma A.9, there exists Λ depending only on the C^2 norm of the maps f_i, which is uniformly bounded by D_1, such that

$$\|f^i\gamma\|_{C^2} \le e^{\Lambda i}\max\{\|\gamma\|_{C^2}, 1\}. \tag{A.48}$$

Hence $\|\gamma_{N_0}\| \le e^{\Lambda N_0}\max\{\|\gamma\|_{C^2}, 1\}$. We then need an estimate on $\hat{\gamma}_{N_0}$. Note that in the Lyapunov coordinates that $\hat{\gamma}_{N_0}$, which as a graph over E^u_0 has slope at most $\epsilon_0 < 1/\sqrt{3}$. Thus by Lemma A.6 ,

$$\|\hat{\gamma}_{N_0}\|_2 \le \sin(\operatorname{arccot}(\epsilon_0))^{-3} e^{N_0\Lambda}\max\{\|\gamma\|_{C^2}, 1\} \le 2e^{\Lambda N_0}\max\{\|\gamma\|_{C^2}, 1\},$$

because $\epsilon_0 < 1/\sqrt{3} = \tan(\pi/6) = \cot(\pi/3)$. Combining this with (A.47),

$$\|\hat{\gamma}_n\|_2 \le 2e^{\Lambda N_0} e^{-2.99\lambda\cdot .999M}\max\{\|\gamma\|_{C^2}, 1\} + C_6. \tag{A.49}$$

But we also have a straightforward estimate for the cutoff N_0. From equation (A.44), we know that $N_0 \le (\ln(2) - \ln(\theta))/\lambda$. Hence because $N_0 \approx -\ln(\theta)$, it is straightforward to see that there exist C_7, C_8 such that

$$\|\hat{\gamma}_n\|_2 < C_6 + C_7 e^{-2.9\lambda n} e^{C_8 \ln\theta}\max\{\|\gamma\|_{C^2}, 1\}. \tag{A.50}$$

In the case that there is no index i such that $\|\hat{\gamma}_i\|_1 \le \epsilon_0$, we may conclude similarly as equation (A.44) implies that $n \le (\ln(2) - \ln(\theta))/\lambda$. Thus we have finished with Step 2 and conclude item (3).

Before going to Step 3, we record an additional more precise estimate on the rate that $\|\hat{\gamma}_i\|_2$ improves. Similar to above, we find:

$$\begin{aligned}
\|\hat{\gamma}_{N_0+i}\|_2 &\le \|\hat{\gamma}_{N_0}\|_2 e^{-2.99\lambda\cdot .999i} + \sum_{j=1}^{i} D_1 e^{6C} e^{6(n-N_0-j)\epsilon} e^{-1.99\lambda\cdot .999} e^{-2.99\lambda\cdot .999(i-j-1)}, \\
&\le \|\hat{\gamma}_{N_0}\|_2 e^{-2.99\lambda\cdot .999i} + e^{6(n-N_0)\epsilon} e^{-1.99\lambda\cdot .999} e^{-6i\epsilon} D_1 e^{6C}\sum_{k=1}^{i} e^{-(2.99\lambda\cdot .999-6\epsilon)(k-1)}, \\
&\le 2e^{\Lambda N_0} e^{-2.99\lambda\cdot .999i}\max\{\|\gamma\|_{C^2}, 1\} + C_9 e^{6\epsilon(n-N_0-i)},
\end{aligned} \tag{A.51}$$

for some $C_9 > 0$.

Step 3. We now show item (4), i.e. we obtain estimates for smoothing a density along γ. We let $\hat{\rho}_i$ be the function giving the density ρ on $\hat{\gamma}_i$ in the Lyapunov coordinates.

We now apply the smoothing estimate. As in Step 2, supposing it exists, let N_0 be the first index such that $\|\hat{\gamma}_i\|_1 \leq \epsilon_0$. If such an index N_0 does not exist, then we may conclude similarly to in Step 2. Then for any $i \geq N_0$, by (A.14),

$$\|\ln \hat{\rho}_{i+1}\|_{C^\alpha} \leq \lambda^{-.9\alpha}(\|\ln \hat{\rho}_i\|_{C^\alpha} + \|\hat{f}_i\|_2 + \|\hat{\gamma}_i\|_2). \tag{A.52}$$

As before, let $M = n - N_0$. By a bookkeeping similar to Step 2, we find that

$$\|\ln \hat{\rho}_n\|_1 \leq e^{-.9 \cdot .999\lambda\alpha M}\|\ln \hat{\rho}_{N_0}\|_{C^\alpha} + \sum_{i=1}^{M} e^{-.9\cdot.999\lambda\alpha(M-i)}(\|\hat{f}_{N_0+i}\|_{C^2} + \|\hat{\gamma}_{N_0+i}\|_{C^2}).$$

By (A.51) and (A.37), we see that there exists C_{11} such that

$$\|\ln \hat{\rho}_n\|_{C^\alpha} \leq e^{-.9\cdot.999\lambda\alpha M}\|\ln \hat{\rho}_{N_0}\|_{C^\alpha} + 2e^{\Lambda N_0}\|\gamma\|_{C^2}e^{-.9\lambda\cdot.999M} + C_{11}. \tag{A.53}$$

We now estimate $\|\ln \hat{\rho}_{N_0}\|$. We first obtain an estimate without the use of the Lyapunov charts. By Lemma A.9 because of the uniform C^2 bound D_1, there exist $C_{12}, \Lambda > 0$ such that

$$\|\ln(f^{N_0})_*\rho\|_{C^\alpha} \leq C_{12}(e^{\Lambda N_0} + e^{\Lambda N_0}\|\ln \rho\|_{C^\alpha}).$$

Next, we push forward γ_{N_0} and ρ_{N_0} by L_{N_0} to obtain a density in the Lyapunov coordinates. Because $\max\{\|L_{N_0}\|, \|L_{N_0}\|^{-1}\} \leq C_1e^{2C}e^{2\epsilon n}$, Lemma A.7 gives that there exists C_{13} such that

$$\|\ln(L_{N_0}) * (f_1^{N_0})_*\rho\|_{C^\alpha} \leq C_{13}e^{(2+2\alpha)\epsilon n}\left(e^{\Lambda N_0} + e^{\Lambda N_0}\|\ln \rho\|_{C^\alpha} + e^{2\epsilon n}(1 + \|\gamma_{N_0}\|_{C^2})\right).$$

For the application we are then interested in the regularity of $(Lf_1^{N_0})_*\rho$ as a function parametrized by E_0^u. As at time N_0, γ_{N_0} is uniformly transverse to E_0^s, this projection has uniformly bounded norm. From before, we have the C^2 bound on γ_{N_0} following (A.48), which gives that there exists C_{14} such that:

$$\|\ln \hat{\rho}_{N_0}\|_{C^\alpha} \leq C_{14}e^{7\epsilon n}e^{\Lambda N_0}(1 + \|\ln \rho\|_{C^\alpha} + \|\gamma\|_{C^2}). \tag{A.54}$$

Combining this with (A.53), we find

$$\|\ln \hat{\rho}_n\| \leq e^{-.9\cdot.999\lambda\alpha M}(C_{14}e^{7\epsilon n}e^{\Lambda N_0}(1 + \|\ln \rho\|_{C^\alpha} + \|\gamma\|_{C^2})) + 2e^{\Lambda N_0}\|\gamma\|_{C^2}e^{-.9\lambda\cdot.999M} + C_{11}.$$

Then as before, because N_0 is order $\ln(\theta)$ and $M = n - N_0$,

$$\|\ln \hat{\rho}_n\| \leq C_{15}e^{-.9\lambda\alpha n}e^{C_{16}\ln(\theta)}(1 + \|\ln \rho\|_{C^\alpha} + \|\gamma\|_{C^2}). \tag{A.55}$$

As $\|\gamma_n\|_{C^2} < D_9$ for some fixed D_9 by assumption, then (A.55) gives the corresponding estimate on ρ with respect to the arclength parameters on γ_n, and we conclude item (4). □

A.6. **Loss of regularity.** In this subsection, we prove some additional estimates that will be used later in the proof of mixing but not the proof of the coupling lemma. These estimates say that for all but an exponentially small amount of the curve γ, typically the images of points in $f^n_\omega(\gamma)$ are in a neighborhood that is at least $n\epsilon$-good. First we introduce in Definition A.14, a notion of a forward tempered point relative to a curve. Then, in Proposition A.15, we show that the image of a curve at a forward tempered time will be $18\epsilon n$ good.

We begin by stating the main definition of this section. Note that it is similar to definitions we also considered for backwards good points (Definition 8.1).

Definition A.14. For a standard pair $\hat{\gamma} = (\gamma, \rho)$ and a word $\omega \in \Sigma$, we say that n is a $(C, \lambda, \epsilon, \theta)$-forward tempered time for $x \in \gamma$ if the sequence of maps $(D_xf^i_\omega)_{1\leq i\leq n}$ is (C, λ, ϵ)-subtempered and the most contracted direction of $D_xf^n_\omega$ exists and is at least θ-transverse to γ. Similarly, we speak of a trajectory being forward tempered relative to a vector $v \in T_xM$.

The following lemma gives a quantitative estimate on the length of an image of a curve experiencing a forward tempered time.

Proposition A.15. *Suppose that M is a closed surface and that $(f_1, \ldots, f_m)$ is a tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. Then for any $\lambda > 0$ and $C_1 > 0$ there exist $D_0, D_1 > 0$ and $N \in \mathbb{N}$, such that for all $\theta > 0$ and $\lambda/30 > \epsilon > 0$, if $\hat{\gamma} = (\gamma, \rho)$ is a C_1-good standard pair, $\omega \in \Sigma$ and $x \in \gamma$ has a $(C, \lambda, \epsilon, \theta)$ forward tempered time at time*

$$n \geq N + D_0 \ln \theta \tag{A.56}$$

then

(1) The pushforward $f^n_\omega(\gamma)$ contains a neighborhood of $f^n_\omega(x)$, $B(x)$, such that denoting by $\hat{B}(x)$ the restriction of the standard pair $f^n_\omega(\hat{\gamma})$ to $B(x)$, then $\hat{B}(x)$ is an $(18\epsilon n + 18\max\{C, 0\} + D_1)$-good standard pair.

(2) The preimage of $\hat{B}(x)$, $(f^n_\omega)^{-1}(\hat{B}(x))$, has length at most $e^{-(\lambda/2)n}$.

Proof. As before, we will use the deterministic smoothing lemmas. We begin by first picking a choice of Lyapunov metrics to use. Applying Lemma A.1 with $\lambda' = .999\lambda$ we get, since the trajectory is forward tempered, that

$$\frac{1}{\sqrt{2}}\|\xi\|_i \leq \|\xi\|'_i \leq 4e^{2C+2\epsilon i}\left(1 - e^{2(\lambda'-\lambda)}\right)^{-1/2}\|\xi\|_i. \tag{A.57}$$

As in the proof of Proposition A.13, using the Lyapunov metric, we obtain new dynamics $\hat{f}_i$ in the Lyapunov coordinates, which are given by composing with a sequence of maps L_i. Crucially, these dynamics satisfy that $D_0\hat{f}_i|_{E^s_i} \leq e^{-.999\lambda}$ and $D_0\hat{f}_i|_{E^u_i} \geq e^{.999\lambda}$. Moreover, we can write:

$$\hat{f}_i(x, y) = (\sigma_{1,i}x + \hat{f}_{i,1}(x, y), \sigma_{2,i}y + \hat{f}_{i,2}(x, y)).$$

Further there exists C_2 such that

$$\max\{\|L_i\|, \|L_i\|^{-1}\} \leq C_2 e^{2C+2\epsilon i}. \tag{A.58}$$

Proceeding as in (A.37), there exists D such that:

$$\|\hat{f}_i\|_{C^2} \leq De^{6C+6i\epsilon}. \tag{A.59}$$

From here, we set up the constants in a manner similar to before. Things are slightly simpler because by assumption the standard pair is C_1-good and hence uniformly long and good. We will take some small $C_3 > 0$ that we will choose later. Set

$$r_i = C_3 e^{-10\epsilon n - 6C}\min\{\theta^{-1}, e^{-.9\lambda(n-i)}\}.$$

As before, we let B_i be the square of side length r_i centered at 0 with respect to the $\|\cdot\|'_i$ metric. As in the previous argument, we let $\hat{\gamma}_i$ denote the portion of $f^{i-1}(\hat{\gamma})$ lying in B_i and we let $\hat{\rho}_n$ denote the density along $\hat{\gamma}_n$. Let $\epsilon_0 > 0$ be the cutoff so that (A.13) holds in Proposition A.12

As above, we denote $\epsilon_{1,i} = \|\hat{f}_{i,1}|_{B_i}\|_1$ and $\epsilon_{2,i} = \|\hat{f}_{2,i}|_{B_i}\|_1$. Because $\hat{f}_i = D_0\hat{f}_i + (\hat{f}_1, \hat{f}_2)$, we see that from the C^2 bound on $\hat{f}_i$ that on B_i,

$$\max\{\epsilon_{1,i}, \epsilon_{2,i}\} \leq r_i\|\hat{f}_i\|_2 \leq C_3 e^{-10\epsilon n - 6C}e^{-.9\lambda(n-i)}e^{6C}e^{6(n-i)\epsilon} \leq C_3 e^{-.9\lambda(n-i)}. \tag{A.60}$$

In particular, note that by choosing C_3 sufficiently small in a manner that only depends on λ, we may ensure that for all i that $\max\{\epsilon_{1,i}, \epsilon_{2,i}\} < \epsilon_0$.

We then carry out an inductive argument to determine the regularity of $\hat{\gamma}_n$. In order, we obtain estimates on the length, the C^2 norm, and then $\|\ln f^n(\rho)\|_{C^\alpha}$.

Step 1. (Length of the curve) As in the proof of Proposition A.13, we see that from the choice of constants $\hat{\gamma}_n$ is uniformly transverse to E^s_n and the projection of its graph to the E^u_n axis fills $E^u_n \cap B_n$. Thus there exists C_4, depending only on C_3 such that $\hat{\gamma}_n$ has length at least $C_4e^{-\epsilon n}$ in the Lyapunov charts. By equation (A.58), this implies that, in the ambient metric, $f^n_\omega(x)$ lies in a neighborhood of length at least

$$C_2^{-1}C_4e^{-2C-3\epsilon n}. \tag{A.61}$$

Step 2. (C^2 norm of the curve) We now turn to an estimate on the C^2 norm of $\hat{\gamma}_n$. This is perhaps the most complicated part of the argument along with the estimate on smoothing the density. We apply the estimate (A.13) from Proposition A.12. Let N_0 be the first iterate such that $\|\hat{\gamma}_{N_0}\|_1 < \epsilon_0$. From our choice of the size of the neighborhood and the comment on the size of C_3 immediately after (A.60), we have that for all $i \geq N_0$, the estimate (A.13) holds, i.e. the C^2 smoothing estimate is valid. Thus we find that:

$$\|\hat{\gamma}_{i+1}\|_2 \leq \lambda^{-1.99}\|\hat{f}_i\|_2 + \lambda^{-2.99}\|\hat{\gamma}_i\|_2. \tag{A.62}$$

From (A.59), it follows inductively that:

$$\|\hat{\gamma}_n\|_2 \leq e^{-2.99\lambda(n-N_0)}\|\hat{\gamma}_{N_0}\|_2 + De^{6C}e^{6\epsilon n}\sum_{j=0}^{n-N_0-1} e^{-1.99\lambda j}e^{6\epsilon j}. \tag{A.63}$$

We then need to estimate N_0. As in the proof of Proposition A.13, we get $N_0{=}O_\lambda(-\ln(\theta))$. Thus there exists $C_4, C_5 > 0$, such that

$$\|\hat{\gamma}_n\|_2 \leq e^{-2.99n\lambda}e^{-C_4\ln\theta}\|\hat{\gamma}_{N_0}\|_2 + C_5e^{6C}e^{6\epsilon n}.$$

As in the proof of Proposition A.13, after (A.48), we see that there exists $\Lambda > 0$ such that $\|f^i\gamma\|_{C^2}$, with respect to the ambient metric is at most $e^{\Lambda i}$. Using (A.56) and the fact that the angle between γ and $E^s_{N_0}$ is uniformly large, we see that there exists C_6 such that with respect to the Lyapunov metric,

$$\|\hat{\gamma}_{N_0}\|_2 \leq e^{-C_6\ln\theta}.$$

Thus for some C_7,

$$\|\hat{\gamma}_n\|_2 \leq e^{-2.99n\lambda}e^{-C_7\ln\theta} + C_5e^{6C}e^{6\epsilon n}. \tag{A.64}$$

We now record an intermediate estimate that will be useful later. By possibly increasing the constants, for each $N_0 \leq i \leq n$, we find:

$$\|\hat{\gamma}_i\|_2 \leq e^{-2.99i\lambda}e^{-C_7\ln\theta} + C_5e^{6C}e^{6\epsilon i}. \tag{A.65}$$

Equation (A.64) is an estimate in the Lyapunov chart, but we need the estimate with respect to the original metric. The C^2 norm of $\hat{\gamma}_n$ as a curve is uniformly comparable to $\|\hat{\gamma}_n\|_2$ because $\hat{\gamma}_n$ is uniformly transverse to E^s_n. By Lemma A.3 there exists C_7, such that, letting γ_n be the segment of γ lying in B_n, we get the following bound in the ambient metric

$$\|\gamma_n\|_{C^2} \leq (e^{-2.99n\lambda}e^{-C_7\ln\theta} + C_5e^{6C}e^{6\epsilon n})C_2^3e^{6C+6\epsilon n}.$$

This is the bound required by the proposition. Indeed for D_0 sufficiently large we have:

$$\|\gamma_n\|_{C^2} \leq C_8e^{12\max\{C,0\}+12\epsilon n}. \tag{A.66}$$

Step 3. (Regularity of the density) Finally, we turn to estimating the Hölder norm of the pushed density. At the same iterate N_0 from Step 2, we have that $\epsilon_{N_0,1}, \epsilon_{N_0,2}, \|\hat{\gamma}_{N_0}\|_1 \leq \epsilon_0$ and that these estimates hold for all future iterates. Consequently, estimate (A.14) applies, hence for $N_0 \leq i \leq n-1$,

$$\|\ln\widetilde{\rho}_{i+1}\|_{C^\alpha} \leq e^{-.9\alpha\lambda}(\|\ln\widetilde{\rho}_i\|_{C^\alpha} + \|\hat{f}_i\|_2 + \|\hat{\gamma}_i\|_2).$$

This leads inductively to the estimate that

$$\|\ln\widetilde{\rho}_n\|_{C^\alpha}\le e^{-.9\alpha\lambda(n-N_0)}\|\ln\widetilde{\rho}_{N_0}\|_{C^\alpha}+\sum_{i=N_0}^{n-1}e^{-.9\lambda\alpha(n-i)}(\|\hat{f}_i\|_2+\|\hat{\gamma}_i\|_2). \tag{A.67}$$

We then need some further estimates in order to simplify this.

We start with an estimate on $\|\ln\widetilde{\rho}_{N_0}\|_{C^\alpha}$. A similar argument to that giving (A.54) yields that $\|\ln\widetilde{\rho}_{N_0}\|_{C^\alpha}\le e^{\Lambda N_0}$, where $\Lambda>0$ only depends on the C^2 norm of the diffeomorphisms and the initial regularity of γ. Hence as long as D_0 is large enough, it follows that the first term is uniformly bounded.

For the other terms, we already have estimates for $\|\hat{f}_i\|_2$ and $\|\hat{\gamma}_i\|_2$, (A.59) and (A.65). These yield a bound on the sum in (A.67):

$$\sum_{i=N_0}^{n-1}e^{-.9\lambda\alpha(n-i)}(\|\hat{f}_i\|_2+\|\hat{\gamma}_i\|_2)$$

$$\le\sum_{j=0}^{n-N_0-1}e^{-.9\lambda\alpha j}\left(De^{6C}e^{6\epsilon(n-j)}+e^{-2.99\lambda(n-j)}e^{-C_7\ln(\theta)}+C_5e^{6C}e^{6\epsilon(n-j)}\right).$$

The sum of the first and third terms inside the parentheses is straightforward to evaluate. There is a constant C_9 such that each is bounded by $C_9e^{6C}e^{6\epsilon n}$. The terms involving the $\ln(\theta)$ are only slightly more complicated as either j or $n-j$ is large, hence the terms involving λ dominate the $e^{-C_7\ln\theta}$ term as long as D_0 is large enough. Thus by the above estimates, it follows that as long as D_0 is sufficiently large that there exists C_{10} such that

$$\|\ln\widetilde{\rho}_n\|_{C^\alpha}\le C_{10}e^{6\max\{C,0\}}e^{6\epsilon n}.$$

This is the form of the estimate in the Lyapunov charts. We then need to pass back to the original metric. Applying Lemma A.7 we see that letting C' denote the constant from that lemma and using (A.58) and (A.66)) we get,

$$\begin{aligned}\|\ln\rho_n\|_{C^\alpha}&\le e^{(1+\alpha)(2C+2\epsilon n)}\left(C_{10}e^{6\max\{C,0\}+6\epsilon n}+C'e^{2C+\epsilon n}(1+C_8e^{12\max\{C,0\}+12\epsilon n})\right)\\&\le C_{11}e^{18\max\{C,0\}}e^{18\epsilon n}.\end{aligned}$$

This is the needed conclusion, so we are done. □

Appendix B. Finite time Pesin theory and fake stable manifolds

B.1. **Fake stable manifolds.** In the proof of the coupling lemma, we will use the holonomies of some "fake" stable manifolds $W^s_n(\omega,x)$. These manifolds behave for finite a time like a true stable manifold insofar as they contract. We then prove some lemmas about fake stable curves. Some of the results below are variants on standard facts in Pesin theory, however, some of the proofs are a little different due to us only using a finite portion of an orbit. For other facts that look standard we needed to supply our own proofs because we could not find a similar enough statement in the literature.

For a given word ω and $n\in\mathbb{N}$ the fake stable manifolds are curves that have analogous properties to the stable manifolds up until time n. So, unlike true stable manifolds, they are not canonically defined.

Before we begin we recall some notation. Throughout this section we will write $\Lambda^\omega_n(C,\lambda,\epsilon)$ for the set of (C,λ,ϵ)-tempered points $x\in M$ at time n for the word $\omega\in\Sigma$. This is essentially the finite time version of a Pesin block. For many of the results there is a lower bound on n, which is required to ensure that the orbit is actual experiencing hyperbolicity.

Below we will make a number of arguments concerning these fake stable manifolds. The main properties we need concern the holonomies between two transversals to the W_n^s lamination. We need to know that the W_n^s holonomies have a uniformly Hölder continuous Jacobian independent of n. In addition we would like to know that as $n \to \infty$ that the holonomies are converging exponentially quickly to the true stable holonomy.

Before proceeding to the proof, we remark that there are other approaches to fake stable manifolds that are adapted to different sorts of dynamical problems and may differ from each other substantially. For example, Burns and Wilkinson [BW10], which originated the term fake manifold, use fake center and stable manifolds where a potentially different fake foliation is defined at every point in the manifold. A different approach in Dolgopyat, Kanigowski, Rodriguez-Hertz [DKRH24] uses a fake foliation that is globally defined but does not cover the entire Pesin regular set. Note that, in contrast with our setting, [BW10] and [DKRH24] allow systems with some zero exponents, and so the invariant manifolds need not be unique in their settings. One benefit of the construction described below is that it applies to *every* point in a Pesin block and further gives a single fake stable lamination defined on the manifold rather than a collection of different overlapping laminations. While this makes the fake stable lamination simple to think about, it requires more work to show that it exists.

B.2. **Preliminaries.** Here we present some background that will be used in the next subsection to study the regularity of E_n^s.

We start with a useful fact for showing that the limit of a sequence of functions is Hölder continuous. This fact is completely standard. Note that the statement is false if the diameter of M_2 is unbounded. Also, recall that in our setup, the Hölder constant only applies to estimates on the distance between $g(x)$ and $g(y)$ for points with $d(x, y) \leq 1$.

Lemma B.1. *Suppose that M_2 is a metric space with bounded diameter. Fix $\eta, \lambda, \delta, \beta > 0$. Then there exists $0 < \alpha < \beta$ and $D(\eta, \lambda, \delta, \beta, \alpha)$ such that for any metric space M_1 the following holds. Let $g_n \colon M_1 \to M_2$, $1 \leq n \leq N$ be a finite or infinite sequence of β-Hölder continuous functions such that:*
(1) For $1 \leq n < N$, $d_{C^0}(g_n, g_{n+1}) \leq C_1 e^{-\delta n}$.
(2) The function g_n is $C_3 e^{\eta n}$ β-Hölder continuous at scale $e^{-C_2} e^{-\lambda n}$, i.e., if $d(x, y) \leq e^{-C_2} e^{-\lambda n}$ then $d(g_n(x), g_n(y)) \leq C_3 e^{\eta n} d(x, y)^\beta$.
Then the functions $g_1, \ldots, g_N$ in the sequence, as well as the possible limiting value of the sequence are all uniformly α-Hölder with constant at most

$$\max\left\{De^{C_2\alpha}, 2C_1(1 - e^{-\delta})^{-1} e^{\beta C_2} D^\beta + C_3 e^{-C_2(\beta-\alpha)}\right\}.$$

Proof. We will assume throughout the proof that g_N is fixed and obtain an estimate for g_N that is independent of N. As the resulting estimate is independent of N, the conclusion holds for infinite sequences as well.

To begin we pick some constants. First, for fixed $\eta > 0$ and any $0 < \alpha_1 < \beta$ there exists $\gamma \geq \lambda$ such that

$$\eta - \gamma\beta \leq -\alpha_1\gamma \quad \text{and} \quad \eta < \gamma\alpha_1. \tag{B.1}$$

Note that γ only depends on $\eta, \alpha_1, \lambda, \beta$, but not on C_1, C_2, C_3.

Next given δ, let $0 < \alpha_2 < \beta$ be sufficiently small that we have

$$\delta \geq \alpha_2\gamma. \tag{B.2}$$

Due the first assumption, we have a uniform estimate independent of N:

$$|g_N - g_n| \le \sum_{i=n}^{N-1} C_1 e^{-i\delta} \le \frac{C_1 e^{-n\delta}}{1-e^{-\delta}}.$$

Having picked those constants, now consider a pair of points $x, y \in M_1$. We consider three cases depending on how far apart x and y are. We proceed from closest to furthest away.

(1) First suppose that $d(x,y) < \min\{e^{-C_2}e^{-\gamma N}, 1\}$. Then

$$\begin{aligned} d(g_N(x), g_N(y)) &\le C_3 e^{\eta N} d(x,y)^{\beta} \le C_3 e^{\eta N} d(x,y)^{\alpha_1} d(x,y)^{\beta-\alpha_1} \\ &\le C_3 e^{\eta N} e^{-C_2\alpha_1} e^{-\gamma\alpha_1 N} d(x,y)^{\beta-\alpha_1} \le C_3 e^{-C_2\alpha_1} d(x,y)^{\beta-\alpha_1}, \end{aligned}$$

where we have used (B.1).

(2) Next, we consider the case where $e^{-C_2}e^{-\gamma}e^{-\gamma n} \le d(x,y) \le \min\{1, e^{-C_2}e^{-\gamma n}\}$ for some $1 \le n < N$. By the choice of constants α_1, α_2 and γ in the first part of the proof we find:

$$\begin{aligned} d(g_N(x), g_N(y)) &\le d(g_N(x), g_n(x)) + d(g_n(x), g_n(y)) + d(g_n(y), g_N(y)) \\ &\le C_1 e^{-n\delta}(1-e^{-\delta})^{-1} + C_1 e^{\eta n} d(x,y)^{\beta} + C_1 e^{-n\delta}(1-e^{-\delta})^{-1} \\ &\le 2C_1 e^{-n\gamma\alpha_2}(1-e^{-\delta})^{-1} + C_3 e^{\eta n} e^{-n\gamma\alpha_1} e^{-C_2\alpha_1} d(x,y)^{\beta-\alpha_1}. \end{aligned}$$

Then due to the lower bound on $d(x,y)$ and $\eta < \gamma\alpha_1$ from (B.1):

$$\begin{aligned} d(g_N(x), g_N(y)) &\le 2C_1(1-e^{-\delta})^{-1} e^{\alpha_2 C_2} e^{\alpha_2\gamma} d(x,y)^{\alpha_2} + C_3 e^{-C_2\alpha_1} d(x,y)^{\beta-\alpha_1} \\ &\le \left(2C_1(1-e^{-\delta})^{-1} e^{\alpha_2 C_2} e^{\alpha_2\gamma} + C_3 e^{-C_2\alpha_1}\right) d(x,y)^{\min\{\alpha_2, \beta-\alpha_1\}}. \end{aligned}$$

(3) Finally we consider the case where $e^{-C_2}e^{-\gamma} < d(x,y)$. Then we use a trivial bound

$$d(g_N(x), g_N(y)) \le \operatorname{diam}(M_2) \le \left(\frac{\operatorname{diam}(M_2)}{(e^{-C_2}e^{-\gamma})^{\beta-\alpha_1}}\right) d(x,y)^{\beta-\alpha_1}.$$

Now using all three cases above, we may conclude. Note that the $(\beta-\alpha_1)$-Hölder constant obtained in the second item above is at least as big as the constant obtained in the first item in the list. Thus the function $g_N(x)$ is uniformly $(\beta-\alpha_1)$-Hölder with constant at most

$$\max\left\{(\operatorname{diam} M_2) e^{(C_2+\gamma)(\beta-\alpha_1)}, 2C_1(1-e^{-\delta})^{-1} e^{\alpha_2 C_2} e^{\alpha_2\gamma} + C_3 e^{-C_2\alpha_1}\right\}.$$

As the choice of constants $\alpha_1, \alpha_2, \gamma$ depend only on δ, η we obtain the needed conclusion. □

We will apply Lemma B.1 to obtain regularity of E^s_n after we obtain small scale Hölder continuity of E^s_n.

Next we present a perturbation result on the singular subspaces of linear transformations called Wedin's theorem. This theorem gives a bound on the change in the angle between the singular directions. We state a specialized version of this theorem adapted from the presentation in [Ste91, Thm. 4]. First we describe the theorem in some generality, but below we give a precise statement for $\mathrm{SL}(2,\mathbb{R})$ independent of the discussion and definitions mentioned below. If A and $\widetilde{A}$ are two $n\times n$ matrices then we may list their singular values as $\sigma_1 \ge \cdots \ge \sigma_n$ and $\widetilde{\sigma}_1 \ge \cdots \widetilde{\sigma}_n$. Write $\|E\|_F$ for the Frobenius norm of the matrix E, i.e. the L^2 norm of its entries viewed as a vector. Fix some index k such that $\sigma_k \ge \widetilde{\sigma}_{k+1}$. If $|\sigma_k - \widetilde{\sigma}_{k+1}| \ge \delta$, and $\widetilde{\sigma}_k \ge \delta$, then Wedin's theorem implies that:

$$\|\sin\Phi\|_F \le \frac{\sqrt{2}\|E\|_F}{\delta},$$

where $\|\sin\Phi\|_F$ denotes the Frobenius norm of the matrix that defines the canonical angles between the right singular subspace associated to $\sigma_1,\dots,\sigma_k$ and $\widetilde{\sigma}_1,\dots,\widetilde{\sigma}_k$. (The matrix $\sin\Phi$ is defined by taking the inner products between an orthonormal basis of the right singular subspaces of A and $\widetilde{A}$.) Note that the statement in [Ste91, Thm. 4] is in terms of certain residuals, but by the comment before the theorem, these are bounded by $\|E\|_F$. Below we will use that the Frobenius norm of a 2 by 2 matrix satisfies the bound $\|E\|_F \le \sqrt{2}\|E\|$, where $\|E\|$ is the usual operator norm of the matrix [HJ13, 5.6.P23].

Although the statement from the above paragraph is somewhat technical, when both the matrix A and its perturbation $A+E$ are in $\mathrm{SL}(2,\mathbb{R})$, as is the case for us, the statement simplifies considerably. This is because for such a matrix $\sigma_1 = \sigma_2^{-1}$ and the top singular value of a matrix in $\mathrm{SL}(2,\mathbb{R})$ can change by at most $\|E\|$ when we perturb by E. If $\|A\| \ge 2$ and E is a perturbation with $\|E\| \le \|A\|/2$, then

$$\|A+E\| \ge \frac{1}{2}\|A\|,$$
$$\|A\| - (\|A+E\|)^{-1} \ge \|A\| - \frac{2}{\|A\|} \ge \frac{1}{2}\|A\|,$$

as long as $\|A\| \ge 2$. So, we may apply Wedin's theorem with $\delta = \|A\|/2$. In this case, the matrix of canonical angles described above consists of a single number: the angle between the original most expanded singular direction and the new one. Thus we obtain the following proposition.

Proposition B.2. *Suppose that A is a matrix in $\mathrm{SL}(2,\mathbb{R})$ with $\|A\| \ge 2$. Consider a perturbation $A+E \in \mathrm{SL}(2,\mathbb{R})$ with $\|E\| \le \|A\|/2$. Denote by v_A and v_{A+E} the most expanded singular vectors of A and $A+E$. Then*

$$|\sin\angle(v_A, v_{A+E})| \le \frac{2\sqrt{2}\|E\|}{\|A\|}.$$

B.3. **Regularity of the most contracting direction.** We now estimate the regularity of $E_n^s(x)$, the most contracted direction of $D_x f_\omega^n$, on the set of (C,λ,ϵ)-tempered points at time n in terms of C. The approach to studying Hölder regularity here may be contrasted with the approach in Shub [Shu87, Thm. 5.18(c)]. That approach establishes Hölder regularity for an invariant section of a bundle automorphism under an appropriate bunching condition by comparing the contraction in the fiber with the strength of hyperbolicity in the base. In some sense the approach is similar: it uses the dynamics to study the Hölder regularity at different scales. One can compare equation (***) there with our Lemma B.1.

Proposition B.3. *Suppose that $(f_1,\dots,f_m)$ is a tuple of diffeomorphisms in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$ of a closed surface M. Fix $\lambda > 0$ then there exists $\epsilon_0, \beta > 0$ such that for any $0 \le \epsilon \le \epsilon_0$ there exists D_1 such that if for $C \ge 0$, $\Lambda_\omega^n(C)$ denotes the (C,λ,ϵ) tempered points at time n for $\omega \in \Sigma$, and $n \ge N_0(C) = \lceil (C+\ln(2))/\lambda \rceil$, then restricted to $\Lambda_\omega^n(C)$, E_n^s is β-Hölder with constant $e^{D_1 C}$.*

Proof. We may always study the dynamics in an atlas of uniformly smooth volume preserving charts on M. So, in what follows we will implicitly be working with such charts.

The first claim is an immediate analog of [BP07, Lem. 5.3.4]. There exists $\Lambda > 0$ such that for $n \in \mathbb{N}$, if $x, y \in M$ with $d(x,y) \le e^{-\Lambda n}$, then (as viewed in charts),

$$\|D_x f_\omega^n - D_y f_\omega^n\| \le e^{\Lambda n} d(x,y). \tag{B.3}$$

Our plan is to apply Lemma B.1, so we need to estimate the regularity of E^s_n. The first thing we need is a lower bound on n for the subspace E^s_n to necessarily exist. From the definition of (C,λ,ϵ) tempered, we see that as long as

$$n \geq \left\lceil \frac{C+\ln 2}{\lambda} \right\rceil = N_0(C), \tag{B.4}$$

then $\|D_x f^n_\omega\| \geq 2$ and hence there is a well defined most contracted subspace.

Next we estimate the Hölder regularity of E^s_n on Λ^N_ω for $N \geq n \geq N_0$. If $x \in \Lambda^n_\omega$ and $d(x,y) \leq e^{-\Lambda n}/2$, then it follows from (B.3) that

$$\|D_x f^n_\omega - D_y f^n_\omega\| \leq e^{\Lambda n} d(x,y) \leq 1/2.$$

Thus, from Proposition B.2, as $\|D_x f^n\| \geq 2$, it follows that for $d(x,y) \leq e^{-\Lambda n}/2$ that

$$d(E^s_n(x), E^s_n(y)) < \sqrt{2} e^{\Lambda n} d(x,y), \tag{B.5}$$

which is the small scale Hölder estimate we were seeking.

Next, we study how fast E^s_n fluctuates as we increase n. By assumption the sequence of points is (C,λ,ϵ)-tempered. Hence by Proposition 4.6, there exists D_8 depending only on λ,ϵ such that for n greater than or equal to our same N_0 it follows that on $\Lambda^\omega_n(C,\lambda,\epsilon)$

$$\angle(E^s_n(x), E^s_{n+1}(x)) \leq e^{4C+D_8} e^{-2(\lambda-\epsilon)n}. \tag{B.6}$$

We can now apply Lemma B.1 to the sequence of distributions E^s_n, for $N_0 \leq n \leq N$ by combining estimates (B.5) and (B.6). Thus there exists $0<\beta<1$ and C_3 such that the E^s_n are β-Hölder with constant

$$\max\{C_3 e^{\Lambda N_0}, 2e^{4C+D_8}(1-e^{-2(\lambda-\epsilon)})^{-1} e^{\Lambda N_0} C_3 + C_3 e^{-\Lambda N_0}\}$$

But by our choice of $N_0 \approx C/\lambda$ and absorbing some constants into each other, we find that there is some C_4 such that the β-Hölder constant of E^s_n is at most $C_4 e^{((\Lambda/\lambda)+4)C}$, which gives the needed conclusion. □

The above lemma will give us a Hölder estimate on the regularity of $Df^n(E^s_n)$ as well, which will allow us to define the fake stable manifolds. Before proceeding, we use the above results to record another useful fact about the continuity of the distribution of the stable directions.

Proposition B.4. *Suppose that $(f_1,\ldots,f_m)$ is an expanding on average tuple of diffeomorphisms on a surface M in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$. Let ν^s_x be the distribution of stable subspaces through x, which is a probability measure on $\mathbb{P}T_xM$, the projectivization of T_xM. Then if we identify nearby fibres by parallel transport, the map $x \mapsto \nu^s_x$ is continuous in the weak* topology.*

Proof. Let $\nu^s_x(C,\lambda,\epsilon,n)$ denote the distribution of $E^s_n(\omega)$ for words ω that are (c,λ,ϵ)-tempered for some c in $[C,C+1)$. Then by Proposition B.3, the distribution E^s_n for such words ω is uniformly Hölder continuous in n for fixed C. So, if $\nu^s_x(C,\lambda,\epsilon)$ denotes the distribution of $E^s_\omega(x)$ for (C,λ,ϵ)-tempered ω, we see that the measures $\nu^s_x(C,\lambda,\epsilon)$ vary weak* continuously. Almost every word ω is (C,λ,ϵ)-tempered for some C. Thus we see that

$$\nu^s_x = \sum_{C=0}^{\infty} \nu^s_x(C,\lambda,\epsilon).$$

Note that each partial sum of this series varies weak* continuously and that the mass is uniformly absolutely summable pointwise. Thus the limiting family ν^s_x is seen to vary weak* continuously. □

B.4. **Construction of fake stable manifolds.** As mentioned above, we will define the fake stable manifolds by taking curves tangent to a smooth approximation to the distribution V_n, which is defined to equal $Df^n_\omega(E^s_n)$ as above. First, we note that Lemma B.3 above will be applicable to studying the regularity of V_n due to the following.

Lemma B.5. *Suppose that $A_1,\dots,A_n$ is a sequence of linear transformations that are (C,λ,ϵ)-tempered. Then the sequence $A_n^{-1},\dots,A_1^{-1}$ is $(C+\epsilon n,\lambda,\epsilon)$-tempered, and the corresponding splitting is the splitting with the stable and unstable subspaces from the original splitting swapped.*

Using Lemmata B.3 and B.5 we can estimate the regularity of V_n.

Lemma B.6. *Suppose that $(f_1,\dots,f_m)$ is a tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$ where M is a compact surface. Fix $C,\lambda>0$, then there exist $\beta,\eta>0$ such that for any sufficiently small $\epsilon>0$ there exists $D_1,N\in\mathbb{N}$, such that if Λ^n_ω is the set of points that are (C,λ,ϵ)-tempered at some time $n\geq N$, then the distribution V_n defined on $f^n_\omega(\Lambda^n_\omega)$ by $D_xf^n_\omega(E^s_n(x))$, is β-Hölder with constant $D_1e^{\eta\epsilon n}$.*

Proof. Apply Proposition B.3 with λ as above to the diffeomorphisms $(f_1^{-1},\dots,f_m^{-1})$. Then there exist β and ϵ_0 such that restricted to the set of (C,λ,ϵ)-tempered points at time $n\geq O_\lambda(C)$, E^s_n is β-Hölder with constant at most e^{D_1C}. From Lemma B.5, we see that for the backwards dynamics $(f_{\sigma^{n-i}(\omega)})^{-1}$, the points in $f^n_\omega(\Lambda^n_\omega)$ are $(C+\epsilon n,\lambda,\epsilon)$-tempered. Note that V_n is equal to the distribution of the most expanded direction for $(f^n_\omega)^{-1}$ and that $V_n^\perp$ is the most contracted direction of $(f^n_\omega)^{-1}$. As the set $f^n_\omega(\Lambda^n_\omega)$ is $(C+\epsilon n,\lambda,\epsilon)$-tempered for the backwards dynamics, it follows that as long as ϵ is sufficiently small and N_0 is sufficiently large, for all $n\geq N_0$, $V_n^\perp$ is $e^{D_1(C+\epsilon n)}$ β-Hölder. The statement of the lemma now follows. □

Next we take a smooth approximation $\widetilde{V}_n$ to the distribution V_n that will be defined in an open neighborhood of $f^n_\omega(\Lambda^n_\omega)$. First we extend the domain of V_n, and then we smooth the extension. If we do not extend the domain, then we won't be able to integrate the distribution. If we do not do this smoothing, then we will have little control over the norm of the integral curves to V_n rather than tempered growth in n.

Lemma B.7. *Suppose that M is a smooth closed surface. There exist D_1,D_2 such that if $K\subseteq M$ is a subset and E is a distribution defined over K that is (C,α)-Hölder then E admits a (D_1C,α)-Hölder extension to a neighborhood of K of size $\delta=D_2\min\{1,C^{-1/\alpha}\}$.*

Proof. We first prove the result with vector fields instead of distributions. Cover M by finitely many charts. In each chart the vector field X is represented as a map $\phi_0\colon K\to S^1\subset\mathbb{R}^2$. The McShane extension theorem [McS34, Cor. 1] says that a (C,α)-Hölder function defined from a subset X of an arbitrary metric space to $\mathbb{R}$ admits a (C,α)-extension to all of X. Then we glue the maps from different charts using a partition of unity. This proves the result for vector fields. Note that the resulting vector field is defined on the whole manifold. To obtain the result for distributions, we take a unit vector field on K in the direction of E, extend it to a vectorfield $\tilde{X}$ as above and note that the resulting extension is nonzero inside the δ neighborhood of K, so we can take $\tilde{E}$ to be the direction of $\tilde{X}$. □

The content of the following lemma is item (2), the C^2 estimate on $\widetilde{V}_n$. While V_n could be seen to be C^2, we have little ability to control its norm; thus we need to produce a more regular approximation to this distribution.

Lemma B.8. *Let $(f_1,\dots,f_m)$ be a tuple of diffeomorphisms in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$, for M a closed surface. Fix $\lambda>0$. Then there exists $\epsilon_1>0$, $\nu_1,\nu_2>0$ and $N\in\mathbb{N}$, D_1,D_2,D_3, such that if* *for $\epsilon < \epsilon_1$ Λ_n^ω denotes the set of (C,λ,ϵ)-tempered points, then there exists a distribution $\widetilde{V}_n$ such that*

1. *The domain of $\widetilde{V}_n$ contains all points within distance $D_1e^{-\epsilon\nu_1 n}$ of the domain of V_n.*
2. *$\widetilde{V}_n$ is C^2 with $\|\widetilde{V}_n\|_{C^2} \le D_2e^{\epsilon\nu_1 n}$.*
3. *At each x in the domain of V_n, $d(\widetilde{V}_n(x), V_n(x)) < D_3e^{-\epsilon\nu_2 n}$.*

Proof. First, from Lemma B.6, given $\epsilon > 0$ we may choose ϵ_1 sufficiently small that V_n is β-Hölder with constant $D_1e^{\eta\epsilon n}$. Let $\hat{V}_n$ be an extension of V_n obtained from Lemma B.7, then from the Hölder estimate on V_n, $\hat{V}_n$ is defined in a neighborhood of $Df_\omega^n(\Lambda_\omega^n)$ of size at least $D_1^{-1/\beta}e^{-\eta\epsilon n/\beta}$.

We now take a smooth approximation to $\hat{V}_n$. For this we can represent $\hat{V}_n$ in charts as a function $\phi\colon U \to S^1 \subset \mathbb{R}^2$, then mollify ϕ. From [FKS13, Eq. (11)], we have estimates for convolution $f_\epsilon = f * \psi_\epsilon$ of a standard mollifier ψ_ϵ with a compactly supported function $f\colon \mathbb{R}^2 \to \mathbb{R}$:

$$\|f_\epsilon\|_2 \le \epsilon^{\alpha-4}\|f\|_\alpha \quad \text{and} \quad \|f - f_\epsilon\|_0 \le \epsilon^\alpha\|f\|_\alpha. \tag{B.7}$$

As domain of $\hat{V}_n$ has size at least $D_1^{-1/\beta}e^{-\eta\epsilon n/\beta}$, we can mollify with any $\epsilon' < D_1^{-1/\beta}e^{-\eta\epsilon n/\beta}/100$ and obtain a function that is well defined at all points at least distance $D_1^{-1/\beta}e^{-\eta\epsilon n/\beta}/100$ from the boundary of the domain of $\hat{V}_n$. Let $\widetilde{V}_n$ denote the mollified function restricted to the points in the domain of $\hat{V}_n$ of distance at most $D_1^{-1/\beta}e^{-\eta\epsilon n/\beta}/100$ from the domain of V_n. Then taking $\epsilon' = e^{-\nu\epsilon}$ for some large ν, mollifying with $\psi_{\epsilon'}$, and applying the estimates in (B.7) gives that there exist constants D_2, D_3, D_4, D_5 such that

$$\|\widetilde{V}_n\|_2 \le D_2e^{-D_3\epsilon n} \text{ and } d(\widetilde{V}_n, V_n) < D_4e^{-D_5\epsilon n}.$$

This gives the needed conclusion. □

The use of the distributions $\widetilde{V}_n$ is that they are integrable and their C^2 norm is well controlled. This implies that if we take a holonomy along the distribution, then we will have good control of the norm of the Jacobian.

Definition B.9. Fix $\lambda > 0$ and sufficiently small $\epsilon > 0$. Then take $\epsilon_1 < \epsilon/\max\{\nu_1, \nu_2\}$ where ν_1, ν_2 are as in Proposition B.8. We consider a collection of (C,λ,ϵ_1)-tempered points. Let $\widetilde{W}_n$ be the foliation defined by the integral curves to $\widetilde{V}_n$. The fake stable leaf through $x \in \Lambda_\omega^n$ is then defined to be $W_n^s(\omega, x) = (f_\omega^n)^{-1}(\widetilde{W}_n(f_\omega^n(x)))$.

We will now state basic facts about the fake stable manifolds. In particular, we show that the fake stable manifolds of sufficiently small size enjoy uniform contraction.

Proposition B.10. *Suppose that $(f_1, \dots, f_m)$ is an expanding on average tuple of diffeomorphisms in $\mathrm{Diff}^2(M)$, where M is a closed surface. Fix $\lambda > 0$. Then there exists $\lambda', \epsilon_0 > 0$ such that for any $0 \le \epsilon \le \epsilon_0$ and any C, there exist $N_0, \delta_0, C_0, \alpha > 0$ such that if $\Lambda_n^\omega \subset M$ is any collection of (C,λ,ϵ)-tempered points at time $n \ge N_0$ lying in some ball $B_{\delta_0} \subset M$. Then*

1. *For $N_0 \le i \le n$ the fake stable manifolds $W_{i,\delta_0}^s(\omega, x)$ exist and have C^2 norm at most C_0.*
2. *$d(T_xW_i^s, E_i^s(x)) \le e^{-\lambda i/2}$.*
3. *The fake stable direction E_i^s is (C_0, α)-Hölder continuous on Λ_n^ω.*
4. *The fake stable leaves $W_{i,\delta_0}^s(\omega, x)$ vary Hölder continuously in the C^1 topology, and the Hölder constants are independent of $N_0 \le i \le n$.*

(5) The fake stable leaves $W^s_{i,\delta_0}(\omega,x)$ are contracting, i.e. for $y,z\in W^s_{i,\delta_0}(\omega,x)$, for each $0\le k\le i$, $d_{W^s_{i,\delta_0}(x)}(f^k_\omega(y),f^k_\omega(z))\le C_0e^{-\lambda' k}$.

Proof Sketch. The claim about the existence and regularity of the fake stable manifolds in (1) essentially follows from the construction of the stable manifolds described in Section 5 or Proposition A.13, depending on taste. An integral curve to the $\widetilde{V}_n$ distribution has C^2 norm that is order $e^{O(\epsilon)}$, and is almost tangent to the most expanded direction of $(Df^n_\omega)^{-1}$ allowing us to apply those lemmas. Similarly, the final item in the lemma says that the dynamics on the fake stable manifolds is contracting. This also follows from the graph transform argument. Specifically one can produce this statement by a generalization of Step 1 in the proof of Proposition A.13, which studies the growth in length of curves in the Lyapunov charts.

The statement (2) saying that $T_xW^s_i$ is near to E^s_i is immediate because $\|Df^n_\omega\|\ge Ce^{\lambda n}$ by assumption. Since $Df^n_\omega E^s_i$ and $\widetilde{V}_n$ are exponentially close, they will attract further under $(Df^n_\omega)^{-1}$.

The statements about Hölder-ness are standard facts; it follows from the same argument as in [BP07, Sec. 5.3] applied for only finitely many iterations. Alternatively, Lemma 10.2 contains an explicit computation showing that nearby points inherit a nearby splitting. The proof of that lemma does not rely on any of the claims from this section. We will not use (4) as everything we need for the main result of this paper follows from (1), (2), and (3). So will will omit detailed proof. The claim essentially follows Hölder continuity of the stable distribution, Hölder continuity of the holonomies, which will be obtained in Proposition B.13, and Lemma B.1. Compare for example, with [BP07, Sec. 8.1.5], which describes a similar argument. □

B.5. **Rate of convergence of fake stable manifolds.** Proposition B.12, proven in this section, is one of the key estimates in this paper playing an important role in the local coupling procedure.

The main crucial feature that the fake stable leaves exhibit is that the fluctuations in W^s_i as we increase i decay exponentially fast. In fact, we have a quantitative estimate that directly relates the speed of convergence of $W^s_i(\omega,x)$ with the hyperbolicty of $D_xf^i_\omega$.

In the following proposition, we will use an additional refinement of (C,λ,ϵ)-tempered points that also requires that the stable direction points in a particular direction. The definition below is structured so that it is hopefully straightforward to think about. When a point is (C,λ,ϵ)-tempered, there is a definite rate at which E^s_n converges to E^s. Thus if E^s_n happens to lie sufficiently far from the boundary of a cone $\mathcal{C}$ at a sufficiently large time n_1, then $E^s_i\in\mathcal{C}$ for all $i\ge n_1$.

Definition B.11. Suppose $x\in M$ and $\mathcal{C}\subset T_xM$ is a cone. We say that a word ω is $(C,\lambda,\epsilon,\mathcal{C},n_1,n_2)$-tempered if for all $n_1\le i\le n_2$, E^s_i is defined and lies in $\mathcal{C}$. We may also speak of being $(C,\lambda,\epsilon,\mathcal{C})$-tempered at a time n, in which case we mean $n_1=n_2=n$ in the previous sentence.

We now estimate how much the fake stable leaves fluctuate. The requirements on the cone are, strictly speaking, not necessary in order to state the theorem below: as long as N is chosen sufficiently large, one can use $E^s_N(x)$ to define the cone $\mathcal{C}$ in the following proposition and obtain the same result.

Proposition B.12. *(Fluctuations in fake-stable leaves) Let $(f_1,\ldots,f_m)$ be a tuple in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$ for a closed surface M. Fix $\lambda,C_1,\theta_0>0$, then there exists $\epsilon_0>0$ such that for all $0\le\epsilon<\epsilon_0$ and $C>0$ there exist $D_1,N,\delta_0>0$ such that for any $\delta\le\delta_0$ the following holds. Given*

$x \in M$ and a cone $\mathcal{C} \subset T_xM$, extend $\mathcal{C}$ by parallel transport to a conefield $\mathcal{C}$ defined over $B_{2\delta}(x)$. Suppose that γ is a C-good curve with distance $d(x,\gamma) < \delta$ and γ is θ_0 transverse to $\mathcal{C}$. If ω is a $(C_1, \lambda, \epsilon, \mathcal{C}, n, n+1)$-tempered with $n \geq N$, then

$$d_\gamma(W^s_n(x) \cap \gamma, W^s_{n+1}(x) \cap \gamma) \leq e^{-1.99 \ln \|D_x f^n_\omega\|}, \tag{B.8}$$

where $W^s_n(x), W^s_{n+1}(x)$ are the fake stable manifolds from Definition B.9.

The reason the proposition follows is evident in the case of a linear map. Consider the action of the map $L = \operatorname{diag}(\sigma, \sigma^{-1})$ on $\mathbb{R}\mathrm{P}^1$ where $\sigma > 1$. Note that the map L has an attracting fixed point of multiplier σ^{-2}, which suggests the asymptotic in the theorem. Consider what happens if we apply L to two curves tangent at $(0,0)$ to the expanded direction of L: the distance between them will contract by a factor of σ^{-2}. The result for a sequence of maps will follow because the temperedness assures a uniform E^s, E^u splitting. When we work with this splitting, the full strength of the hyperbolicity will be available allowing us to recover almost $e^{-2 \ln \|D_x f^n_\omega\|}$ contraction as in the theorem.

The formal proof will rely on the study of the graph transform. The argument for this proposition is simpler than the argument in the recovery lemma since the curves we consider in this lemma are (by assumption) well positioned with respect to the stable and unstable splitting.

There are three steps in the proof. We have two curves at $f^n_\omega(x)$, one corresponding to the time n fake stable manifolds and one corresponding to the time $n+1$ fake stable manifolds. In the first step, we iterate the graph transform until these curves look uniformly Lipschitz in the Lyapunov charts. In the second step, we iterate the graph transform to see that these two curves approach each other at the appropriate exponential rate. In the third step, we do some bookkeeping to conclude.

Proof. Recall that, by definition, the fake stable manifold $W^s_n(x)$ is given by taking a curve γ_n tangent to the distribution $\widetilde{V}_n$ from Lemma B.8 and letting $W^s_n(x)$ equal $(f^n_\omega)^{-1}(\gamma_n)$ restricted to a segment of length δ_0 about x where δ_0 is chosen as in Proposition B.10. Note that we need not take the δ_0 in this proposition to be the same as the one in Proposition B.10. Indeed, at certain points in the analysis below it may be convenient to decrease δ_0 in a way that depends only on the parameters of the proposition.

The proposition is comparing $(f^n_\omega)^{-1}(\gamma_n)$ and $(f^{n+1}_\omega)^{-1}(\gamma_{n+1})$. As in previous sections, we will view both of these curves as graphs of functions from E^u to E^s in the Lyapunov charts. In this proof we will work with the splitting into stable and unstable subspaces for the subspaces defined by the associated splitting for Df^n_ω rather than Df^{n+1}_ω. Recall that E^s_n denotes the most contracted subspace for Df^n_ω and E^s_{n+1} denotes the most contracted subspace for Df^{n+1}_ω.

In the Lyapunov charts at $f^j_\omega(x)$, we write $(f^{n-j}_{\sigma^j(\omega)})^{-1}(\gamma_n)$ as the graph of the function ϕ^1_j and we write $(f^{n-j+1}_{\sigma^j(\omega)})^{-1}(\gamma_{n+1})$ as the graph of $\phi^2_j(x)$. Let e^Λ be an upper bound on $\|Df_i\|$, $1 \leq i \leq m$, with $\Lambda > 100$.

With respect to the Lyapunov metrics, we use the similar choices as in previous arguments, specifically Proposition A.15, and thereby obtain essentially identical intermediate estimates. View the sequence of maps f^j_ω as being reversed tempered starting at $f^{n+1}_\omega(x)$ and ending at x. So, set $\lambda' = .9999\lambda$ and take the finite time Lyapunov metrics as in Lemma A.1 for this sequence. In particular, note that from the construction of the Lyapunov metrics, the $e^{O(\epsilon n)}$ bound on the C^2 norm of the curves γ_n from Lemma B.8 and the angle $\widetilde{V}_n$ makes with V_n of $O\left(e^{O(-\epsilon n)}\right)$ combine to show that there exist $C_2, \nu > 0$ such that $\|\phi^1_n\|_1, \|\phi^2_n\|_1 \leq C_2 e^{\nu\epsilon n}$. We now proceed with the proof.

Step 1. (Lipschitzness) In this step, we will identify $N_l \approx (1 - O(\epsilon))n$ such that for $j \leq N_l$, ϕ_j^1 and ϕ_j^2 are C^0 close.

To begin we estimate how far apart $Df_\omega^n(E_n^s)$ and $Df_\omega^n(E_{n+1}^s)$ are. We claim that there exists N_0 such that for $n \geq N_0$, then $\angle(Df_\omega^n(E_n^s), Df_\omega^n(E_{n+1}^s)) \leq 1/4$. Note that if N_0 is sufficiently large that both $\|Df_\omega^n\|$ and $\|Df_\omega^{n+1}\|$ are at least $e^{10\Lambda}$ and $\angle(E_n^s, E_{n+1}^s) < 1/100$ both of which follow from the (C_1, λ, ϵ)-temperedness (The latter claim is part of Proposition 4.6). As in previous computations, it follows that if $\angle(Df_\omega^n E_n^s, Df_\omega^n E_{n+1}^s) > 1/4$, then $\|Df_\omega^n(E_{n+1}^s)\| > 2$ because Df_ω^n expands E_n^u and contracts E_n^s. Consequently, $\|Df_\omega^{n+1}(E_{n+1}^s)\| > 2e^{-\Lambda}$. But this is not less than $e^{-10\Lambda}$, so it is impossible that $\angle(Df_\omega^n(E_n^s), Df_\omega^n(E_{n+1}^s)) > 1/4$.

Note that in Proposition A.13, we considered smoothing estimates for a reverse tempered point. In the case of this theorem, we may consider x as a reverse tempered point for the sequence of maps $(f^j_{\sigma^{n-j}(\omega)})^{-1}$ beginning at $f_\omega^n(x)$. Consequently, we may read off the intermediate estimates from the proof of that theorem. In particular, as in equation (A.44) by possibly restricting the domain of ϕ_j^1 and ϕ_j^2 as in that proposition, it follows that there exists C_3 such that for $i \in \{1, 2\}$ that

$$\|\phi_{n-j}^i\|_1 \leq C_3 e^{\nu\epsilon n} e^{-j\lambda}.$$

In particular this shows that if we let $N_l = \lfloor n - \nu\epsilon/\lambda n \rfloor$, then because both curves pass through 0 and our choice of N_l, we see that there exists C_4 such that for $i \in \{1, 2\}$, $\|\phi_j^i\|_1 \leq C_4$. Because both pass through 0, the following estimate holds for all $N_0 \leq j \leq N_l$:

$$\left|\phi_j^1(x) - \phi_j^2(x)\right| \leq 2C_4 |x|, \tag{B.9}$$

which is the desired estimate for this step in the proof.

Step 2. (Contraction) In this step, we study how fast the curves ϕ_j^1 and ϕ_j^2 attract as we apply the dynamics $(f_{\sigma^j(\omega)})^{-1}$. Our goal is to show that the C^0 distance between these functions is rapidly decreasing, which is the content of (B.14).

First, in the Lyapunov chart we have

$$\hat{f}^{-1}_{\sigma^j(\omega)} = (e^{\sigma_j^1} x + \hat{f}_{j,1}(x, y), e^{\sigma_j^2} y + \hat{f}_{j,2}(x, y)), \tag{B.10}$$

where $\min\{\sigma_j^1, -\sigma_j^2\} \geq .999\lambda$. Then in the Lyapunov charts, the differential is

$$D\hat{f}^{-1}_{\sigma^j(\omega)} = \begin{bmatrix} e^{\sigma_{j,1}} + \partial_x \hat{f}_{j,1} & \partial_y \hat{f}_{j,1} \\ \partial_x \hat{f}_{j,2} & e^{\sigma_{j,2}} + \partial_y \hat{f}_{j,2} \end{bmatrix}. \tag{B.11}$$

In addition, write

$$\Lambda_j = \sum_{i=j}^{N_l} \sigma_{j,1} - \sigma_{j,2}. \tag{B.12}$$

As in Proposition A.13, we have a C^2 estimate in the Lyapunov charts. There exists $C_5 > 0$ such that

$$\|(\hat{f}_{\sigma^i(\omega)})^{-1}\|_{C^2} \leq C_5 e^{6C_1} e^{6i\epsilon}. \tag{B.13}$$

We will now verify inductively that a strengthening of (B.9) holds for $N_0 < j < N_l$. We now show that by possibly increasing N_0, which is fixed and does not depend on n, that for all $|x| < e^{-(\lambda/2)j}$, and $N_0 \leq j < N_l$,

$$\left|\phi_j^1(x) - \phi_j^2(x)\right| \leq C_4 e^{-1.999\Lambda_j} |x|. \tag{B.14}$$

To show (B.14), we measure the distance between ϕ_j^1 and ϕ_j^2 using a piece of the vertical curve $V(t)$ parallel to E^s between $\phi_{j+1}^1(x)$ and $\phi_{j+1}^2(x)$. We then apply $(f_{\sigma^j(\omega)})^{-1}$ to the curve and estimate its length. We then use the Lipschitzness of ϕ_j^1 and ϕ_j^2 to obtain (B.14). Let $V(t)$ be a vertical curve (parallel to E^s) defined on $[-1,1]$ taking values in the Lyapunov charts such that $V(-1)\in\phi_{j+1}^1$ and $V(1)\in\phi_{j+1}^2$ passing through the point $(x,0)$. Then from the inductive hypothesis, we see that $\operatorname{len}(V)\le C_4e^{-1.999\Lambda_j}\,|x|$.

By applying the differential to V, we see by (B.11), $(\hat f_{\sigma^j(\omega)})^{-1}(V)$ is tangent to a vector of the form

$$\partial_t((\hat f_{\sigma^j(\omega)})^{-1}V(t))=\begin{bmatrix}\partial_y\hat f_{j,1}\\ e^{\sigma_{j,2}}+\partial_y\hat f_{j,2}\end{bmatrix}. \tag{B.15}$$

In particular, for C_5 as before if we are restricted to a ball of radius $C_5^{-1}e^{-(\lambda/2)j}$, then as the C^2 norm of $(\hat f_{\sigma^j(\omega)})^{-1}$ is $O(e^{6j\epsilon})$, it follows that

$$\left|\partial_y\hat f_{j,i}\right|<e^{-(\lambda/4)j} \tag{B.16}$$

for $i\in\{1,2\}$. Let π_u be the projection onto the E^u direction in the Lyapunov coordinates and let π_s be the projection onto the E^s direction in the Lyapunov coordinates. We see that there exists C_6 such that:

$$\left|\pi_s((\hat f_{\sigma^j(\omega)})^{-1}V(-1))-\pi_s((\hat f_{\sigma^j(\omega)})^{-1}V(1)))\right|\le C_4e^{-1.999\Lambda_j}e^{(1-\epsilon_j)\sigma_{j,2}}\,|x| \tag{B.17}$$

where $|\epsilon_j|\le C_6e^{-\lambda/4j}$.

We now use (B.17) to estimate the C^0 norm of ϕ_j^1 and ϕ_j^2, rather than just the distance between two points along these curves. The endpoints of $(\hat f_{\sigma^j(\omega)})^{-1}V(t)$ lie in ϕ_j^1 and ϕ_j^2. Note that when $(\hat f_{\sigma^j(\omega)})^{-1}V$ is viewed as a graph over the vertical line parallel to E^s through $\pi_u(\hat f_{\sigma^j\omega})^{-1}(x,0)$, that $(\hat f_{\sigma^j(\omega)})^{-1}V$ is distance at most $e^{-\lambda/4j}\operatorname{len}(V)$ from a vertical line by (B.15) and (B.16). Thus as ϕ_j^1 and ϕ_j^2 are both C_4 Lipschitz for $N_0\le j\le N_l$, we see that

$$\begin{aligned}\left|\phi_j^1(\pi_1(\hat f_{\sigma^j\omega})^{-1}(x,0))-\phi_j^2(\pi_1(\hat f_{\sigma^j\omega})^{-1}(x,0))\right|&<C_4e^{-1.999\Lambda_j}e^{(1-\epsilon_j)\sigma_{j,2}}\,|x|+C_4e^{-\lambda/4j}\operatorname{len}(V)\\ &\le(e^{(1-\epsilon_j)\sigma_{j,2}}+C_4e^{-\lambda/4j})e^{-1.999\Lambda_j}\,|x|\,.\end{aligned} \tag{B.18}$$

As long as N_0 is sufficiently large, for $j\ge N_0$,

$$|x|\le e^{-(1-\epsilon_j)\sigma_{j,1}}\left|\pi_1(\hat f_{\sigma^j(\omega)})^{-1}(x,0))\right|. \tag{B.19}$$

Note that if j is larger than some fixed N_0 and ϵ_j is sufficiently small relative to λ, then

$$(e^{(1-\epsilon_j)\sigma_{j,2}}+C_4e^{-\lambda/4j})e^{-(1-\epsilon_j)\sigma_{j,1}}\le e^{1.999(\sigma_{j,2}-\sigma_{j,1})}. \tag{B.20}$$

Combining (B.18), (B.19), and (B.20), we get $\left|\phi_j^1(x)-\phi_j^2(x)\right|\le C_4e^{-1.999\Lambda_{j-1}}$, as required.

Step 3. (Bookkeeping and Conclusion) So far, we have obtained that for some N_0 and C_4 depending only on the constants in the theorem

$$\left|\phi_{N_0}^1(x)-\phi_{N_0}^2(x)\right|\le C_4e^{-1.999\Lambda_{N_0}}$$

Thus as ϕ_0^1 and ϕ_0^2 are related to $\phi_{N_0}^1$ and $\phi_{N_0}^2$ by applying only the fixed number N_0 more maps, we see that there exists C_7 and $\delta_2>0$ such that on a ball of radius δ_2 in the Lyapunov charts at x:

$$\left|\phi_0^1(x)-\phi_0^2(x)\right|\le C_7e^{-1.999\Lambda_{N_0}}.$$

Consider a nearby C-good curve γ that is θ_0-transverse to $\mathcal{C}$ and hence to E^s, ϕ_0^1, and ϕ_0^2. It then follows easily from transversality, that as ϕ_0^1 is nearly tangent to E^s by Proposition B.10(2) and ϕ_0^1, ϕ_0^2 are uniformly Lipschitz, there exists C_8 such that

$$d_\gamma(\phi_0^1 \cap \gamma, \phi_0^2 \cap \gamma) \leq C_8 e^{-1.999\Lambda_{N_0}}.$$

The only remaining thing we need is to know that Λ_{N_0} is within a factor of $.001\Lambda$ of $\ln \|Df_\omega^n\|$. This will follow as long as we take ϵ sufficiently small relative to λ, ν_1, ν_2 and the maximum of the norm of the differentials of $f_1, \ldots, f_m$. We omit the computation of exactly how small ϵ must be. Such sufficiently small ϵ exists because when we look in the Lyapunov charts, we obtain the straightforward bound that there exists C_9 such that

$$\ln \|Df_\omega^n\| \leq C_9 + 4\epsilon n + \sum \sigma_{j,1}.$$

But Λ_{N_0} includes only the hyperbolicity for the iterates $N_0 \leq j \leq N_l$. From volume preservation of the f_i, it similarly follows that $\ln \|Df^n\| \leq C_{10} + 4\epsilon n - \sum \sigma_{j,2}$ for some C_{10}. As $N_l = (1 - O(\epsilon))n$ and N_0 is a fixed independent of n, it follows that for sufficiently small ϵ and sufficiently large n that $e^{1.99 \ln \|Df^n\|} \leq e^{1.999\Lambda_{N_0}}$, which is the needed conclusion. □

B.6. **Jacobian of the fake stable holonomies.** Now that we have defined the fake stable manifolds and have an estimate for the rate at which their holonomies converge, we study the Jacobian of their holonomies, whose properties are crucial in the coupling argument. The next quantity of interest is the fluctuations in the Jacobian of the holonomies for the fake stable manifolds.

Proposition B.13. *Suppose that $(f_1, \ldots, f_m)$ is a tuple of diffeomorphisms in $\mathrm{Diff}^2_{\mathrm{vol}}(M)$ for a closed surface M. For $\lambda > 0$ there exists $\epsilon_0 > 0$ such that for all $0 \leq \epsilon \leq \epsilon_0$ and $C > 0$, there exists $N \in \mathbb{N}$ and $\delta, \eta, \alpha > 0$ such that for any $n \geq N$, and any $\omega \in \Sigma$, if Λ_n^ω is the set of (C, λ, ϵ)-tempered points up to time n then for any ball $B_\delta \subseteq M$ of radius δ, the following holds for $x \in \Lambda_n^\omega \cap B_\delta$.*

For any two uniform transversals T_1 and T_2 to the W_N^s laminations of $B_\delta(x)$, T_1 and T_2 will be uniform transversals to the W_i^s lamination for $N \leq i \leq n$. Where defined, consider the holonomies H_i^s between T_1 and T_2 and moreover the Jacobian $\mathrm{Jac}\, H_i^s$, which is defined on a subset of T_1. Then we have the following for all $N \leq i \leq n$:

(1) The Jacobians of the holonomies between uniform transversals are uniformly α-Hölder and bounded away from zero. In particular, this implies that these Jacobians are uniformly log-α-Hölder between uniform transversals. Specifically, for fixed (C_1, δ_1), there exist D_1, D_2, D_3 such that if γ_1 and γ_2 are a (C_1, δ_1)-configuration in the sense of Definition 7.8 with γ_1 and γ_2 uniformly transverse to the $E_N^s(x)$ extended by parallel transport in a small neighborhood, and $I \subseteq \Lambda_n^\omega$ is a subset of γ_1 then, for $x, y \in I$,

$$|\log \mathrm{Jac}\, H_n^s(x) - \log \mathrm{Jac}\, H_n^s(y)| \leq D_1 d_{\gamma_1}(x, y)^\alpha. \tag{B.21}$$

(2) The Jacobians from item (1) converge exponentially quickly, i.e.

$$\left|\mathrm{Jac}\, H_{i-1}^s - \mathrm{Jac}\, H_i^s\right| \leq D_2 e^{-\eta i}, \tag{B.22}$$

and

$$\left|\frac{\mathrm{Jac}\, H_i^s}{\mathrm{Jac}\, H_{i-1}^s} - 1\right| \leq D_3 e^{-\eta i}. \tag{B.23}$$

(3) The true stable holonomy restricted to $\Lambda_\infty^\omega \cap T_1$ is absolutely continuous. The Jacobian of the fake stable holonomies converges to the Jacobian of the true stable holonomies restricted

to the set $\Lambda^\omega_\infty \cap T_1$*. Namely, for almost every point of this intersection,* $\operatorname{Jac} H^s_n \to \operatorname{Jac} H^s$*, this convergence is uniform, and the limit is uniformly Hölder and bounded away from zero.*

Proof. **Part 1.** (Formula for Jacobian) We begin by exhibiting a formula for the Jacobian of the stable holonomies. This may be compared with [BP07, Sec. 8.6.4], which uses a similar formula though analyzes it differently. Suppose that T_1 and T_2 are the two transversals we are considering as in the statement of the proposition. Then write Π^s_i for the holonomy along $f^i_\omega(W^s_i) = \widetilde{W}^s_i$, the smooth integral curves to $\widetilde{V}_i$ we used when defining the fake stable foliation. Then we have the following formula for the Jacobian of H^s_i:

$$\operatorname{Jac}(H^s_i)(y) = \prod_{k=0}^{i-1} \frac{\operatorname{Jac}(D(f_{\sigma^k\omega})^{-1}|T_{f^k_\omega H^s_i(y)} f^k_\omega(T^2))}{\operatorname{Jac}(D(f_{\sigma^k\omega})^{-1}|T_{f^k_\omega(y)} f^k_\omega(T^1))} \operatorname{Jac}(\Pi^s_i(y)). \tag{B.24}$$

For finite time this formula is evident because all of the foliations we are considering are smooth: it is just the change of variables formula.

Part 2. (Exponential convergence) Applying Lemma B.1 we will obtain Hölder continuity for the Jacobians once we know that $\operatorname{Jac}(H^s_i)$ is converging exponentially fast.

To see that (B.24) converges exponentially quickly, two estimates are needed.

(1) The first is showing that for some $\eta > 0$

$$|\operatorname{Jac}(\Pi^s_n) - 1| \le C_1 e^{-n\delta_1}. \tag{B.25}$$

This is the Jacobian of the foliation holonomy of $\widetilde{W}^s_n$. The foliation holonomy is between two transversals that are distance $e^{-(\lambda/2)n}$ apart. By working in Lyapunov charts, it is straightforward to see that the $f^n_\omega(T_1)$ and $f^n_\omega(T_2)$ make angle at least $Ce^{-\epsilon n}$ with $\widetilde{W}^s_n$. As $\widetilde{W}^s_n$ itself has C^2 norm at most $e^{O(\epsilon n)}$ from Lemma B.8, it is easy to see that there exists some $C_1, \delta_1 > 0$ such that (B.25) holds.

(2) Next we estimate the rate of convergence of:

$$\prod_{k=0}^{i-1} \frac{\operatorname{Jac}(D(f_{\sigma^k\omega})^{-1}|T_{f^k_\omega H^s_i(y)} f^k_\omega(T^2))}{\operatorname{Jac}(D(f_{\sigma^k\omega})^{-1}|T_{f^k_\omega(y)} f^k_\omega(T^1))} = \exp\left(\sum_{k=0}^{i-1} P(k,i)\right), \tag{B.26}$$

where $P(k,i)$ is the logarithm of the kth term of the product.

We claim that there exist C_2, δ_2, N_2, such that for $i \ge N_2$ and $k \ge 0$,

$$\left| \frac{\operatorname{Jac}(D(f_{\sigma^k\omega})^{-1}|T_{f^k_\omega H^s_i(y)} f^k_\omega(T^2))}{\operatorname{Jac}(D(f_{\sigma^k\omega})^{-1}|T_{f^k_\omega(y)} f^k_\omega(T^1))} - 1 \right| \le C_2 e^{-\delta_2 k}. \tag{B.27}$$

We will not give a detailed proof of this estimate because it standard. The key claim is that if V_1 and V_2 are the tangent vectors to γ_1 and γ_2 at y and $H^s_i(y)$, respectively, then there exists a uniform constant C_2' and $\varpi > 0$ such that when we identify $Df^k_\omega V_1$ and $Df^k_\omega V_2$ by parallel transport along the distance minimizing geodesic between their basepoints, then

$$d(Df^k_\omega V_1, Df^k_\omega V_2) \le C_2' e^{-k\varpi}. \tag{B.28}$$

One can deduce this in a very similar way to the argument for [Mn87, Lem. III.3.7], which inductively checks that as one applies more iterates of the dynamics that these two vectors attract exponentially quickly by using that the basepoints of the vectors do as well; this argument is similar to the proof of our Proposition 10.3. Once (B.28) is known, then it is straightforward to conclude (B.27) because the Jacobian of a diffeomorphism $f\colon M \to M$ restricted to a curve $\gamma \subset M$ depends Hölder continuously on the direction of $\dot\gamma$.

(B.27) shows that the product (B.26) is uniformly bounded. It then suffices to estimate:

$$\sum_{k=0}^{i-1} P(k,i) - \sum_{k=0}^{i} P(k,i+1).$$

We will pick some $0<\theta<1$, and split this sum as follows:

$$\sum_{k=0}^{\theta i} (P(k,i) - P(k,i+1)) + \left[\sum_{k\geq \theta i}^{i} P(k,i) - \sum_{k\geq \theta i}^{i} P(k,i+1)\right] = I + II.$$

For any such θ, it follows from (B.27) that there exists $C_3, \delta_3 > 0$ such that $|II| \leq C_3 e^{-\delta_3 i}$. Thus to conclude we need only bound term I. From Proposition B.12 and the temperedness, we know that there exists C_4, δ_4 such that

$$d_{T_2}(H_i^s(y), H_{i+1}^s(y)) \leq C_4 e^{-\delta_4 i}. \tag{B.29}$$

It is straightforward to see that there exists $\beta, \beta_1 > 0$ such that the function

$$\frac{\mathrm{Jac}(D(f_{\sigma^k\omega})^{-1}|T_{f_\omega^k H_i^s(y)} f_\omega^k(T^2))}{\mathrm{Jac}(D(f_{\sigma^k\omega})^{-1}|T_{f_\omega^k(y)} f_\omega^k(T^1))}$$

viewed as a function of $H_i^s(y)$ is β-Hölder with the Hölder constant at most $e^{\beta_1 k}$ for all $k \leq i$. Thus by combining (B.29) with the Hölder continuity, we see that $|P(k,i) - P(k,i+1)| \leq e^{\beta k} e^{-\delta i}$. Thus as long as $\theta > \beta/\delta$, we see that there exists C_5, δ_5, such that

$$|I| \leq C_5 e^{-\delta_5 i}.$$

Combining the estimates on I and II implies that there exists C_6, δ_6 so that (B.26) is converging exponentially fast, as desired.

Thus we see that the Jacobian of the holonomies converges exponentially fast pointwise and is uniformly positive. Thus we have concluded (2) of the statement of the proposition.

Part 3. (Uniform Hölderness) We now apply Lemma B.1. We have just shown that the Jacobian of the holonomies is converging exponentially fast, and certainly the Hölder norm of the terms is growing at most exponentially fast as well as it is the composition of diffeomorphisms along with a holonomy, whose Hölder norm is also growing at most exponentially fast. Thus we conclude (1) above.

Part 4. The final claim (3) about the holonomies is fairly standard. The following lemma implies the conclusion:

Lemma B.14. *Let γ_1 and γ_2 be two curves with finite Lebesgue measure and for $n \in \mathbb{N}$ let $\Omega_n \subseteq \gamma_1$ be a decreasing sequence of subsets, each of which is a union of intervals. Suppose that $K := \bigcap_{n\geq\mathbb{N}} \Omega_n$ is compact. Let $\phi_n\colon \Omega_n \to \gamma_2$ be a sequence of absolutely continuous maps with uniformly continuous, equicontinuous Jacobians J_n. If (ϕ_n) converges uniformly to an injective map $\phi\colon K \to \gamma_2$, and $J_n|_k$ converges uniformly to an integrable function $J\colon K \to \mathbb{R}$, then ϕ is absolutely continuous with Jacobian J.*

We will not include a proof of the above lemma since it is a variant of a lemma in Mañé [Mn87, Thm. 3.3] and the proof of [Mn87] can be modified to obtain a proof of this lemma. □

References

[ABR22] José F. Alves, Wael Bahsoun, and Marks Ruziboev, *Almost sure rates of mixing for partially hyperbolic attractors*, J. Differential Equations **311** (2022), 98–157. MR 4354854

[ABRV23] José F. Alves, Wael Bahsoun, Marks Ruziboev, and Paulo Varandas, *Quenched decay of correlations for nonuniformly hyperbolic random maps with an ergodic driving system*, Nonlinearity **36** (2023), no. 6, 3294–3318. MR 4588339

[AGT06] Artur Avila, Sébastien Gouëzel, and Masato Tsujii, *Smoothness of solenoidal attractors*, Discrete Contin. Dyn. Syst. **15** (2006), no. 1, 21–35. MR 2191383

[Arn98] Ludwig Arnold, *Random dynamical systems*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 1998. MR 1723992

[AV10] Artur Avila and Marcelo Viana, *Extremal Lyapunov exponents: an invariance principle and applications*, Invent. Math. **181** (2010), no. 1, 115–189. MR 2651382

[Bal00] Viviane Baladi, *Positive transfer operators and decay of correlations*, Advanced Series in Nonlinear Dynamics, vol. 16, World Scientific Publishing Co., Inc., River Edge, NJ, 2000. MR 1793194

[BC91] Michael Benedicks and Lennart Carleson, *The dynamics of the Hénon map*, Ann. of Math. (2) **133** (1991), no. 1, 73–169. MR 1087346

[BCS22] Jérôme Buzzi, Sylvain Crovisier, and Omri Sarig, *Measures of maximal entropy for surface diffeomorphisms*, Ann. of Math. (2) **195** (2022), no. 2, 421–508. MR 4387233

[BCS23] ———, *On the existence of SRB measures for C^∞ surface diffeomorphisms*, Int. Math. Res. Not. IMRN (2023), no. 24, 20812–20826. MR 4681273

[BCZG23] Alex Blumenthal, Michele Coti Zelati, and Rishabh S. Gvalani, *Exponential mixing for random dynamical systems and an example of Pierrehumbert*, Ann. Probab. **51** (2023), no. 4, 1559–1601. MR 4597327

[BG20] Michael Björklund and Alexander Gorodnik, *Central limit theorems for group actions which are exponentially mixing of all orders*, J. Anal. Math. **141** (2020), no. 2, 457–482. MR 4179768

[BL85] Philippe Bougerol and Jean Lacroix, *Products of random matrices with applications to Schrödinger operators*, Progress in Probability and Statistics, vol. 8, Birkhäuser Boston, Inc., Boston, MA, 1985. MR 886674

[BO21] Snir Ben Ovadia, *Hyperbolic SRB measures and the leaf condition*, Comm. Math. Phys. **387** (2021), no. 3, 1353–1404. MR 4324380

[Bow75] Rufus Bowen, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Mathematics, vol. Vol. 470, Springer-Verlag, Berlin-New York, 1975. MR 442989

[BP07] Luis Barreira and Yakov Pesin, *Nonuniform hyperbolicity*, Encyclopedia of Mathematics and its Applications, vol. 115, Cambridge University Press, Cambridge, 2007, Dynamics of systems with nonzero Lyapunov exponents. MR 2348606

[BRH17] Aaron Brown and Federico Rodriguez Hertz, *Measure rigidity for random dynamics on surfaces and related skew products*, J. Amer. Math. Soc. **30** (2017), no. 4, 1055–1132. MR 3671937

[Bur24] David Burguet, *SRB measures for C^∞ surface diffeomorphisms*, Invent. Math. **235** (2024), no. 3, 1019–1062. MR 4701884

[BW10] Keith Burns and Amie Wilkinson, *On the ergodicity of partially hyperbolic systems*, Ann. of Math. (2) **171** (2010), no. 1, 451–489. MR 2630044

[BXY17] Alex Blumenthal, Jinxin Xue, and Lai-Sang Young, *Lyapunov exponents for random perturbations of some area-preserving maps including the standard map*, Ann. of Math. (2) **185** (2017), no. 1, 285–310. MR 3583355

[BXY18] ———, *Lyapunov exponents and correlation decay for random perturbations of some prototypical 2D maps*, Comm. Math. Phys. **359** (2018), no. 1, 347–373. MR 3781453

[CE80] Pierre Collet and Jean-Pierre Eckmann, *Iterated maps on the interval as dynamical systems*, Progress in Physics, vol. 1, Birkhäuser, Boston, MA, 1980. MR 613981

[Che06] N. Chernov, *Advanced statistical properties of dispersing billiards*, J. Stat. Phys. **122** (2006), no. 6, 1061–1094. MR 2219528

[Chu20] Ping Ngai Chung, *Stationary measures and orbit closures of uniformly expanding random dynamical systems on surfaces*, 2020, `https://arxiv.org/abs/2006.03166`.

[CL22] Roberto Castorrini and Carlangelo Liverani, *Quantitative statistical properties of two-dimensional partially hyperbolic systems*, Adv. Math. **409** (2022), Paper No. 108625, 122. MR 4469072

[CLP22] Vaughn Climenhaga, Stefano Luzzatto, and Yakov Pesin, *SRB measures and Young towers for surface diffeomorphisms*, Ann. Henri Poincaré **23** (2022), no. 3, 973–1059. MR 4396671

[CM06] Nikolai Chernov and Roberto Markarian, *Chaotic billiards*, Mathematical Surveys and Monographs, vol. 127, American Mathematical Society, Providence, RI, 2006. MR 2229799

[CV13] A. Castro and P. Varandas, *Equilibrium states for non-uniformly expanding maps: decay of correlations and strong stability*, Ann. Inst. H. Poincaré C Anal. Non Linéaire **30** (2013), no. 2, 225–249. MR 3035975

[dCJ02] Augusto Armando de Castro Júnior, *Backward inducing and exponential decay of correlations for partially hyperbolic attractors*, Israel J. Math. **130** (2002), 29–75. MR 1919371

[DeW24] Jonathan DeWitt, *Simultaneous linearization of diffeomorphisms of isotropic manifolds*, J. Eur. Math. Soc. (JEMS) **26** (2024), no. 8, 2897–2969. MR 4756948

[DFL22] Dmitry Dolgopyat, Bassam Fayad, and Sixu Liu, *Multiple Borel-Cantelli lemma in dynamics and multilog law for recurrence*, J. Mod. Dyn. **18** (2022), 209–289. MR 4447598

[DK07] Dmitry Dolgopyat and Raphaël Krikorian, *On simultaneous linearization of diffeomorphisms of the sphere*, Duke Math. J. **136** (2007), no. 3, 475–505. MR 2309172

[DKK04] Dmitry Dolgopyat, Vadim Kaloshin, and Leonid Koralov, *Sample path properties of the stochastic flows*, Ann. Probab. **32** (2004), no. 1A, 1–27. MR 2040774

[DKRH24] D. Dolgopyat, A. Kanigowski, and F. Rodriguez Hertz, *Exponential mixing implies Bernoulli*, Ann. of Math. (2) **199** (2024), no. 3, 1225–1292. MR 4740539

[DL23] Mark F. Demers and Carlangelo Liverani, *Projective cones for sequential dispersing billiards*, Comm. Math. Phys. **401** (2023), no. 1, 841–923. MR 4604909

[Dol00] Dmitry Dolgopyat, *On dynamics of mostly contracting diffeomorphisms*, Comm. Math. Phys. **213** (2000), no. 1, 181–201. MR 1782146

[EL] Alex Eskin and Elon Lindenstrauss, *Random walks on locally homogeneous spaces.*

[ES23] Rosemary Elliott Smith, *Uniformly expanding random walks on manifolds*, Nonlinearity **36** (2023), no. 11, 5955–5972. MR 4656974

[FKS13] David Fisher, Boris Kalinin, and Ralf Spatzier, *Global rigidity of higher rank Anosov actions on tori and nilmanifolds*, J. Amer. Math. Soc. **26** (2013), no. 1, 167–198, With an appendix by James F. Davis. MR 2983009

[Gal10] Stefano Galatolo, *Hitting time in regular sets and logarithm law for rapidly mixing dynamical systems*, Proc. Amer. Math. Soc. **138** (2010), no. 7, 2477–2487. MR 2607877

[GL06] Sébastien Gouëzel and Carlangelo Liverani, *Banach spaces adapted to Anosov systems*, Ergodic Theory Dynam. Systems **26** (2006), no. 1, 189–217. MR 2201945

[GS14] Alexander Gorodnik and Ralf Spatzier, *Exponential mixing of nilmanifold automorphisms*, J. Anal. Math. **123** (2014), 355–396. MR 3233585

[HJ13] Roger A. Horn and Charles R. Johnson, *Matrix analysis*, second ed., Cambridge University Press, Cambridge, 2013. MR 2978290

[Hör76] Lars Hörmander, *The boundary problems of physical geodesy*, Arch. Rational Mech. Anal. **62** (1976), no. 1, 1–52. MR 602181

[Kif86] Yuri Kifer, *Ergodic theory of random transformations*, Progress in Probability and Statistics, vol. 10, Birkhäuser Boston, Inc., Boston, MA, 1986. MR 884892

[KM96] D. Y. Kleinbock and G. A. Margulis, *Bounded orbits of nonquasiunipotent flows on homogeneous spaces*, Sinaĭ's Moscow Seminar on Dynamical Systems, Amer. Math. Soc. Transl. Ser. 2, vol. 171, Amer. Math. Soc., Providence, RI, 1996, pp. 141–172. MR 1359098

[Liu16] Xiao-Chuan Liu, *Lyapunov exponents approximation, symplectic cocycle deformation and a large deviation theorem*, ProQuest LLC, Ann Arbor, MI, 2016, Thesis (Ph.D.)–IMPA.

[Liv04] Carlangelo Liverani, *On contact Anosov flows*, Ann. of Math. (2) **159** (2004), no. 3, 1275–1312. MR 2113022

[LQ95] Pei-Dong Liu and Min Qian, *Smooth ergodic theory of random dynamical systems*, Lecture Notes in Mathematics, vol. 1606, Springer-Verlag, Berlin, 1995. MR 1369243

[McS34] E. J. McShane, *Extension of range of functions*, Bull. Amer. Math. Soc. **40** (1934), no. 12, 837–842. MR 1562984

[Mn87] Ricardo Mañé, *Ergodic theory and differentiable dynamics*, Ergebnisse der Mathematik und ihrer Grenzgebiete (3) [Results in Mathematics and Related Areas (3)], vol. 8, Springer-Verlag, Berlin, 1987, Translated from the Portuguese by Silvio Levy. MR 889254

[OP22] Davi Obata and Mauricio Poletti, *Positive exponents for random products of conservative surface diffeomorphisms and some skew products*, J. Dynam. Differential Equations **34** (2022), no. 3, 2405–2428. MR 4482258

[Pal00] Jacob Palis, *A global view of dynamics and a conjecture on the denseness of finitude of attractors*, Géométrie complexe et systèmes dynamiques (Orsay, 1995), no. 261, SMF, 2000, pp. xiii–xiv, 335–347. MR 1755446

[Pot22] Rafael Potrie, *A remark on uniform expansion*, Rev. Un. Mat. Argentina **64** (2022), no. 1, 11–21. MR 4477288

[PP90] William Parry and Mark Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque (1990), no. 187-188, 268. MR 1085356

[Roh64] V.A. Rohlin, *Exact endomorphisms of a Lebesgue space*, 15 papers on topology and logic, American Mathematical Society Translations. Series 2, vol. 39, American Mathematical Society, Providence, RI, 1964, pp. 1–36.

[Roh67] V. A. Rohlin, *Lectures on the entropy theory of transformations with invariant measure*, Uspehi Mat. Nauk **22** (1967), no. 5(137), 3–56. MR 217258

[Rue78] David Ruelle, *Thermodynamic formalism*, Encyclopedia of Mathematics and its Applications, vol. 5, Addison-Wesley Publishing Co., Reading, MA, 1978, The mathematical structures of classical equilibrium statistical mechanics, With a foreword by Giovanni Gallavotti and Gian-Carlo Rota. MR 511655

[Shu87] Michael Shub, *Global stability of dynamical systems*, Springer-Verlag, New York, 1987, With the collaboration of Albert Fathi and Rémi Langevin, Translated from the French by Joseph Christy. MR 869255

[Shu06] ———, *All, most, some differentiable dynamical systems*, International Congress of Mathematicians. Vol. III, Eur. Math. Soc., Zürich, 2006, pp. 99–120. MR 2275672

[Sin72] Ja. G. Sinaĭ, *Gibbs measures in ergodic theory*, Uspehi Mat. Nauk **27** (1972), no. 4(166), 21–64. MR 399421

[Ste91] Gilbert W. Stewart, *Perturbation theory for the singular value decomposition*, SVD and signal processing, II: algorithms, analysis, and applications (Richard J. Vaccaro, ed.), Elsevier, Amsterdam, 1991.

[Ste97] J. Michael Steele, *Probability theory and combinatorial optimization*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 69, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997. MR 1422018

[Tsu01] Masato Tsujii, *Fat solenoidal attractors*, Nonlinearity **14** (2001), no. 5, 1011–1027. MR 1862809

[TZ23] Masato Tsujii and Zhiyuan Zhang, *Smooth mixing Anosov flows in dimension three are exponentially mixing*, Ann. of Math. (2) **197** (2023), no. 1, 65–158. MR 4513143

[Via98] Marcelo Viana, *Dynamics: a probabilistic and geometric perspective*, Proceedings of the International Congress of Mathematicians, Vol. I (Berlin, 1998), 1998, pp. 557–578. MR 1648047

[Via99] ———, *Lecture notes on attractors and physical measures*, Monografías del Instituto de Matemática y Ciencias Afines [Monographs of the Institute of Mathematics and Related Sciences], vol. 8, Instituto de Matemática y Ciencias Afines, IMCA, Lima, 1999, A paper from the 12th Escuela Latinoamericana de Matemáticas (XII-ELAM) held in Lima, June 28–July 3, 1999. MR 2007887

[Via08] ———, *Almost all cocycles over any hyperbolic system have nonvanishing Lyapunov exponents*, Ann. of Math. (2) **167** (2008), no. 2, 643–680. MR 2415384

[Via14] ———, *Lectures on Lyapunov exponents*, Cambridge Studies in Advanced Mathematics, vol. 145, Cambridge University Press, Cambridge, 2014. MR 3289050

[WY01] Qiudong Wang and Lai-Sang Young, *Strange attractors with one direction of instability*, Comm. Math. Phys. **218** (2001), no. 1, 1–97. MR 1824198

[You98] Lai-Sang Young, *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. (2) **147** (1998), no. 3, 585–650. MR 1637655

[You99] ———, *Recurrence times and rates of mixing*, Israel J. Math. **110** (1999), 153–188. MR 1750438

Department of Mathematics, The University of Maryland, College Park, MD 20742, USA
Email address: dewitt@umd.edu, dolgop@umd.edu