NON-STATIONARY COMPOSITIONS OF ANOSOV DIFFEOMORPHISMS

MIKKO STENLUND

Abstract. Motivated by non-equilibrium phenomena in nature, we study dynamical systems whose time-evolution is determined by non-stationary compositions of chaotic maps. The constituent maps are topologically transitive Anosov diffeomorphisms on a 2-dimensional compact Riemannian manifold, which are allowed to change with time — slowly, but in a rather arbitrary fashion. In particular, such systems admit no invariant measure. By constructing a coupling, we prove that any two sufficiently regular distributions of the initial state converge exponentially with time. Thus, a system of the kind loses memory of its statistical history rapidly.

Acknowledgements. The author has received financial support from the Academy of Finland and the Väisälä Fund. He wishes to thank Lai-Sang Young for many discussions.

1. Introduction

1.1. **Motivation.** Statistical properties of dynamical systems are traditionally studied in a stationary context. Let us elaborate briefly, discussing only discrete time for simplicity. Suppose $\mathcal{M}$ is the set of all possible states of the system. Given the state $x_n \in \mathcal{M}$ at some time $n \geq 0$, the state of the system at time $n+1$ is assumed to be either (i) $x_{n+1} = Tx_n$, where $T : \mathcal{M} \to \mathcal{M}$ is an *a priori* specified map used at every time step or (ii) $x_{n+1} = T_{n+1}x_n$, where the maps $T_i : \mathcal{M} \to \mathcal{M}$ are drawn randomly and independently of each other and of $x_0, \ldots, x_{i-1}$ from a set of maps $\mathfrak{T}$ according to a distribution η. Now, suppose the initial point x_0 has random distribution μ: $\mathrm{Prob}(x_0 \in E) = \mu(E)$, for all measurable sets $E \subset \mathcal{M}$. By stationarity we mean that $\mathrm{Prob}(x_n \in E) = \mu(E)$, for *all* $n \geq 0$, for all measurable sets $E \subset \mathcal{M}$. This condition translates to $\mu(T^{-1}E) = \mu(E)$ in case (i) and to $\int_{\mathfrak{T}} \mu(T^{-1}E)\, d\eta(T) = \mu(E)$ in case (ii). In either case the measure μ is called invariant. The reader may verify that these definitions result, indeed, in a *strictly* stationary process in that all finite dimensional distributions are shift-invariant: $\mathrm{Prob}(x_{k_1+n} \in E_1, \ldots, x_{k_m+n} \in E_m) = \mathrm{Prob}(x_{k_1} \in E_1, \ldots, x_{k_m} \in E_m)$, for all choices of the indices and of the measurable sets. Let f be a measurable function on $\mathcal{M}$, which represents a quantity whose observed values $f(x_n)$ at different times one is interested in. Given an invariant measure one may, for example, study the statistical behavior of the sums $\sum_{i=0}^{n-1} f(x_i)$ of observations, making use of the fact that $(f(x_n))_{n\geq 0}$ is a stationary sequence of random variables.

A key ingredient in obtaining advanced statistical results on interesting systems is chaos, that is to say the dynamical complexity due to sensitive dependence of the trajectories $(x_n)_{n\geq 0}$ on the initial point x_0. In this paper we initiate a program to free ourselves from the standard constraint of stationarity, advocating the following view:

Much of the statistical theory of stationary dynamical systems can be carried over to sufficiently chaotic non-stationary systems.

2000 *Mathematics Subject Classification.* 37D20; 60F05, 82C05.

Key words and phrases. Hyperbolic dynamical systems, non-stationary compositions, non-equilibrium.

The deliberately imprecise statement above is proposed as a guideline and challenge instead of a theorem. We believe that a result obtained for a strongly chaotic stationary system quite generically has a non-stationary counterpart if the corresponding non-stationary system continues to be sufficiently chaotic.

The inspiration for undertaking the program stems from non-equilibrium processes in nature where it is often unfounded or simply false to assume that an observed system is driven by stationary forces. For example, it is conceivable that an ambient system governed in principle by measure preserving dynamics is, for all practical time scales, in a non-equilibrium state, so that the subsystem actually being observed is better modeled separately in terms of non-stationary dynamical rules. The remark is by no means limited to situations of physical interest alone, but seems to lend itself rather universally to applied sciences. Second, from a purely theoretical point of view it appears very restrictive to focus only on stationary dynamical models.

In order to advance the program in a meaningful way, we need a concrete model to work with. Deferring technical definitions till later, let the state $x_n \in \mathcal{M}$ of the system at time n be determined by the action of the composition $T_n \circ \cdots \circ T_1$ on the initial state $x_0 \in \mathcal{M}$, where each map $T_i : \mathcal{M} \to \mathcal{M}$ describes the dynamical rules at time i. For us, the constituent maps T_i are topologically transitive Anosov diffeomorphisms on a 2-dimensional compact Riemannian manifold $\mathcal{M}$, which form a prime class of nontrivial chaotic maps. It is clear that some additional control is needed; for instance, an alternating sequence of an Anosov diffeomorphism T and its inverse T^{-1} would yield $T_n \circ \cdots \circ T_1 = \mathrm{id}$ for even values of n, which does not result in chaotic dynamics. To that end, the maps T_i are here assumed to evolve slowly with time i, but otherwise they may do so in a rather arbitrary fashion. We point out that the maps T_i need not be randomly picked, there is no assumption of stationarity, and for large n the map T_n may be far from the map T_1. Even if all the maps T_i preserved the same initial measure, so that the random variables x_n were identically distributed, the process $(x_n)_{n\geq 0}$ would typically fail to be stationary.

We call such compositions *non-stationary* and think of them as descriptions of *dynamical systems out of equilibrium*.

We prove in this paper that the system at issue loses memory of its initial state exponentially. More accurately, assume x_0 has either distribution μ^1 or μ^2 and call μ^1_n and μ^2_n, respectively, the corresponding distributions of x_n. Our main result states that if μ^1 and μ^2 are sufficiently regular, then the difference $\int f\,d\mu^1_n - \int f\,d\mu^2_n$ tends to zero at an exponential rate with increasing n, provided f is a suitable test function. This type of weak convergence is natural due to the invertibility of the dynamics: the supports of the measures μ^1_n and μ^2_n will never overlap unless they did so initially. Instead, they tend to concentrate increasingly on unstable manifolds due to the contracting direction of the maps T_i and then wind wildly around the phase space $\mathcal{M}$ due to the expansion on unstable manifolds. Hence, one cannot hope to identify ever-increasing portions of μ^1_n and μ^2_n unless one first integrates against a test function that possesses some regularity along stable manifolds. In spite of the convergence of the difference $\mu^1_n - \mu^2_n$ for arbitrary initial measures μ^1 and μ^2, in general the limit measures $\lim_{n\to\infty} \mu^i_n$ do not exist individually even in the weak sense. It is more appropriate to think that all regular measures are attracted by a moving target in the space of measures.

Finally, let us point out that in the real world, where observations take place on finite time scales, one is not interested in the excessively distant future. To underline this, the results here are finite-time results, in which the sequence $T_1, \ldots, T_n$ is assumed to be known only up to some finite value of n. The lack of infinite future leads to certain technical problems to be discussed and dealt with below.

In [12] analogous results were obtained for uniformly expanding and piecewise expanding maps. The situation of the present paper is markedly more complicated because our Anosov diffeomorphisms have a contracting direction. Some steps in this direction were taken in [2], where mixing for certain arbitrarily ordered compositions of finitely many toral automorphisms was established. There are other studies which contain at least some elements that in spirit are not very far from our setting. In [3,4] compositions of hyperbolic maps — all close to each other — were studied and limit theorems proved. An abstract operator theoretic approach for obtaining limit theorems was described in [9], with applications to piecewise expanding interval maps. Moreover, symbolic dynamics of non-stationary subshifts of finite type was considered in [1]. An extensive literature on random compositions of maps exists. It will not be reviewed here, as the present paper concerns quite a different type of questions. Nevertheless, some of the techniques developed below should be useful in the context of random maps as well.

1.2. **Structure of the paper.** In Section 1.3 we describe the precise setting of the paper. In particular, we explain what kind of compositions of maps we are interested in and discuss our standing assumptions. After that, the main result of the paper, Theorem 2, is formulated. Section 1.4 introduces some basic concepts needed throughout the paper. The Introduction ends with Section 1.5, which discusses what the author perceives as the most important contributions of the paper, including a technical version of Theorem 2.

In Section 2 we define finite-time stable and unstable distributions and stable foliations needed to keep track of the dynamics with appropriate accuracy. We also prove quantitative results concerning the distortion effects of the dynamics. Subsequently, we are able to define in a meaningful way finite-time holonomy maps which satisfy useful bounds.

In Section 3 we formulate the central result of the paper — the Coupling Lemma. It is then used to prove Theorem 4, which subsequently implies Theorem 2. The Coupling Lemma itself is proved in Section 4, which is the most technical part of the paper.

To maintain the flow of the discussion, some key technical facts have been separated from the main text and presented in the appendices. They are cited in the text as needed. Appendix B is of special interest; there we prove the uniform Hölder regularity of the finite-time stable and unstable distributions introduced in Section 2.

1.3. **Compositions of Anosov diffeomorphisms.** Fix $Q \in \mathbb{N}$. For each $1 \le q \le Q$, let $\widetilde{T}_q : \mathcal{M} \to \mathcal{M}$ be a topologically transitive $\mathcal{C}^2$ Anosov diffeomorphism on the 2-dimensional compact Riemannian manifold $\mathcal{M}$ with metric d embedded in an ambient space $\mathbb{R}^M$ [1]. The Riemannian volume is denoted by m. The map $\widetilde{T}_q$ admits an invariant Sinai–Ruelle–Bowen (SRB) measure, μ_q, which is mixing; see for example [5]. In general, such a measure is not absolutely continuous with respect to the Riemannian volume. Let $\mathcal{U}_q = \mathbb{D}(\widetilde{T}_q, \varepsilon_q)$ be disk neighborhoods of small radii $\varepsilon_q > 0$ in the $\mathcal{C}^2$ topology. Now, pick a finite sequence (T_n) of Anosov diffeomorphisms such that

$$T_n \in \mathcal{U}_q \qquad \forall\, n \in I_q = (n_{q-1}, n_q], \tag{1}$$

where $0 = n_0 < n_1 < \cdots \le n_Q$. For technical reasons, also set $T_n = \widetilde{T}_Q$ for all $n > n_Q$. We assume that the intervals I_q are long enough:

$$|I_q| = n_q - n_{q-1} \ge N_q, \tag{2}$$

[1]Such a diffeomorphism is topologically conjugate to an automorphism of the torus.

where the numbers N_q, $1 \le q \le Q$, will be assumed suitably large. We will be interested in the statistical properties of the compositions

$$\mathcal{T}_n = T_n \circ \cdots \circ T_1 \qquad n \le n_Q. \tag{3}$$

The maps $\widetilde{T}_q$ serve as successive guiding points which the sequence (T_n) follows in the space of Anosov diffeomorphisms, spending a sufficiently long time N_q in each neighborhood $\mathcal{U}_q$ before moving on to $\mathcal{U}_{q+1}$. We also write $\mathcal{T}_{n,m} = T_n \circ \cdots \circ T_m$ for $m \le n$.

Each $\widetilde{T}_q$ admits a unique continuous invariant splitting of the tangent bundle: for each $x \in \mathcal{M}$, $T_x\mathcal{M} = E^u_{q,x} \oplus E^s_{q,x}$, where the 1-dimensional linear spaces $E^{u,s}_{q,x}$ depend continuously on the base point x, $D_x\widetilde{T}_q E^u_{q,x} = E^u_{q,\widetilde{T}_q x}$ and $D_x\widetilde{T}_q E^s_{q,x} = E^s_{q,\widetilde{T}_q x}$. In fact, in our 2-dimensional setting, the dependence on the base point is $\mathcal{C}^{1+\alpha}$ for some $\alpha > 0$, because the so-called bunching conditions [10] are satisfied. The families $E^u_q = \{E^u_{q,x}\}$ and $E^s_q = \{E^s_{q,x}\}$ are called the unstable and stable distributions of $\widetilde{T}_q$, respectively, and their integral curves are called unstable and stable manifolds of $\widetilde{T}_q$, respectively. By continuity, the angle between E^u_q and E^s_q at each point is uniformly bounded away from zero. The maps also have continuous families of unstable cones, $\{\mathcal{C}^u_{q,x}\}$, and stable cones, $\{\mathcal{C}^s_{q,x}\}$. These can be defined by setting

$$\mathcal{C}^u_{q,x} = \{v^u + v^s \,:\, v^u \in E^u_{q,x},\ v^s \in E^s_{q,x},\ \|v^s\| \le a_q\|v^u\|\},$$
$$\mathcal{C}^s_{q,x} = \{v^u + v^s \,:\, v^u \in E^u_{q,x},\ v^s \in E^s_{q,x},\ \|v^u\| \le a_q\|v^s\|\},$$

for some constants $a_q > 0$ such that

(C1) $D_x\widetilde{T}^n_q \mathcal{C}^u_{q,x} \subset \{0\} \cup \operatorname{int} \mathcal{C}^u_{q,\widetilde{T}^n_q x}$ and $D_x\widetilde{T}^{-n}_q \mathcal{C}^s_{q,x} \subset \{0\} \cup \operatorname{int} \mathcal{C}^s_{q,\widetilde{T}^{-n}_q x}$ if $n \ge p_q$,

(C2) $\|D_x\widetilde{T}^n_q v\| \ge \widetilde{C}_q\widetilde{\Lambda}^n_q\|v\|$ if $v \in \mathcal{C}^u_{q,x}$ and $\|D_x\widetilde{T}^{-n}_q v\| \ge \widetilde{C}_q\widetilde{\Lambda}^n_q\|v\|$ if $v \in \mathcal{C}^s_{q,x}$,

for constants $p_q \ge 1$, $0 < \widetilde{C}_q < 1$, and $\widetilde{\Lambda}_q > 1$.

We make the following standing assumptions:

(A0) $p_q = 1$ in condition (C1) above.
(A1) $D_xT\mathcal{C}^u_{q,x} \subset \{0\} \cup \operatorname{int} \mathcal{C}^u_{q,Tx}$ and $D_xT^{-1}\mathcal{C}^s_{q,x} \subset \{0\} \cup \operatorname{int} \mathcal{C}^s_{q,T^{-1}x}$ for all $T \in \mathcal{U}_q$.
(A2) There exist constants $0 < C_q < 1$ and $\Lambda_q > 1$ such that, if each $T_i \in \mathcal{U}_q$ for a fixed q, $\|D_x\mathcal{T}_n v\| \ge C_q\Lambda^n_q\|v\|$ if $v \in \mathcal{C}^u_{q,x}$ and $\|D_x\mathcal{T}^{-1}_n v\| \ge C_q\Lambda^n_q\|v\|$ if $v \in \mathcal{C}^s_{q,x}$.
(A3) $D_xT\mathcal{C}^u_{q,x} \subset \{0\} \cup \operatorname{int} \mathcal{C}^u_{q+1,Tx}$ if $T \in \mathcal{U}_{q+1}$ and $D_xT^{-1}\mathcal{C}^s_{q+1,x} \subset \{0\} \cup \operatorname{int} \mathcal{C}^s_{q,T^{-1}x}$ if $T \in \mathcal{U}_q$.
(A4) The numbers a_q can be assumed small.

Convention 1. *From now on we will assume that Q reference Anosov diffeomorphisms $\widetilde{T}_1, \dots, \widetilde{T}_Q$ have been fixed. When we say that a result does not depend on the choice of the sequence (T_i), we mean that the result holds true uniformly for all finite sequences $(T_i)_{i=1}^{n_Q}$ of any length n_Q, provided (1) and Assumptions (A) are satisfied and the numbers N_q appearing in (2) are large enough.*

Given $0 < \gamma < 1$, we say that a function $f : \mathcal{M} \to \mathbb{R}$ is a γ-Hölder continuous observable, if

$$|f|_\gamma \equiv \sup_{x \ne y} \frac{|f(x) - f(y)|}{d(x,y)^\gamma} < \infty.$$

We are now in position to state our main theorem, which is reminiscent of weak convergence of measures in probability theory.

Theorem 2 (Weak convergence)**.** *There exist constants $0 < \eta < 1$ and $C > 0$, for which the following statements hold. Let $d\mu^i = \rho^i dm$ $(i = 1, 2)$ be two probability measures, absolutely continuous with respect to the Riemannian volume m, such that ρ^i are strictly positive and η-Hölder continuous on $\mathcal{M}$. If f is continuous, then*

$$\left| \int_{\mathcal{M}} f \circ \mathcal{T}_n \, d\mu^1 - \int_{\mathcal{M}} f \circ \mathcal{T}_n \, d\mu^2 \right| \le A_f(n), \qquad n \le n_Q,$$

where $A_f(n) = o(1)$. Given $0 < \gamma < 1$, there exist constants $0 < \theta_\gamma < 1$ and $C_\gamma = C_\gamma(\rho^1, \rho^2) > 0$ such that, if f is γ-Hölder, then $A_f(n) = C_\gamma B_f \theta_\gamma^n$ with $B_f = C(\sup f - \inf f) + |f|_\gamma$. In either case, the various constants do not depend on the choice of the sequence (T_i), in particular its length n_Q, as long as the earlier assumptions hold and the numbers N_q appearing in (2) are large enough. Among the constants only C_γ depends on the densities ρ^i, and in fact it only depends on the Hölder constants of $\ln \rho^i$.

In other words, if f is continuous, the difference between the two integrals $\int_{\mathcal{M}} f \circ \mathcal{T}_n \, d\mu^i$ is eventually arbitrarily small, assuming there are sufficiently many maps in the finite sequence (T_i) of n_Q maps. The latter means that at least one of the intervals I_q in (2) is sufficiently long, and consequently n_Q is large. What is more, the rate of convergence is exponential, if f is Hölder continuous. By approximation, one can get an $o(1)$ estimate also for general continuous densities ρ^i.

Let us emphasize once more that despite such convergence or pairs of measures, it does not make any sense to speak of a limit measure, because the maps T_n keep evolving with time — possibly drifting very far from T_1. Furthermore, all observations in our theorems are restricted to times not exceeding (the arbitrarily large but finite) n_Q.

Theorem 2 remains true for much more general, SRB-like, initial measures. It is enough that each measure μ^i can be disintegrated relative to a measurable partition $\mathcal{P}^i$ such that the partition elements $W \in \mathcal{P}^i$ are smooth unstable curves with respect to the cones $\{\mathcal{C}^u_{1,x}\}$ with uniformly bounded curvatures and the conditional measures $\mu^i|W$ have regular densities. See below for details.

In our formulation of the theorem, the convergence rate θ_γ is constant. The latter depends on the reference diffeomorphisms $\widetilde{T}_q$, $1 \le q \le Q$. A sharper, variable, convergence rate that depends also on the time interval I_q that n belongs to, can be deduced from the proof.

We finish the section by discussing Assumptions (A) and how they could be relaxed.

Assumption (A0) is one of convenience; we could as well assume that $T_{p_q} \circ \cdots \circ T_1$ is sufficiently close to $\widetilde{T}^{p_q}$, but have opted for a streamlined presentation. Assumptions (A1) and (A2) state that compositions of maps belonging to $\mathcal{U}_q$ have similar hyperbolicity properties as powers of $\widetilde{T}_q$. The following lemma is proved after a few paragraphs:

Lemma 3. *Assumptions (A1) and (A2) are satisfied if ε_q is sufficiently small.*

Assumption (A3) guarantees that hyperbolicity prevails when a transition from $\mathcal{U}_q$ to $\mathcal{U}_{q+1}$ occurs. The first part of (A3) could be relaxed by replacing the map $T \in \mathcal{U}_{q+1}$ by sufficiently long compositions $\mathcal{T}_n = T_n \circ \cdots \circ T_1$ of maps with each $T_i \in \mathcal{U}_{q+1}$: given a sufficiently large $r_q > 0$, $D_x \mathcal{T}_n \mathcal{C}^u_{q,x} \subset \{0\} \cup \operatorname{int} \mathcal{C}^u_{q+1,\mathcal{T}_n x}$ if $n \ge r_q$ and $T_i \in \mathcal{U}_{q+1}$ for $1 \le i \le n$. The assumption is then satisfied, for example, if $\mathcal{U}_q \cap \mathcal{U}_{q+1} \ne \varnothing$ and if ε_q is small, for all q. However, $\mathcal{U}_q$ and $\mathcal{U}_{q+1}$ need not overlap or even be close to each other, as most vectors in the tangent space $T_x\mathcal{M}$ get eventually mapped by $\mathcal{T}_n = T_n \circ \cdots \circ T_1$ into $\mathcal{C}^u_{q+1,\mathcal{T}_n x}$ if each $T_i \in \mathcal{U}_{q+1}$ and if ε_{q+1} is small.

Similar remarks hold for the second part of (A3). This way, the sequence (T_n) used to build up the compositions (3) might, without affecting our analysis, involve occasional long jumps from one neighborhood $\mathcal{U}_q$ to the next, as long as the number of steps $|I_q|$ spent in each neighborhood $\mathcal{U}_q$, see (2), is sufficiently large.

Assumption (A4) means that the cones can be assumed narrow. This is not restrictive for our purposes either, as it follows from (C2) that arbitrarily narrow cones can be treated by considering sufficiently long compositions of maps in a given $\mathcal{U}_q$ with a sufficiently small ε_q.

Recapitulating, it would be adequate to assume that the properties above hold eventually, for sufficiently long compositions of maps, and in this case the assumptions are very natural and easily fulfilled. For technical convenience and notational ease, we assume from now on that all the nice properties hold immediately, after the application of just one map.

Proof of Lemma 3. We will prove the claims for unstable cones, going forward in time. Similar arguments work for the stable cones, by reversing time.

(A1): Suppose $T \in \mathcal{U}_q$. By the chain rule $D_xT\mathcal{C}^u_{q,x} = D_{\widetilde{T}_q x}(T\widetilde{T}_q^{-1})D_x\widetilde{T}_q\,\mathcal{C}^u_{q,x}$, where $D_x\widetilde{T}_q\mathcal{C}^u_{q,x} \subset \{0\}\cup \operatorname{int}\mathcal{C}^u_{q,\widetilde{T}_q x}$. By the continuity of the cones with respect to the base point and the fact that $D_{\widetilde{T}_q x}(T\widetilde{T}_q^{-1}) = \mathbb{1} + \mathcal{O}(\varepsilon_q)$ [2], we have $D_xT\mathcal{C}^u_{q,x} \subset \{0\}\cup \operatorname{int}\mathcal{C}^u_{q,Tx}$, provided ε_q is sufficiently small. Compactness guarantees that ε_q can be chosen independently of x.

(A2): For each i and x, $D_xT_i = D_x\widetilde{T}_q + \mathcal{E}_{i,x}$, where $\sup_{i,x}\|\mathcal{E}_{i,x}\| = \mathcal{O}(\varepsilon_q)$. We can bound $\|D_x\mathcal{T}_N - D_x\widetilde{T}_q^N\| \le C(N)\varepsilon_q$. If ε_q is sufficiently small, we have $\|D_x\mathcal{T}_N v\| \ge \|D_x\widetilde{T}_q^N v\| - C(N)\varepsilon_q\|v\| \ge \frac{1}{2}\widetilde{C}_q\widetilde{\Lambda}_q^N\|v\|$ for $v \in \mathcal{C}^u_{q,x}$. Now assume $N = N(q)$ is so large that $\frac{1}{2}\widetilde{C}_q\widetilde{\Lambda}_q^N > 1$. Here N depends neither on x, on v, nor on the choice of the maps $T_i \in \mathcal{U}_q$. Let us set $\Lambda_q = (\frac{1}{2}\widetilde{C}_q\widetilde{\Lambda}_q^N)^{1/N}$. The uniform estimate $\|D_x\mathcal{T}_n v\| \ge c_q\|v\|$ holds with some $c_q = c_q(N)$ for $1 \le n < N$. Now, assume $n = kN + l$, $0 \le l < N$. Then $\|D_x\mathcal{T}_n v\| \ge c_q\|D_x\mathcal{T}_{kN} v\| \ge c_q\Lambda_q^{kN}\|v\|$. Thus, we can take $C_q = c_q/\Lambda_q^N$. □

1.4. **Unstable curves with smooth measures.** We call a smooth curve $W \subset \mathcal{M}$ unstable with respect to $\{\mathcal{C}^u_{q,x}\}$ if its tangent space at each point $x \in W$ is contained in the unstable cone $\mathcal{C}^u_{q,x}$, *i.e.*, $T_xW \subset \mathcal{C}^u_{q,x}$. Stable curves are defined similarly. Let $W(x,y) \subset W$ denote the subcurve of W whose end points are $x, y \in W$. The length $|W|$ of a curve W is given by

$$|W| = \int_W dm_W,$$

where m_W stands for the measure m_W on W induced by the Riemannian metric. Also, let $\kappa(W)$ stand for the maximum curvature of W: if $u(x)$ is a unit tangent vector of W at x depending smoothly on x, then $\kappa(W) = \sup\|u \cdot \nabla u\|$.

It is convenient to consider curves of bounded length and curvature only. Hence, we introduce two length caps, L and $\ell < L$, and a curvature cap K, and say that a smooth curve W is *standard*, if $\ell \le |W| \le L$ and if $\kappa(\mathcal{T}_n W) \le K$ for all $n \ge 0$. If $|W| > L$, we can always "standardize" it by cutting it into shorter subcurves. If an unstable curve is of length less than ℓ, it will eventually grow under the application of the sequence (T_i), such that $|\mathcal{T}_n W| \ge \ell$ for sufficiently large n. The dynamics also flattens unstable curves, such that $\kappa(\mathcal{T}_n W) \le K$ for all sufficiently large n, even if $\kappa(W) > K$. We will confirm these last two facts in the following. Finally, there

[2]Here it is understood that $\mathcal{M}$ is embedded in the ambient space $\mathbb{R}^M$ and that $D_{\widetilde{T}_q x}(T\widetilde{T}_q^{-1})$ acts between the linear subspaces $T_{\widetilde{T}_q x}\mathcal{M}$ and $T_{Tx}\mathcal{M}$ of $\mathbb{R}^M$.

are no discontinuities which would introduce more short curves under the dynamics by cutting longer ones. Taking these considerations into account it is quite natural to commit to the mild constraint that all curves are standard curves to begin with. This will help keep the somewhat technical discussion as clear as possible.

A *standard pair* (W,ν) (w.r.t. $\{\mathcal{C}^u_{q,x}\}$) consist of a standard unstable curve (w.r.t. $\{\mathcal{C}^u_{q,x}\}$), W, and a probability measure, ν, on W. The measure ν is assumed absolutely continuous with respect to m_W on W with a density, ρ, that is regular in the following sense: for some global constants $C_{\mathrm{r}} > 0$ and $\eta_{\mathrm{r}} \in (0,1]$ to be fixed later [3],

$$|\ln \rho(x) - \ln \rho(y)| \le C_{\mathrm{r}} |W(x,y)|^{\eta_{\mathrm{r}}} \tag{4}$$

for all $x, y \in W$. In particular,

$$\frac{\sup \rho}{\inf \rho} \le e^{C_{\mathrm{r}}|W|^{\eta_{\mathrm{r}}}},$$

which implies $\inf \rho \ge \frac{1}{|W|} e^{-C_{\mathrm{r}} L^{\eta_{\mathrm{r}}}} \ge \frac{1}{L} e^{-C_{\mathrm{r}} L^{\eta_{\mathrm{r}}}} > 0$, since $\int_W \rho \, dm_W = 1$. Moreover, $\inf \rho \le \frac{1}{|W|}$. If $W' \subset W$, we obtain by using the previous facts that

$$e^{-C_{\mathrm{r}} L^{\eta_{\mathrm{r}}}} \le \frac{|W|}{|W'|} \nu(W') \le e^{C_{\mathrm{r}} L^{\eta_{\mathrm{r}}}}.$$

Hence, if $D = e^{2C_{\mathrm{r}} L^{\eta_{\mathrm{r}}}}$ and $W', W'' \subset W$,

$$D^{-1} \frac{\nu(W'')}{|W''|} \le \frac{\nu(W')}{|W'|} \le D \, \frac{\nu(W'')}{|W''|}. \tag{5}$$

Formally, a standard family is a family $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}_{\alpha \in \mathfrak{A}}$ of standard pairs together with a probability factor measure $\lambda_{\mathcal{G}}$ on the (possibly uncountable) index set $\mathfrak{A}$ and a probability measure $\mu_{\mathcal{G}}$ satisfying

$$\mu_{\mathcal{G}}(B) = \int_{\mathfrak{A}} \nu_\alpha(B \cap W_\alpha) \, d\lambda_{\mathcal{G}}(\alpha)$$

for each Borel measurable set $B \subset \mathcal{M}$. The measure $\mu_{\mathcal{G}}$ is supported on $\cup_\alpha W_\alpha$ and

$$\mathbb{E}_{\mathcal{G}}(f) = \int_{\mathcal{M}} f \, d\mu_{\mathcal{G}} = \int_{\mathfrak{A}} \int_{W_\alpha} f(x) \, d\nu_\alpha(x) d\lambda_{\mathcal{G}}(\alpha)$$

for each Borel measurable function f on $\mathcal{M}$. In Theorem 4 we assume that a standard family is associated to a measurable partition.

A standard family can, for example, consist of just one standard pair $\{(W,\nu)\}$ and the Dirac point mass factor measure δ_W. Another natural example of a standard family $\{(W_\alpha,\nu_\alpha)\}_{\alpha \in \mathfrak{A}}$ is obtained by considering an Anosov diffeomorphism and taking as $\{W_\alpha\}_{\alpha \in \mathfrak{A}}$ a measurable partition consisting of unstable manifolds of bounded length and letting the Riemannian volume induce the factor measure and the conditional measures ν_α.

1.5. **Main contributions.** A technical version of our main result is the following theorem. It states that for reasonable initial distributions $\mu_{\mathcal{G}}$ and $\mu_{\mathcal{E}}$, the images $\mathcal{T}_n \mu_{\mathcal{G}}$ and $\mathcal{T}_n \mu_{\mathcal{E}}$ converge exponentially in a weak sense.

[3] C_{r} has to satisfy the condition in Lemma 9 and η_{r} is determined in Lemma 22. Both depend on the reference sequence $\widetilde{T}_1, \dots, \widetilde{T}_Q$, but not on the choice of (T_i).

Theorem 4. *There exist constants $C > 0$ and $0 < \vartheta < 1$, and $0 < \lambda < 1$, such that the following holds. For any standard families $\mathcal{G}$ and $\mathcal{E}$, any $\gamma > 0$, and any γ-Hölder observable f* [4],

$$\left| \int_{\mathcal{M}} f \circ \mathcal{T}_n \, d\mu_{\mathcal{G}} - \int_{\mathcal{M}} f \circ \mathcal{T}_n \, d\mu_{\mathcal{E}} \right| \le B_f \theta_\gamma^n, \qquad n \le n_Q,$$

where,

$$B_f = C(\sup f - \inf f) + |f|_\gamma \quad \text{and} \quad \theta_\gamma = \max(\vartheta, \lambda^\gamma)^{1/2}.$$

The various constants do not depend on the choice of the sequence (T_i), in particular its length n_Q, as long as the earlier assumptions hold and the numbers N_q appearing in (2) are large enough.

As a consequence of Theorem 4, we prove the earlier Theorem 2, which is stated in terms of less technical notions and is closer in spirit to the weak convergence of probability theory.

The proofs of Theorems 4 and 2 rely on a coupling method that has its roots in probability theory. It was carried over to the study of dynamical systems by Lai-Sang Young [13, 11] who used it to prove exponential decay of correlations for Sinai Billiards and uniqueness of invariant measures for randomly perturbed dissipative parabolic PDEs. Bressaud and Liverani [5] also used coupling to give explicit estimates on the decay of correlations for Anosov diffeomorphisms. The present paper takes advantage of a version of Young's coupling method introduced by Dmitry Dolgopyat and Nikolai Chernov [8, 7].

A considerable amount of work is devoted to obtaining *uniform bounds*, which is more involved than in the case of iterating a single map. A central issue is that the finite sequences $(T_1, \dots, T_{n_Q})$ of maps that we consider do not possess stable and unstable manifolds, because defining such objects requires an infinite future and an infinite past, respectively. Thus, we have to resort to artificial, finite-time, foliations that describe the dynamics sufficiently faithfully but are by no means unique. Moreover, in the single map case the regularity properties of the foliations of the manifold into stable and unstable manifolds play an important role. Our construction should therefore also yield regular foliations. In addition, the amount of regularity must not depend on the choice of the sequence $(T_1, \dots, T_{n_Q})$ (as long as $Q \ge 1$ and the maps $\widetilde{T}_q$, $1 \le q \le Q$, have been fixed and the earlier assumptions are satisfied), since the goal is to prove the uniform convergence result in Theorem 4.

At the heart of Dolgopyat's and Chernov's method lies the Coupling Lemma (corresponding to Lemma 13). In its proof, one constructs a special reference set called the magnet. By mixing, any standard pair will ultimately cross the magnet as if it was attracted by the latter. Once two standard pairs cross the magnet, parts of them can be coupled to each other using the stable foliation. In this paper, we generalize the idea by considering time-dependent magnets and time-dependent, finite-time, foliations for the coupling construction.

2. Distortions and holonomy maps

2.1. **Stable foliations $\mathcal{W}^n$.** As pointed out above, there is no well-defined sequence of stable foliations associated to the finite sequence $(T_1, \dots, T_{n_Q})$ of maps. A way around this is to try to augment the sequence with a fake future consisting of infinitely many maps — in our case $T_n = \widetilde{T}_Q$ for $n > n_Q$ — and to consider the uniquely defined stable foliations of the resulting infinite sequence of maps. This sequence of stable foliations naturally depends on the chosen

[4]Given a sequence (T_i), it is in fact enough to assume that f is Hölder continuous along the finite-time stable leaves associated to that particular sequence; see Section 2.

future and it is not *a priori* clear whether they have very much to do with the finite-time dynamics ($1 \le n \le n_Q$) which is the only thing we are interested in.

For a sequence (T_i) satisfying the earlier assumptions, we can define a sequence of stable distributions on the manifold $\mathcal{M}$, by pulling back the stable distribution $E^s_{Q,x}$ of $\widetilde{T}_Q$. More precisely, let us first define $E^n_x = E^s_{Q,x}$ for $n \ge n_Q + 1$ and then

$$E^n_x = D_{T_{n+1}x} T^{-1}_{n+1} E^{n+1}_{T_{n+1}x}, \qquad 0 \le n \le n_Q.$$

With this definition,

$$D_x \mathcal{T}_{n,m} E^{m-1}_x = E^n_{\mathcal{T}_{n,m}x}, \qquad n \ge m \ge 1.$$

Assumptions (A1) and (A3) guarantee that $E^n_x \subset \{0\} \cup \operatorname{int} \mathcal{C}^s_{q,x}$ for $n+1 \in I_q$ and $1 \le q \le Q$. By Assumption (A4), the angle between E^n_x and $E^s_{q,x}$ can be assumed uniformly small, for $n \in I_q$ and $1 \le q \le Q$.

The distributions E^n above are the tangent distributions to the stable foliations $\mathcal{W}^n$ of the sequences $(T_i)_{i>n}$. If $\mathcal{W}^s_{Q,x}$ is the stable leaf of $\widetilde{T}_Q$ at x, then $\mathcal{W}^n_x = \mathcal{W}^s_{Q,x}$ for $n \ge n_Q + 1$ and

$$\mathcal{W}^n_x = T^{-1}_{n+1} \mathcal{W}^{n+1}_{T_{n+1}x}, \qquad 0 \le n \le n_Q.$$

Notice that $y \in \mathcal{W}^n_x$ if and only if $\lim_{N\to\infty} d(\mathcal{T}_{N,n+1}x, \mathcal{T}_{N,n+1}y) = 0$.

For technical reasons, we also define $F^n_x = E^u_{1,x}$ for $n \le 0$ and then

$$F^n_x = D_{T_n^{-1}x} T_n F^{n-1}_{T_n^{-1}x}, \qquad 1 \le n \le n_Q. \tag{6}$$

Assumptions (A1) and (A3) guarantee that $F^n_x \subset \{0\} \cup \operatorname{int} \mathcal{C}^u_{q,x}$ for $n \in I_q$ and $1 \le q \le Q$. By Assumption (A4), the angle between F^n_x and $E^u_{q,x}$ can be assumed uniformly small, for $n \in I_q$ and $1 \le q \le Q$. The distributions F^n are in fact the unstable distributions of the sequence (T_i) augmented with the past $T_i = \widetilde{T}_1$ for $i \le 0$. They serve as *Hölder continuous* reference distributions that allow us to accurately compare different unstable vectors.

In Appendix B we show that the distributions F^n and E^n for all n are uniformly Hölder continuous.

2.2. **Distortion.** It is necessary to control the distortion and growth of curves under maps T. Given a curve W we denote by $\mathcal{J}_W T$ the Jacobian of the restriction of T to W. If v is any nonzero tangent vector of W at x, then

$$\mathcal{J}_W T(x) = \frac{\|D_x T v\|}{\|v\|}.$$

Lemma 5 (Growth of unstable curves)**.** *Fix $1 \le q \le Q$ and let $T_i \in \mathcal{U}_q$ for each i. If W is an unstable curve with respect to $\{\mathcal{C}^u_{q,x}\}$, setting $\bar{\Lambda}_q = \sup_x \|D_x \widetilde{T}_q\| + \varepsilon_q$,*

$$C_q \Lambda^n_q |W| \le |\mathcal{T}_n W| \le \bar{\Lambda}^n_q |W|. \tag{7}$$

If $W, \mathcal{T}_1 W, \dots, \mathcal{T}_n W$ are stable curves with respect to $\{\mathcal{C}^s_{q,x}\}$, then

$$|\mathcal{T}_n W| \le \frac{|W|}{C_q \Lambda^n_q}. \tag{8}$$

Proof. Since $|\mathcal{T}_n W| = \int_{\mathcal{T}_n W} dm_{\mathcal{T}_n W} = \int_W \mathcal{J}_W \mathcal{T}_n \, dm_W = \int_W \|D_x \mathcal{T}_n v_x\| \, dm_W(x)$, where v_x is a unit vector tangent to W at x, it suffices to observe that $C_q \Lambda^n_q \le \|D_x \mathcal{T}_n v_x\| \le \bar{\Lambda}^n_q$ in the "unstable case" and $\|D_x \mathcal{T}_n v_x\| \le \frac{1}{C_q \Lambda^n_q}$ in the "stable case". □

Lemma 6 (Curvature of unstable curves)**.** *Fix $1 \le q \le Q$ and let $T_i \in \mathcal{U}_q$ for each i. There exist K_1 and, for any $K' > 0$, $K_2(K')$ and $n_\kappa(K')$, such that*

$$\kappa(\mathcal{T}_n W) \le \begin{cases} K_1, & n \ge n_\kappa, \\ K_2, & n \ge 0, \end{cases}$$

holds if W is an unstable curve with respect to $\{\mathcal{C}^u_{q,x}\}$ and $\kappa(W) \le K'$. Notice that $K_1 \le K_2$ is independent of K'.

Remark 7. *We can now fix some K' and set $K = K_2(K')$ in the definition of standard pairs. In particular, this means that any unstable curve W with length between ℓ and L and curvature $\kappa(W) \le K'$ is a standard curve.*

Proof of Lemma 6. Let W be an unstable curve and γ its parametrization by arc length, such that $u(x) = \dot\gamma(t) \in \mathcal{C}^u_{q,x}$ with $x = \gamma(t)$. Note $\|u\| = 1$. The curvature of W at x is the length of

$$\ddot\gamma(t) = u(x) \cdot \nabla u(x).$$

Setting $V = D\mathcal{T}_n u$, $v(y) = V(x)/\|V(x)\|$ is the unit tangent of $\mathcal{T}_n W$ at $y = \mathcal{T}_n x$. The curvature of $\mathcal{T}_n W$ at y is thus obtained from

$$v(y) \cdot \nabla v(y) = Dv(y)v(y) = \|V(x)\|^{-1} Dv(y) D\mathcal{T}_n(x) u(x) = \|V(x)\|^{-1} D_x(v(y)) u(x).$$

Here the chain rule $D_x(v(y)) = Dv(y) D\mathcal{T}_n(x)$ was used. Now

$$\begin{aligned} D_x(v(y))u(x) &= D(\|V(x)\|^{-1}V(x))u(x) = \|V(x)\|^{-1}DV(x)u(x) + V(x)D(\|V(x)\|^{-1})u(x) \\ &= \|V(x)\|^{-1}DV(x)u(x) + V(x)\left(-\|V(x)\|^{-3}V(x)\cdot DV(x)\right)u(x) \\ &= \|V(x)\|^{-1}DV(x)u(x) - \|V(x)\|^{-1}v(y)\left(v(y)\cdot DV(x)u(x)\right), \end{aligned}$$

such that

$$v(y)\cdot\nabla v(y) = \|V(x)\|^{-2}\left(DV(x)u(x) - v(y)\left(v(y)\cdot DV(x)u(x)\right)\right),$$

or compactly

$$v\cdot\nabla v = \|V\|^{-2}\left(DVu - v\left(v\cdot DVu\right)\right). \tag{9}$$

Notice that $DVu - v\,(v\cdot DVu)$ is the component of DVu orthogonal to v and hence $\|DVu - v\,(v\cdot DVu)\| \le \|DVu\|$. Furthermore, as $Du\,u = u\cdot\nabla u$, which we recognize to be the curvature of W at x, we have

$$DVu = D^2\mathcal{T}_n(u,u) + D\mathcal{T}_n(u\cdot\nabla u). \tag{10}$$

Using Lemma 26 and $\|D\mathcal{T}_n u\| \ge C_q\Lambda_q^n\|u\|$, we see from (9) and (10) that

$$\begin{aligned} \|v\cdot\nabla v\| \le \|V\|^{-2}\|DVu\| &\le \frac{\|D^2\mathcal{T}_n(u,u)\|}{\|D\mathcal{T}_n u\|^2} + \frac{\|D\mathcal{T}_n(u\cdot\nabla u)\|}{\|D\mathcal{T}_n u\|^2} \\ &\le (C_q\Lambda_q^n)^{-2}\|D^2\mathcal{T}_n\|_\infty + (C_q\Lambda_q^n)^{-1}C_\#\|u\cdot\nabla u\|. \end{aligned} \tag{11}$$

Fix an N such that $(C_q\Lambda_q^N)^{-1}C_\# < 1$. Iterating (11),

$$\kappa(\mathcal{T}_{kN+l}W) \le \frac{(C_q\Lambda_q^N)^{-2}\sup_{(T_i)}\|D^2\mathcal{T}_N\|_\infty}{1-(C_q\Lambda_q^N)^{-1}C_\#} + ((C_q\Lambda_q^N)^{-1}C_\#)^k\kappa(\mathcal{T}_l W).$$

A uniform bound $\max_{0\le l<N}\kappa(\mathcal{T}_l W) \le a + b\cdot\kappa(W)$ is also obtained, so we are done. ∎

If W carries a measure ν with density ρ, then $\mathcal{T}_n W$ carries the measure $\mathcal{T}_n\nu$ whose density, which we denote $\mathcal{T}_n\rho$, is

$$(\mathcal{T}_n\rho)(x) = \frac{\rho(\mathcal{T}_n^{-1}x)}{\mathcal{J}_W\mathcal{T}_n(\mathcal{T}_n^{-1}x)} = \mathcal{J}_{\mathcal{T}_nW}\mathcal{T}_n^{-1}(x)\cdot\rho(\mathcal{T}_n^{-1}x).$$

For controlling the regularity of such densities, we have the following result.

Lemma 8 (Distortion bound). *Fix $1\le q\le Q$ and let $T_i\in\mathcal{U}_q$ for each i. If $\mathcal{T}_n^{-1}W$ is a standard unstable curve with respect to $\{\mathcal{C}^u_{q,x}\}$ for all $0\le n\le N$ and if $x,y\in W$, then*

$$\left|\ln\frac{\mathcal{J}_W\mathcal{T}_n^{-1}(x)}{\mathcal{J}_W\mathcal{T}_n^{-1}(y)}\right| \le C_{\mathrm{d},q}|W(x,y)|\,, \qquad n\le N.$$

Here $C_{\mathrm{d},q}>0$ is a constant that do not depend on W or the choice of (T_i).

Proof. We first prove that the distortion factor of any map $T\in\mathcal{U}_q$ is close to that of $\widetilde{T}_q$. To this end, let W be an unstable curve with respect to $\{\mathcal{C}^u_{q,x}\}$, $x\in W$, and v a unit vector tangent to W at x. Then

$$\begin{aligned}\left|\mathcal{J}_WT^{-1}(x)-\mathcal{J}_W\widetilde{T}_q^{-1}(x)\right| &= \left|\|D_xT^{-1}v\|-\|D_x\widetilde{T}_q^{-1}v\|\right| \le \|(D_xT^{-1}-D_x\widetilde{T}_q^{-1})v\|\\ &= \|D_xT^{-1}(D_{\widetilde{T}_q^{-1}x}\widetilde{T}_q - D_{T^{-1}x}T)D_x\widetilde{T}_q^{-1}v\|\\ &\le C\|D_{\widetilde{T}_q^{-1}x}\widetilde{T}_q - D_{T^{-1}x}T\|\mathcal{J}_W\widetilde{T}_q^{-1}(x) \le C\varepsilon_q\,\mathcal{J}_W\widetilde{T}_q^{-1}(x).\end{aligned}$$

Next, let γ parametrize $W(x,y)$ according to arc length. Because

$$\begin{aligned}\left|\frac{d}{dt}\mathcal{J}_WT^{-1}(\gamma(t))\right| &= \left|\frac{d}{dt}\big\|D_{\gamma(t)}T^{-1}\dot\gamma(t)\big\|\right| = \left|\frac{D_{\gamma(t)}T^{-1}\dot\gamma(t)}{\big\|D_{\gamma(t)}T^{-1}\dot\gamma(t)\big\|}\cdot\frac{d}{dt}\big(D_{\gamma(t)}T^{-1}\dot\gamma(t)\big)\right|\\ &\le \left\|\frac{d}{dt}\big(D_{\gamma(t)}T^{-1}\dot\gamma(t)\big)\right\| = \big\|D^2_{\gamma(t)}T^{-1}(\dot\gamma(t),\dot\gamma(t)) + D_{\gamma(t)}T^{-1}\ddot\gamma(t)\big\|\\ &\le \sup_x\big\|D^2_xT^{-1}\big\| + \sup_x\big\|D_xT^{-1}\big\|\|\ddot\gamma(t)\|,\end{aligned}$$

and because the curvature $\|\ddot\gamma\|\le K$ for all standard unstable curves,

$$\begin{aligned}\big|\ln\mathcal{J}_WT^{-1}(x)-\ln\mathcal{J}_WT^{-1}(y)\big| &= \left|\int_0^{|W(x,y)|}\frac{d}{dt}\big(\ln\mathcal{J}_WT^{-1}(\gamma(t))\big)\,dt\right|\\ &= \left|\int_0^{|W(x,y)|}\frac{\frac{d}{dt}\mathcal{J}_WT^{-1}(\gamma(t))}{\mathcal{J}_WT^{-1}(\gamma(t))}\,dt\right| \le \widetilde{C}_{\mathrm{d},q}|W(x,y)|,\end{aligned}$$

where $\widetilde{C}_{\mathrm{d},q}$ is independent of the choice of T. The desired estimate follows. Indeed, writing $x^{-j}=(\mathcal{T}_{n,n-j+1})^{-1}x$, $y^{-j}=(\mathcal{T}_{n,n-j+1})^{-1}y$, and $W^{-j}=(\mathcal{T}_{n,n-j+1})^{-1}W$ (with $\mathcal{T}_{n,n+1}=\mathrm{id}$),

$$\left|\ln\frac{\mathcal{J}_W\mathcal{T}_n^{-1}(x)}{\mathcal{J}_W\mathcal{T}_n^{-1}(y)}\right| \le \sum_{j=0}^{n-1}\left|\ln\frac{\mathcal{J}_{W^{-j}}T_{n-j}^{-1}(x^{-j})}{\mathcal{J}_{W^{-j}}T_{n-j}^{-1}(y^{-j})}\right| \le \sum_{j=0}^{n-1}\widetilde{C}_{\mathrm{d},q}|W^{-j}(x^{-j},y^{-j})|.$$

Moreover, by (7), $|W^{-j}(x^{-j},y^{-j})|\le C_q^{-1}\Lambda_q^{-j}|W(x,y)|$. □

2.3. **Image of a standard family.**

Lemma 9. *Fix $1 \le q \le Q$ and let $T_i \in \mathcal{U}_q$ for each i. Let $\mathcal{G} = (W, \nu)$ be a standard pair with respect to $\{\mathcal{C}^u_{q,x}\}$ and assume that C_r satisfies*

$$2C_{\mathrm{d,q}}L^{1-\eta_\mathrm{r}} \le C_\mathrm{r}.$$

For $n \ge \ln \frac{2}{C_q} / \ln \Lambda_q$, denote by W_i the (finitely many) standard pieces of the image $\mathcal{T}_n W$ after it has been standardized by cutting into shorter pieces and split the image measure $\mathcal{T}_n \nu$ into the sum $\sum_i c_i \nu_i$, where ν_i is a probability measure on W_i and $\sum_i c_i = 1$. Then each (W_i, ν_i) is a standard pair w.r.t. $\{\mathcal{C}^u_{q,x}\}$.

Proof. We only need to check that the density, ρ_i, of ν_i is regular. For $x \in W_i$, $\rho_i(x) = \mathcal{J}_{W_i}\mathcal{T}_n^{-1}(x) \cdot \rho(\mathcal{T}_n^{-1}x)/c_i$. Thus, for any pair $x, y \in W_i$,

$$\begin{aligned}
|\ln \rho_i(x) - \ln \rho_i(y)| &\le |\ln \rho(\mathcal{T}_n^{-1}x) - \ln \rho(\mathcal{T}_n^{-1}y)| + \left| \ln \frac{\mathcal{J}_{W_i}\mathcal{T}_n^{-1}(x)}{\mathcal{J}_{W_i}\mathcal{T}_n^{-1}(y)} \right| \\
&\le C_\mathrm{r}|W(\mathcal{T}_n^{-1}x, \mathcal{T}_n^{-1}y)|^{\eta_\mathrm{r}} + C_{\mathrm{d,q}}|W_i(x,y)| \\
&\le (C_\mathrm{r} C_q^{-\eta_\mathrm{r}} \Lambda_q^{-\eta_\mathrm{r}\cdot n} + C_{\mathrm{d,q}}L^{1-\eta_\mathrm{r}})|W_i(x,y)|^{\eta_\mathrm{r}} \le C_\mathrm{r}|W_i(x,y)|^{\eta_\mathrm{r}}.
\end{aligned}$$

We used Lemma 8 and also $W_i(x,y) = \mathcal{T}_n(W(\mathcal{T}_n^{-1}x, \mathcal{T}_n^{-1}y))$ together with Lemma 5. □

Thus, $\mathcal{G}_n = \{(W_i, \nu_i)\}$ is a standard family equipped with the factor measure $\lambda_{\mathcal{G}_n}(i) = c_i$. More generally, if $\mathcal{G}$ is a standard family, $\mathcal{G}_n$ obtained by processing each standard pair in a similar fashion is a standard family.

2.4. **Holonomy maps.** A holonomy map is a device needed in the coupling construction for coupling some of the probability masses on different points. Let W_1 and W_2 be two unstable curves (w.r.t. $\{\mathcal{C}^u_{1,x}\}$) connected by the stable foliation $\mathcal{W}^0$. In other words, for each point $x \in W_1$ the leaf $\mathcal{W}^0_x$ intersects W_2 and conversely for each point $y \in W_2$ the leaf $\mathcal{W}^0_y$ intersects W_1. We assume that the curves W_i are close enough and not too long, so that the connected pairs $(x, y) \in W_1 \times W_2$ are uniquely defined by demanding that the connecting leaf be shorter than a small number $\ell_0 < 1$. Then the holonomy map $\mathbf{h} : W_1 \to W_2$ is defined by sliding along the leaf: $\mathbf{h}x = y$. Since the images $\mathcal{T}_n W_i$ are connected by the stable foliation $\mathcal{W}^n$, one can define the holonomy map $\mathbf{h}_n = \mathcal{T}_n \circ \mathbf{h} \circ \mathcal{T}_n^{-1} : \mathcal{T}_n W_1 \to \mathcal{T}_n W_2$.

Remark 10. *Notice that if the curves W_i carry measures ν_i that are compatible in the sense that $\nu_2 = \mathbf{h}\nu_1$, then the images $\mathcal{T}_n W_i$ carry compatible measures: $\mathcal{T}_n \nu_2 = \mathbf{h}_n \mathcal{T}_n \nu_1$. This will guarantee in the following that once some of the masses on two points have been coupled to each other, they remain coupled.*

The holonomy map $\mathbf{h}$ is said to be absolutely continuous, if the measure $\mathbf{h}^{-1}m_{W_2}$ is absolutely continuous with respect to the measure m_{W_1}. In this case the Jacobian, which measures distortion under the holonomy map, is defined as the Radon–Nikodym derivative $\mathcal{J}\mathbf{h} = \frac{d(\mathbf{h}^{-1}m_{W_2})}{dm_{W_1}}$. The change-of-variables formula for integrals is $dm_{W_2}(y) = \mathcal{J}\mathbf{h}(x)\, dm_{W_1}(x)$ with $y = \mathbf{h}x$ [5]. For any

[5]Given a Borel set $A \subset W_1$, we have $\int_{\mathbf{h}A} dm_{W_2} = (\mathbf{h}^{-1}m_{W_2})(A) = \int_A \frac{d(\mathbf{h}^{-1}m_{W_2})}{dm_{W_1}}\, dm_{W_1}$.

$x \in W_1$, we have $\mathbf{h}x = \mathcal{T}_n^{-1}\mathbf{h}_n\mathcal{T}_n x$. It is elementary to check that if $\mathbf{h}_n$ is absolutely continuous, then $\mathbf{h}$ inherits this property via the identity

$$\mathcal{J}\mathbf{h}(x) = \mathcal{J}_{\mathbf{h}_n\mathcal{T}_n W_1}\mathcal{T}_n^{-1}(\mathbf{h}_n\mathcal{T}_n x)\cdot\mathcal{J}\mathbf{h}_n(\mathcal{T}_n x)\cdot\mathcal{J}_{W_1}\mathcal{T}_n(x) = \frac{\mathcal{J}_{W_1}\mathcal{T}_n(x)}{\mathcal{J}_{W_2}\mathcal{T}_n(\mathbf{h}x)}\cdot\mathcal{J}\mathbf{h}_n(\mathcal{T}_n x). \tag{12}$$

By reversing the argument, we see that if $\mathbf{h}_m$ is absolutely continuous for some m, then $\mathbf{h}_n$ is absolutely continuous and (12) holds for all values of $n \geq 0$.

Lemma 11 (Absolute continuity of the holonomy map)**.** *Let* $\mathbf{h}$ *be as above. It is absolutely continuous. Moreover, there exist constants* $c_1 \geq 1$ *and* $0 < \mu < 1$*, independent of the curves* W_1 *and* W_2 *and the choice of the sequence* $(T_i)_{i=1}^{n_Q}$*, such that*

$$|\ln\mathcal{J}\mathbf{h}_n(\mathcal{T}_n x)| \leq c_1\mu^n \tag{13}$$

holds for $x \in W_1$ *and* $0 \leq n \leq n_Q$*. In particular,* $e^{-c_1} \leq \mathcal{J}\mathbf{h} \leq e^{c_1}$*.*

As a curiosity, (13) continues to hold for $n > n_Q$ since $T_n = \widetilde{T}_Q$. In particular, the precise value of $\mathcal{J}\mathbf{h}(x)$ could be obtained as the limit $\lim_{n\to\infty}\frac{\mathcal{J}_{W_1}\mathcal{T}_n(x)}{\mathcal{J}_{W_2}\mathcal{T}_n(\mathbf{h}x)}$. However, we only care about $n \leq n_Q$. What is important above is that c_1 and μ do not change when the lengths of the intervals I_q in (2) and hence the value of n_Q are increased arbitrarily.

Proof of Lemma 11. Denote $x^n = \mathcal{T}_n x$ and $y = \mathbf{h}x$ for $x \in W_1$, and $W_i^n = \mathcal{T}_n W_i$ for $i = 1, 2$. We also write $y^n = \mathcal{T}_n y = \mathbf{h}_n x^n$. Since $T_n = \widetilde{T}_Q$ and $\mathcal{W}_x^{n-1} = \mathcal{W}_{Q,x}^s$ for all x, for all $n > n_Q$, we know the following: $\mathbf{h}$ inherits absolute continuity from $\mathbf{h}_{n_Q}$, (12) holds for all $n \geq 0$ as explained above, and $\lim_{n\to\infty}\mathcal{J}\mathbf{h}_n(x^n) = 1$. Therefore, with the aid of the chain rule $\mathcal{J}_W\mathcal{T}_n(x) = \mathcal{J}_{W^{n-1}}T_n(x^{n-1})\cdots\mathcal{J}_{W^0}T_1(x^0)$, we conclude that

$$\mathcal{J}\mathbf{h}_m(x^m) = \prod_{n\geq m}\frac{\mathcal{J}_{W_1^n}T_{n+1}(x^n)}{\mathcal{J}_{W_2^n}T_{n+1}(y^n)}.$$

By Assumption (A3), we may use Lemma 27 on each of the time intervals I_q. Since $|\ln z| \leq \max(z-1, z^{-1}-1)$ for all $z > 0$, we see that (27) implies (13):

$$|\ln\mathcal{J}\mathbf{h}_m(x^m)| \leq \sum_{n\geq m}\left|\ln\frac{\mathcal{J}_{W_1^n}T_{n+1}(x^n)}{\mathcal{J}_{W_2^n}T_{n+1}(y^n)}\right| \leq C'\frac{\mu^m}{1-\mu} = c_1\mu^m.$$

□

Lemma 12 (Regularity of the holonomy map)**.** *There exist* $0 < \eta_{\mathbf{h}} < 1$ *and* $C_{\mathbf{h}} > 0$*, such that the following holds. Let* W_1 *and* W_2 *be standard unstable curves connected by the stable foliation* $\mathcal{W}^0$ *as above. For* $x_1, x_2 \in W_1$ *such that* $|W_1(x_1, x_2)| \leq 1$*,*

$$|\ln\mathcal{J}\mathbf{h}(x_1) - \ln\mathcal{J}\mathbf{h}(x_2)| \leq C_{\mathbf{h}}|W_1(x_1, x_2)|^{\eta_{\mathbf{h}}}.$$

Proof. Denote $x^n = \mathcal{T}_n x$ for all x in $W_1^n = \mathcal{T}_n W_1$. We also set $y_i = \mathbf{h}x_i$, $y_i^n = \mathcal{T}_n y_i$, and $W_2^n = \mathcal{T}_n W_2$. By (12),

$$|\ln\mathcal{J}\mathbf{h}(x_1) - \ln\mathcal{J}\mathbf{h}(x_2)| \leq \left|\ln\frac{\mathcal{J}_{W_1}\mathcal{T}_m(x_1)}{\mathcal{J}_{W_1}\mathcal{T}_m(x_2)}\right| + \left|\ln\frac{\mathcal{J}_{W_2}\mathcal{T}_m(y_2)}{\mathcal{J}_{W_2}\mathcal{T}_m(y_1)}\right| + \sum_{i=1,2}|\ln\mathcal{J}\mathbf{h}_m(x_i^m)|,$$

for all m. The last sum can be bounded with the aid of (13). Using the chain rule $\mathcal{J}_W\mathcal{T}_n(x) = \mathcal{J}_{W^{n-1}}T_n(x^{n-1})\cdots\mathcal{J}_{W^0}T_1(x^0)$, Lemma 8 and the bounds (7),

$$\begin{aligned}
\left|\ln\frac{\mathcal{J}_{W_1}\mathcal{T}_{n_Q}(x_1)}{\mathcal{J}_{W_1}\mathcal{T}_{n_Q}(x_2)}\right| &\le \sum_{0\le q\le Q-1}\left|\ln\frac{\mathcal{J}_{W_1^{n_q}}\mathcal{T}_{n_{q+1},n_q+1}(x_1^{n_q})}{\mathcal{J}_{W_1^{n_q}}\mathcal{T}_{n_{q+1},n_q+1}(x_2^{n_q})}\right| \\
&\le \sum_{0\le q\le Q-1} C_{\mathrm{d},q+1}\left|\mathcal{T}_{n_{q+1},n_q+1}\left(W_1^{n_q}(x_1^{n_q},x_2^{n_q})\right)\right| \\
&\le \sum_{0\le q\le Q-1} C_{\mathrm{d},q+1}\bar\Lambda_{q+1}^{n_{q+1}-n_q}\left|W_1^{n_q}(x_1^{n_q},x_2^{n_q})\right| \\
&\le \sum_{0\le q\le Q-1} C_{\mathrm{d},q+1}\bar\Lambda_{q+1}^{n_{q+1}-n_q}\cdots\bar\Lambda_2^{n_2-n_1}\bar\Lambda_1^{n_1}|W_1(x_1,x_2)| \\
&\le \left(\max_{1\le q\le Q} C_{\mathrm{d},q}\right)|W_1(x_1,x_2)|\sum_{1\le q\le Q}\left(\max_{1\le q\le Q}\bar\Lambda_q\right)^{n_q} \\
&\le \frac{\max_{1\le q\le Q} C_{\mathrm{d},q}}{1-\left(\max_{1\le q\le Q}\bar\Lambda_q\right)^{-1}}\left(\max_{1\le q\le Q}\bar\Lambda_q\right)^{n_Q}|W_1(x_1,x_2)| \\
&= C\bar\Lambda^{n_Q}|W_1(x_1,x_2)|.
\end{aligned}$$

Similarly, for any m,

$$\left|\ln\frac{\mathcal{J}_{W_1}\mathcal{T}_m(x_1)}{\mathcal{J}_{W_1}\mathcal{T}_m(x_2)}\right| \le C\bar\Lambda^m|W_1(x_1,x_2)| \quad\text{and}\quad \left|\ln\frac{\mathcal{J}_{W_2}\mathcal{T}_m(y_2)}{\mathcal{J}_{W_2}\mathcal{T}_m(y_1)}\right| \le C\bar\Lambda^m|W_2(y_1,y_2)|.$$

Notice from the definition of C and $\bar\Lambda$ that they are independent of the curves W_i and of the sequence (T_i). Because $|W_2(y_1,y_2)| \le \sup_{W_1(x_1,x_2)}\mathcal{J}\mathbf{h}\cdot|W_1(x_1,x_2)|$,

$$\left|\ln\frac{\mathcal{J}_{W_1}\mathcal{T}_m(x_1)}{\mathcal{J}_{W_1}\mathcal{T}_m(x_2)}\right| + \left|\ln\frac{\mathcal{J}_{W_2}\mathcal{T}_m(y_2)}{\mathcal{J}_{W_2}\mathcal{T}_m(y_1)}\right| \le C\bar\Lambda^m(1+e^{c_1})|W_1(x_1,x_2)| = c_2\bar\Lambda^m|W_1(x_1,x_2)|.$$

Finally, choose $m = \frac{\ln|W_1(x_1,x_2)|}{\ln\mu}$. Then $\mu^m = |W_1(x_1,x_2)|$, $\bar\Lambda^m = |W_1(x_1,x_2)|^{\ln\bar\Lambda/\ln\mu}$, and

$$|\ln\mathcal{J}\mathbf{h}(x_1) - \ln\mathcal{J}\mathbf{h}(x_2)| \le 2c_1\mu^m + c_2\bar\Lambda^m|W_1(x_1,x_2)| \le (2c_1+c_2)|W_1(x_1,x_2)|^{1-\ln\bar\Lambda/|\ln\mu|}.$$

□

3. Coupling Lemma and the proof of Theorems 4 and 2

Let (W,ν) be a standard pair and $d\nu = \rho\, dm_W$. We will be interested in densities of the form $\tau\rho$ where $\tau : W \to [0,1]$ is a function. These can be considered as portions of the measure ν. In practice, we will replace W by the rectangle $\hat W = W\times[0,1]$ with base W and $d\nu$ by the measure $d\hat\nu = d\nu\otimes dt$, where dt denotes the Lebesgue measure on $[0,1]$, and look at the subdomain $\{(x,t)\in\hat W : 0\le t\le\tau(x)\}$ of $\hat W$. Introducing the rectangle facilitates bookkeeping.

A standard family $\mathcal{G} = \{(W_\alpha,\nu_\alpha)\}_{\alpha\in\mathfrak{A}}$ can similarly be replaced by $\hat{\mathcal{G}} = \{(\hat W_\alpha,\hat\nu_\alpha)\}_{\alpha\in\mathfrak{A}}$. The measure $\mu_{\mathcal{G}}$ induces canonically a measure $\hat\mu_{\mathcal{G}}$ on $\cup_\alpha\hat W_\alpha$. A map T on $\mathcal{M}$ extends to a map on $\mathcal{M}\times[0,1]$ by setting $T(x,t)\equiv(T(x),t)$ and all observables on $\mathcal{M}$ extend to observables on $\mathcal{M}\times[0,1]$ by setting $f(x,t)\equiv f(x)$.

We are now in position to state the following key result.

Lemma 13 (Coupling Lemma)**.** *Consider two standard families $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}_{\alpha\in\mathfrak{A}}$ and $\mathcal{E} = \{(W_\beta, \nu_\beta)\}_{\beta\in\mathfrak{B}}$. There exist an almost everywhere defined bijective map $\Theta : \cup_\alpha \hat{W}_\alpha \to \cup_\beta \hat{W}_\beta$, called the coupling map, that preserves measure, i.e., $\Theta(\hat{\mu}_{\mathcal{G}}) = \hat{\mu}_{\mathcal{E}}$, and an almost everywhere defined function $\Upsilon : \cup_\alpha \hat{W}_\alpha \to \mathbb{N}$, called the coupling time, both depending on the sequence (T_i), such that the following hold:*

1. *Let $(x,t) \in \hat{W}_\alpha$, $\alpha \in \mathfrak{A}$, and $\Theta(x,t) = (y,s) \in \hat{W}_\beta$, $\beta \in \mathfrak{B}$. Then the points x and y lie on the same leaf, say W, of the stable foliation $\mathcal{W}^0$. If $n \ge \Upsilon(x,t)$, then the distance of the points $\mathcal{T}_n x$ and $\mathcal{T}_n y$ along the leaf $\mathcal{T}_n W$ of the stable foliation $\mathcal{W}^n$ satisfies $|\mathcal{T}_n W(\mathcal{T}_n x, \mathcal{T}_n y)| < \ell_0 \lambda^{n-\Upsilon(x,t)}$. Here $\ell_0 > 0$ has been introduced earlier and $\lambda = \max_{1\le q\le Q} \Lambda_q^{-1} < 1$.*
2. *The exponential tail bound*
$$\hat{\mu}_{\mathcal{G}}(\Upsilon > n) \le C_\Upsilon \vartheta_\Upsilon^n \tag{14}$$
holds for uniform constants $C_\Upsilon > 0$ and $\vartheta_\Upsilon \in (0,1)$.

Proof of Theorem 4. We use the coupling between $\mathcal{G}$ and $\mathcal{E}$ given in the Coupling Lemma:

$$\begin{aligned}
&\int_{\mathcal{M}} f\circ\mathcal{T}_n \, d\mu_{\mathcal{G}} - \int_{\mathcal{M}} f\circ\mathcal{T}_n \, d\mu_{\mathcal{E}} \\
&\quad= \int_{\mathcal{M}\times[0,1]} (f\circ\mathcal{T}_n)(x,t)\, d\hat{\mu}_{\mathcal{G}}(x,t) - \int_{\mathcal{M}\times[0,1]} (f\circ\mathcal{T}_n)(y,s)\, d\hat{\mu}_{\mathcal{E}}(y,s) \\
&\quad= \int_{\mathcal{M}\times[0,1]} (f\circ\mathcal{T}_n)(x,t)\, d\hat{\mu}_{\mathcal{G}}(x,t) - \int_{\mathcal{M}\times[0,1]} (f\circ\mathcal{T}_n\circ\Theta)(x,t)\, d\hat{\mu}_{\mathcal{G}}(x,t) \\
&\quad= \int_{\mathcal{M}\times[0,1]} (f\circ\mathcal{T}_n - f\circ\mathcal{T}_n\circ\Theta)\, d\hat{\mu}_{\mathcal{G}} \\
&\quad= \int_{\Upsilon\le n/2} (f\circ\mathcal{T}_n - f\circ\mathcal{T}_n\circ\Theta)\, d\hat{\mu}_{\mathcal{G}} + \int_{\Upsilon> n/2} (f\circ\mathcal{T}_n - f\circ\mathcal{T}_n\circ\Theta)\, d\hat{\mu}_{\mathcal{G}}.
\end{aligned}$$

By (14),
$$\left|\int_{\Upsilon>n/2} (f\circ\mathcal{T}_n - f\circ\mathcal{T}_n\circ\Theta)\, d\hat{\mu}_{\mathcal{G}}\right| \le C_\Upsilon(\sup f - \inf f)\vartheta_\Upsilon^{n/2}.$$

On the other hand, assume $\Upsilon(x,t) \le n/2$. Then $|(f\circ\mathcal{T}_n - f\circ\mathcal{T}_n\circ\Theta)(x,t)| \le |f|_\gamma (\ell_0\lambda^{n-\Upsilon(x,t)})^\gamma$, by the Coupling Lemma, such that
$$\left|\int_{\Upsilon\le n/2} (f\circ\mathcal{T}_n - f\circ\mathcal{T}_n\circ\Theta)\, d\hat{\mu}_{\mathcal{G}}\right| \le \ell_0^\gamma |f|_\gamma \lambda^{\gamma n/2}.$$

Since $\ell_0 < 1$, the proof is complete. □

Proof of Theorem 2. First notice that both of the measures μ^i can be disintegrated using a suitable measurable partition of $\mathcal{M}$ so that we almost obtain two standard families, with the nuisance that the Hölder constants of the logarithms of the conditional measures possibly exceed $C_{\rm r}$ in (4). In the latter case we need a finite waiting time $N = N(\rho^1, \rho^2)$, depending on the Hölder constants of $\ln\rho^i$, until the densities regularize and yield true standard families; see the proof of Lemma 9. For γ-Hölder observables the result then follows immediately from Theorem 4, with the above waiting time giving the constant $C_\gamma(\rho^1,\rho^2) = \theta_\gamma^{-N}$. If f is only continuous, we fix an arbitrarily small $\varepsilon > 0$ and, by Stone–Weierstrass theorem, pick a γ-Hölder f_ε such that $\|f - f_\varepsilon\|_\infty < \varepsilon$. Then $\left|\int_\mathcal{M} f \circ \mathcal{T}_n \, d\mu^1 - \int_\mathcal{M} f \circ \mathcal{T}_n \, d\mu^2\right| < C_\gamma(\rho^1, \rho^2) B_{f_\varepsilon} \theta_\gamma^n + 2\varepsilon < 3\varepsilon$ if $n > \ln(\varepsilon / C_\gamma(\rho^1, \rho^2) B_{f_\varepsilon}) / \ln \theta_\gamma$. □

4. Proof of the Coupling Lemma

4.1. **Outline of the proof.** The idea of the proof is to construct special tiny rectangles, called magnets, which can be thought to attract unstable curves. Mixing guarantees that a small fraction, say 1 percent, of any high enough iterate of any unstable curve will ultimately lie on a magnet. Once two unstable curves from two different standard families cross a magnet, we are able to couple a fraction of their masses by connecting some of their points lying on the magnet with very short stable manifolds. This has to be done with due care, because the resulting coupling has to be measure preserving.

The process is then repeated recursively, and so the construction of the coupling map Θ and the coupling time function Υ is recursive. It can be shown that after a fixed finite number of iterates a fixed fraction of the remaining masses can always be coupled, so that the measures on the unstable curves can be 'drained' at an exponential rate.

Since we are dealing with compositions of diffeomorphisms from the sequence (T_i) rather than iterates of a single diffeomorphism, we need to use time-dependent magnets. For $n \in I_q$, $T_n \in \mathcal{U}_q$, and the magnet to be used should reflect the structure of the reference diffeomorphism $\widetilde{T}_q$. In our time-dependent, finite time, setting it is not even *a priori* clear what coupling should mean. We choose to construct a coupling via the stable foliations $\mathcal{W}^n$ that vary from one point in time to the next. As mentioned earlier, these foliations are artificial in the sense that they depend on the artificial future $T_n = \widetilde{T}_Q$ for times $n > n_Q$ although in reality we only consider the compositions $\mathcal{T}_n = T_n \circ \cdots \circ T_1$ for $n \le n_Q$. We then have to pay special attention to uniformity: our convergence rates, *etc.*, should depend neither on the particular value of n_Q nor on the particular finite sequence $(T_1, \dots, T_{n_Q})$ as long as the reference automorphisms $(\widetilde{T}_1, \dots, \widetilde{T}_Q)$ have been chosen and the earlier assumptions on $(T_1, \dots, T_{n_Q})$ are being respected.

4.2. **Magnets and crossings.** In this subsection $1 \le q \le Q$ is fixed for good. Unstable curves and standard pairs are to be understood as being defined with respect to the cone family $\{\mathcal{C}^u_{q,x}\}$ with q fixed.

Consider the Anosov diffeomorphism $\widetilde{T}_q$. A 'rectangle', $\mathfrak{R} \subset \mathcal{M}$, is a closed and connected region bounded by two stable manifolds and two unstable manifolds of $\widetilde{T}_q$. These are called the *s- and u-sides* of the rectangle, respectively. Recalling that $\mathcal{W}^s_{q,x}$ denotes the stable leaf of $\widetilde{T}_q$ at x, we also assume that the size of the rectangle in the stable direction satisfies $|\mathcal{W}^s_{q,x} \cap \mathfrak{R}| \ll \ell_0$.

We say that an unstable curve W crosses the rectangle *properly*, if

(P1) W crosses $\mathfrak{R}$ completely, *i.e.*, $W \cap \mathfrak{R}$ contains a connected curve W' connecting the two s-sides of the rectangle, and
(P2) both components of $W \setminus W'$ are of length strictly greater than $\ell/10$,

both hold. In other words, a crossing is proper if the curve crosses the rectangle completely and there is a guaranteed amount of excess length beyond each s-side of the rectangle. Here ℓ is the lower bound on the length of a standard curve. Finally, an unstable curve W crosses the rectangle *super-properly*, if (P1),

(P2') both components of $W \setminus W'$ are of length strictly greater than $\ell/5$, and
(P3) each $x \in W \cap \mathfrak{R}$ divides the curve $\mathcal{W}^s_{q,x} \cap \mathfrak{R}$ in a ratio strictly between $1/10$ and $9/10$.

all hold. Thus, in a super-proper crossing there is more guaranteed excess length than in a proper crossing and the curve also stays well clear of the u-sides.

Lemma 14. *There exists a finite set of rectangles, $\{\mathfrak{R}^k : 1 \le k \le k_0\}$, such that each standard unstable curve crosses at least one of the rectangles super-properly.*

Proof. Every closed standard curve crosses some rectangle $\mathfrak{R}$ super-properly. Since crossing a rectangle $\mathfrak{R}$ super-properly is an open condition in the Hausdorff metric, the set $\mathcal{U}_{\mathfrak{R}}$ of all closed standard curves crossing $\mathfrak{R}$ super-properly is an open set. The collection formed by all the sets $\mathcal{U}_{\mathfrak{R}}$ is an open cover of the space of closed standard curves equipped with the Hausdorff metric. The latter space is compact. We can therefore pick a finite subcover and the corresponding rectangles. □

We now pick arbitrarily one of the rectangles $\mathfrak{R}^k$. This special rectangle, that we will denote by $\mathfrak{R}_q$, will be called a *magnet*. It will serve as a reference set on which points will be coupled.

Lemma 15. *Fix $n \ge 1$. By taking ε_q sufficiently small (depending on n) the following holds. If W is an unstable curve and $\widetilde{T}_q^n W$ crosses $\mathfrak{R}_q$ super-properly, then $\mathcal{T}_n W$ crosses $\mathfrak{R}_q$ properly, provided each $T_i \in \mathcal{U}_q$.*

Proof. For any point x, we have the bound $d(\mathcal{T}_n x, \widetilde{T}_q^n x) \le C(n)\varepsilon_q$. □

If W is an unstable curve and n is fixed, let $W_{n,i}^q$, $i \in \mathfrak{I}$, be the connected components of $\widetilde{T}_q^n W \cap \mathfrak{R}_q$, that correspond to super-proper crossings. That is, each $W_{n,i}^q$ is a subset of a longer curve $\widetilde{W}_{n,i}^q \subset \widetilde{T}_q^n W$ which crosses $\mathfrak{R}_q$ super-properly and $\widetilde{W}_{n,i}^q \cap \mathfrak{R}_q = W_{n,i}^q$.

Lemma 16. *There exist a subrectangle $\mathfrak{B}_q \subset \mathfrak{R}_q$ and a number $s' \ge 1$ such that the following holds. Assume that W is an unstable curve that crosses a rectangle $\mathfrak{R}^k$ properly, $n \ge s'$, and $\widetilde{T}_q^n \mathfrak{R}^k \cap \mathfrak{B}_q \ne \varnothing$. Every component of $\widetilde{T}_q^n \mathfrak{R}^k \cap \mathfrak{R}_q$ that intersects $\mathfrak{B}_q$ intersects $W_{n,i}^q$ for precisely one value of the index $i \in \mathfrak{I}$.*

In words, each intersection of $\widetilde{T}_q^n \mathfrak{R}^k$ with $\mathfrak{B}_q$ yields a super-proper crossing of $\widetilde{T}_q^n W$, as long as n is large enough.

Proof. We assume that the magnet $\mathfrak{R}_q$ is so small that the leaves of the unstable foliation of $\widetilde{T}_q$ are almost parallel lines on $\mathfrak{R}_q$. This can be guaranteed by considering only sufficiently small rectangles in the proof of Lemma 14. Now, choose $\mathfrak{B}_q \subset \mathfrak{R}_q$ to be a rectangle whose distance to the u-sides of $\mathfrak{R}_q$ is sufficiently large; say each $x \in \mathfrak{B}_q$ divides the curve $\mathcal{W}_{q,x}^s \cap \mathfrak{R}_q$ in a ratio between 1/5 and 4/5. As $\widetilde{T}_q$ is one-to-one, the components of $\widetilde{T}_q^n \mathfrak{R}^k \cap \mathfrak{R}_q$ are disjoint. Assuming s' is large, these components are very thin strips, almost aligned with the unstable foliation. Pick such a component and assume that it intersects $\mathfrak{B}_q$. It is a safe distance away from the u-sides of the magnet. Inside this component lies a piece V of the curve $\widetilde{T}_q^n W$. The piece V has to extend to a super-proper crossing of $\mathfrak{R}_q$, because W crosses $\mathfrak{R}^k$ properly and because n_1 is large. Thus, V is actually a subcurve of one of the $W_{n,i}^q$. □

Lemma 17. *There exist numbers $d'' > 0$ and $s'' \ge 1$ such that if (W, ν) is a standard pair and $n \ge s''$, then $\nu(\widetilde{T}_q^{-n}(\cup_{i\in\mathfrak{I}} W_{n,i}^q)) \ge d''$.*

In other words, the fraction of W that will cross the magnet $\mathfrak{R}_q$ super-properly after n steps is at least d''.

Proof. Fix a k such that W crosses $\mathfrak{R}^k$ properly. This is possible by Lemma 14. By Lemma 16, if a component of $\widetilde{T}_q^n\mathfrak{R}^k \cap \mathfrak{R}_q$ intersects the subrectangle $\mathfrak{B}_q$, it is crossed by the curve component $W_{n,i}^q$ for precisely one value of the index $i \in \mathfrak{I}$. In this case let $\mathfrak{R}_{n,i}^{k,q}$ denote the former component of $\widetilde{T}_q^n\mathfrak{R}^k \cap \mathfrak{R}_q$. Thus $\mathfrak{R}_{n,i}^{k,q}$ is only defined for a subset $\mathfrak{I}_{\mathfrak{B}_q} \subset \mathfrak{I}$ of indices. We have $\nu(\widetilde{T}_q^{-n}(\cup_{i\in\mathfrak{I}}W_{n,i}^q)) \ge \nu(\widetilde{T}_q^{-n}(\cup_{i\in\mathfrak{I}_{\mathfrak{B}_q}}W_{n,i}^q)) = \nu(\widetilde{T}_q^{-n}(\cup_{i\in\mathfrak{I}_{\mathfrak{B}_q}}W_{n,i}^q \cap \mathfrak{R}_{n,i}^{k,q})) \ge c\mu_q(\widetilde{T}_q^{-n}(\cup_{i\in\mathfrak{I}_{\mathfrak{B}_q}}\mathfrak{R}_{n,i}^{k,q})) = c\mu_q(\cup_{i\in\mathfrak{I}_{\mathfrak{B}_q}}\mathfrak{R}_{n,i}^{k,q}) \ge c\mu_q(\cup_{i\in\mathfrak{I}_{\mathfrak{B}_q}}\mathfrak{R}_{n,i}^{k,q} \cap \mathfrak{B}_q) = c\mu_q(\widetilde{T}_q^n\mathfrak{R}^k \cap \mathfrak{B}_q) \ge \frac{c}{2}\mu_q(\mathfrak{R}^k)\mu_q(\mathfrak{B}_q)$ if $n \ge s''$ and s'' is large. The last step in the estimate follows from mixing of the invariant measure μ_q. The third step relies on the absolute continuity with bounded Jacobians of the holonomy maps of $\widetilde{T}_q$ as well as on the regularity of ν and of the conditional measures of μ_q on the unstable leaves of $\widetilde{T}_q$. To finish, notice that $d'' = \mu_q(\mathfrak{R}^k)\mu_q(\mathfrak{B}_q) > 0$, since the interiors of $\mathfrak{R}^k$ and $\mathfrak{B}_q$ are nonempty. □

For an unstable curve W, let $W_{n,i}$ now be the connected components of $\mathcal{T}_nW \cap \mathfrak{R}_q$, labeled by i, that correspond to proper crossings. That is, each $W_{n,i}$ is a subset of a longer curve $\widetilde{W}_{n,i} \subset \mathcal{T}_nW$ which crosses $\mathfrak{R}_q$ properly and $\widetilde{W}_{n,i} \cap \mathfrak{R}_q = W_{n,i}$.

Corollary 18. *There exist numbers $d_0' > 0$ and $s_0' \ge 1$ such that the following holds. Let $T_i \in \mathcal{U}_q$ for each i and (W,ν) be a standard pair. If $n \ge s_0'$, then $\nu(\mathcal{T}_n^{-1}(\cup_iW_{n,i})) \ge d_0'$.*

Proof. Fix $m \ge \ln\frac{2}{C_q}/\ln\Lambda_q$. By Lemma 9, $(\mathcal{T}_mW, \mathcal{T}_m\nu)$ can be broken into a finite collection of standard pairs (W_j,ν_j) such that $\mathcal{T}_mW = \cup_jW_j$ and $\mathcal{T}_m\nu = \sum_j c_j\nu_j$, where $0 < \nu_j < 1$ and $\sum_j c_j = 1$. For each j, by Lemma 17, there is a finite collection of disjoint (minimal) subcurves $V_{j,k} \subset W_j$ such that $\widetilde{T}_q^{s''}V_{j,k}$ crosses the magnet $\mathfrak{R}_q$ super-properly. Moreover, $\sum_k \nu_j(V_{j,k}) \ge d''$ and, by Lemma 15, the images $\mathcal{T}_{m+s'',m+1}V_{j,k}$ cross the magnet $\mathfrak{R}_q$ properly. We also have $\mathcal{T}_m\nu(\cup_j\cup_kV_{j,k}) = \sum_j c_j\sum_k\nu_j(V_{j,k}) \ge d''$. Let us relabel the collection of the subcurves $\mathcal{T}_m^{-1}V_{j,k} \subset W$ by U_l. We have so far shown that $\nu(\cup_lU_l) \ge d''$ and that $\mathcal{T}_{m+s''}U_l$ crosses the magnet $\mathfrak{R}_q$ properly.

We are almost done, but we still need to truncate each U_l to a subcurve $\tilde{U}_l$ so that $\mathcal{T}_{m+s''}\tilde{U}_l = \mathcal{T}_{m+s''}U_l \cap \mathfrak{R}_q$ and argue that $\nu(\cup_l\tilde{U}_l) \ge \alpha\nu(\cup_lU_l)$ for some constant $\alpha > 0$. Such a truncation amounts to choosing the subcurve $\tilde{V}_{j,k} = \mathcal{T}_m\tilde{U}_l$ of $V_{j,k} \subset W_j$ so that $\mathcal{T}_{m+s'',m+1}\tilde{V}_{j,k} = \mathcal{T}_{m+s'',m+1}V_{j,k} \cap \mathfrak{R}_q$. Now $\nu_j(\tilde{V}_{j,k}) \ge \alpha\nu_j(V_{j,k})$ follows from two observation:

- $|\tilde{V}_{j,k}|$ is bounded uniformly away from zero, because $\mathcal{T}_{m+s'',m+1}\tilde{V}_{j,k}$ crosses $\mathfrak{R}_q$ completely.
- $|\mathcal{T}_{m+s'',m+1}V_{j,k} \setminus \mathfrak{R}_q|$ is bounded uniformly from above, because $|\widetilde{T}_q^{s''}V_{j,k} \setminus \mathfrak{R}_q|$ was assumed to be as small as possible (for a super-properly crossing curve). Therefore $|V_{j,k} \setminus \tilde{V}_{j,k}| = |V_{j,k} \setminus \mathcal{T}_{m+s'',m+1}^{-1}\mathfrak{R}_q|$ is bounded uniformly from above.

Indeed, it is implied that $|\tilde{V}_{j,k}|/|V_{j,k}| \ge \alpha'$ for some $\alpha' \in (0,1]$ so that, by estimate (5), $\nu_j(\tilde{V}_{j,k})/\nu_j(V_{j,k}) \ge D^{-1}\alpha'$ for all j,k.

Notice that only s'' affected the size of ε_q. This happened when Lemma 15 was used. □

4.3. **Time-dependent magnets.** For the rest of the section, let (T_i) be a sequence of the form described in the Introduction, which is not confined to a neighborhood $\mathcal{U}_q$ of any one map $\widetilde{T}_q$.

The Coupling Lemma needs to hold for the compositions in (3). For this reason we cannot use the same magnet for all times. Moreover, the stable foliation $\mathcal{W}^n$ that we will use to couple points (more correctly some of the probability masses carried by these points) on the magnets changes with time. This will guarantee that what has aleady been coupled will always remain coupled.

Therefore, we need to introduce the following time-dependent magnets: For every $q \in \{1, \dots, Q\}$ and every $n \in I_q$, define

$$\mathfrak{M}_n = \{x \in \mathfrak{R}_q \,:\, \mathcal{W}^n_x \cap \mathfrak{R}_q \text{ connects the u-sides of } \mathfrak{R}_q \text{ and has only one component}\}.$$

In other words, $\mathfrak{M}_n$ consists of those leaves of $\mathcal{W}^n$ that connect the u-sides of $\mathfrak{R}_q$ *and* are entirely inside $\mathfrak{R}_q$.

We say that an unstable curve W crosses $\mathfrak{M}_n$ properly if $n \in I_q$ and W crosses $\mathfrak{R}_q$ properly. Assuming that the cones $\mathcal{C}^s_{q,x}$ are narrow enough, $\mathfrak{M}_n$ is close to $\mathfrak{R}_q$, and we have $|W \cap \mathfrak{M}_n| \geq \frac{3}{4}|W \cap \mathfrak{R}_q|$.

Lemma 19. *Assume that W crosses the magnet $\mathfrak{R}_q$ properly and that W carries the measure $d\nu = \rho\, dm_W$, where ρ satisfies (4). Then, for all $n \in I_q$,*

$$\nu(W \cap \mathfrak{M}_n) \geq \frac{1}{2}\nu(W \cap \mathfrak{R}_q).$$

Proof. By (4), $\rho(y) \geq \rho(x)e^{-C_\mathrm{r}|W\cap\mathfrak{R}_q|^{\eta_\mathrm{r}}}$ for any $x, y \in W \cap \mathfrak{R}_q$. In particular, we can average the left side over $W \cap \mathfrak{M}_n$ and the right side over $W \cap \mathfrak{R}_q$, obtaining $\nu(W \cap \mathfrak{M}_n) \geq \frac{|W\cap\mathfrak{M}_n|}{|W\cap\mathfrak{R}_q|}e^{-C_\mathrm{r}|W\cap\mathfrak{R}_q|^{\eta_\mathrm{r}}}\nu(W \cap \mathfrak{R}_q)$. Since $\mathfrak{R}_q$ has small diameter, the exponential factor is close to 1. For $n \in I_q$, we also have $|W \cap \mathfrak{M}_n| \geq \frac{3}{4}|W \cap \mathfrak{R}_q|$. □

For a standard family with respect to $\{\mathcal{C}^u_{1,x}\}$, $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}_{\alpha\in\mathfrak{A}}$, let $W_{\alpha,n,i}$ be the connected components of $\mathcal{T}_n W_\alpha \cap \mathfrak{M}_n$ that correspond to proper crossings, and introduce the notation

$$W_{\alpha,n,\star} = \mathcal{T}_n^{-1}(\cup_i W_{\alpha,n,i}).$$

Next, we generalize Corollary 18 of Lemma 17.

Lemma 20. *There exist numbers $s_0 \geq 1$ and $d_0 > 0$, such that the following holds. If (T_i) is a sequence of the general form described in the Introduction and $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}_{\alpha\in\mathfrak{A}}$ a standard family with respect to $\{\mathcal{C}^u_{1,x}\}$ then, for $1 \leq q \leq Q$ and $n_{q-1} + s_0 \leq n \leq n_q$,*

$$\mu_\mathcal{G}(\cup_\alpha W_{\alpha,n,\star}) = \int_\mathfrak{A} \nu_\alpha(W_{\alpha,n,\star})\, d\lambda_\mathcal{G}(\alpha) \geq d_0.$$

In other words, there are time windows for n, such that a significant fraction of the image under $\mathcal{T}_n$ of the standard family $\mathcal{G}$ lies on the magnet $\mathfrak{M}_n$, ready to be coupled.

Remark 21. *Fixing some $1 \leq q \leq Q$ and $k \in I_q$, Lemma 20 can be applied to the shifted sequence $(T_i)_{i\geq k}$ and $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}_{\alpha\in\mathfrak{A}}$ a standard family with respect to $\{\mathcal{C}^u_{q,x}\}$. Then, if $k - 1 + s_0 \leq n \leq n_q$, as well as if $n_{q'-1} + s_0 \leq n \leq n_{q'}$ for some $q' \in \{q+1, \dots, Q\}$, at least a d_0-fraction of its image under $\mathcal{T}_{n,k}$ lies on the magnet $\mathfrak{M}_n$ as a result of proper crossings.*

Proof of Lemma 20. The s_0' in Corollary 18 depends on q. We take s_0 larger than the maximum of these numbers over $1 \leq q \leq Q$. If $n_{q-1} + s_0 \leq n \leq n_q$, and if s_0 is taken sufficiently larger than s_0', then the image of $\mathcal{G}$ under $\mathcal{T}_{n_{q-1}+s_0-s_0'}$ after standardizing the curves becomes a standard family with respect to $\{\mathcal{C}^u_{q,x}\}$. Applying Corollary 18 to this standard family and the map $\mathcal{T}_{n,n_{q-1}+s_0-s_0'+1}$ yields a lower bound on the $\mathcal{T}_n\mu_\mathcal{G}$-measure of proper crossings of $\mathfrak{R}_q$. From this we infer a lower bound on the proper crossings of $\mathfrak{M}_n$ by the regularity of densities with the aid of Lemma 19. □

4.4. **Coupling step.** Consider first two standard families, $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}$ and $\mathcal{E} = \{(W_\beta, \nu_\beta)\}$, consisting of one standard pair each.

A good fraction of the images $\mathcal{T}_{s_0} W_\alpha$ and $\mathcal{T}_{s_0} W_\beta$ cross the magnet $\mathfrak{M}_{s_0}$ properly, so that $\mathcal{T}_{s_0}\nu_\alpha(\cup_i W_{\alpha,s_0,i}) = \nu_\alpha(W_{\alpha,s_0,\star}) > d_0$ and $\mathcal{T}_{s_0}\nu_\beta(\cup_j W_{\beta,s_0,j}) = \nu_\beta(W_{\beta,s_0,\star}) > d_0$. Here i and j run through some finite index sets and $\mathcal{T}_{s_0}\nu$ is the pushforward of ν.

Recall that, for a curve W, $\hat W$ denotes the rectangle $W \times [0,1]$ with base W, and if W carries a measure $d\nu$ then $\hat W$ carries the measure $d\hat\nu = d\nu \otimes dt$. We will construct a coupling from a subset of $\cup_i \hat W_{\alpha,s_0,i}$ to a subset of $\cup_j \hat W_{\beta,s_0,j}$ and then show that the complements of these subsets can be coupled recursively.

With small preliminary preparations, we can assume that the cardinalities of the index sets for i and j are the same, so that we can pair each $\hat W_{\alpha,s_0,i}$ with precisely one $\hat W_{\beta,s_0,i}$. Furthermore, we can assume that their relative masses agree:

$$\frac{\hat\nu_{\alpha,s_0,i}(\hat W_{\alpha,s_0,i})}{Z_{\alpha,s_0}} = \frac{\hat\nu_{\beta,s_0,i}(\hat W_{\beta,s_0,i})}{Z_{\beta,s_0}} \qquad \forall\, i. \tag{15}$$

Here $\hat\nu_{\bullet,s_0,i}$ is the measure on $\hat W_{\bullet,s_0,i}$ and $Z_{\bullet,s_0} = \sum_l \hat\nu_{\bullet,s_0,l}(\hat W_{\bullet,s_0,l}) = \nu_\bullet(W_{\bullet,s_0,\star})$. To see that no generality is lost making such assumptions, consider a rectangle $\hat W$ with a measure $\hat\nu$ on it. We can subdivide it into lower rectangles $W \times I_k$, where $I_k \subset [0,1]$ is an interval. Then, each $W \times I_k$ is stretched affinely onto $W \times [0,1]$ and equipped with the measure $d\hat\nu_k = |I_k|\, d\hat\nu$. In other words, we end up with replicas of $\hat W$ equipped with lowered measures. Such an operation on $\hat W$ is measure preserving, because the pushforward of the measure $\hat\nu|_{W\times I_k}$ under the affine map $\mathbb{A} : W \times I_k \to W \times [0,1]$ is precisely $\hat\nu_k$. Subdividing the rectangles in the families $\{\hat W_{\alpha,s_0,i}\}$ and $\{\hat W_{\beta,s_0,j}\}$ as necessary, and relabeling the resulting rectangles, we can tune the number of rectangles as well as their relative weights so as to arrive at the convenient situation described above. Each rectangle $\hat W_{\bullet,s_0,i}$ now comes with an associated affine map $\mathbb{A}_{\bullet,s_0,i}$ (which is the identity if no subdivision of the particular rectangle was necessary). Some of the rectangles will have a common curve as their base on the manifold $\mathcal{M}$, but this is not a matter of concern.

For each fixed i, we can couple a subset of $\hat W_{\alpha,s_0,i}$ to a subset of $\hat W_{\beta,s_0,i}$ as follows. Choose a number $\tau_\alpha \in (0, 1/2]$ such that

$$\tau_\alpha \cdot Z_{\alpha,s_0} = \frac{d_0}{2}. \tag{16}$$

Now, fix i. Omitting some ornaments for the sake of readability, let $\mathbf{h}$ stand for the holonomy map from $W_{\alpha,s_0,i}$ to $W_{\beta,s_0,i}$ associated with the stable foliation $\mathcal{W}^{s_0}$ and denote by $\rho_\bullet$ the density of $\hat\nu_{\bullet,s_0,i}$ with respect to $dm_{W_{\bullet,s_0,i}} \otimes dt$.

The subset $\hat W'_{\alpha,s_0,i} = \{(x,t) \in \hat W_{\alpha,s_0,i} \,:\, 0 \le t \le \tau_\alpha\}$ is coupled to a corresponding subset $\hat W'_{\beta,s_0,i} = \{(y,s) \in \hat W_{\beta,s_0,i} \,:\, 0 \le s \le \tau_{\beta,i}(y)\}$ via the coupling map $\Theta'_{s_0,i} : \hat W'_{\alpha,s_0,i} \to \hat W'_{\beta,s_0,i} : (x,t) \mapsto (y,s)$ with

$$y = \mathbf{h}x \qquad \text{and} \qquad s = \frac{\tau_{\beta,i}(y)}{\tau_\alpha} t.$$

Notice, however, that $\tau_{\beta,i}$ is not constant but a function. It is given by the consistency rule

$$\tau_{\beta,i}(y)\rho_\beta(y) = \frac{\tau_\alpha \rho_\alpha(x)}{\mathcal{J}\mathbf{h}(x)}. \tag{17}$$

The expression on the right-hand side of (17) equals the pushforward of the density $\tau_\alpha\rho_\alpha$ under the holonomy map, evaluated at $y = \mathbf{h}x$. This guarantees that the coupling is measure preserving:

if f is a measurable function $\hat{W}'_{\beta,s_0,i} \to \mathbb{R}$, then

$$\begin{aligned}\int_{\hat{W}'_{\alpha,s_0,i}} (f \circ \Theta'_{s_0,i})(x,t)\, d\hat{\nu}_{\alpha,s_0,i}(x,t) &= \int_{W_{\alpha,s_0,i}} \int_0^{\tau_\alpha} f(\Theta'_{s_0,i}(x,t))\, \rho_\alpha(x)\, dm_{W_{\alpha,s_0,i}}(x)dt \\ &= \int_{W_{\alpha,s_0,i}} \left[\int_0^{\tau_\alpha} f(\Theta'_{s_0,i}(x,t))\, dt\right] \rho_\alpha(x)\, dm_{W_{\alpha,s_0,i}}(x) \\ &= \int_{W_{\beta,s_0,i}} \left[\int_0^{\tau_\alpha} f\left(y, \frac{\tau_{\beta,i}(y)}{\tau_\alpha} t\right) dt\right] \frac{\rho_\alpha(x)}{\mathcal{J}\mathbf{h}(x)}\, dm_{W_{\beta,s_0,i}}(y) \\ &= \int_{W_{\beta,s_0,i}} \left[\int_0^{\tau_{\beta,i}(y)} f(y,s)\, ds\right] \frac{\tau_\alpha}{\tau_{\beta,i}(y)} \frac{\rho_\alpha(x)}{\mathcal{J}\mathbf{h}(x)}\, dm_{W_{\beta,s_0,i}}(y) = \int_{\hat{W}'_{\beta,s_0,i}} f(y,s)\, d\hat{\nu}_{\beta,s_0,i}(y,s).\end{aligned}$$

We thus have a coupling for each value of the index i, and have therefore managed to couple exactly $d_0/2$ units of mass between the families $\{\hat{W}_{\alpha,s_0,i}\}$ and $\{\hat{W}_{\beta,s_0,i}\}$ via a measure preserving map.

We are now in position to describe the desired coupling map Θ from a subset $\tilde{W}_\alpha \subset \hat{W}_\alpha$ to a subset $\tilde{W}_\beta \subset \hat{W}_\beta$. Define

$$\begin{aligned}\tilde{W}_\alpha &= \{(x,t) \in \hat{W}_\alpha \,:\, (\mathcal{F}_{s_0}x, \mathbb{A}_{\alpha,s_0,i}t) \in \hat{W}'_{\alpha,s_0,i} \text{ for some } i\}, \\ \tilde{W}_\beta &= \{(y,s) \in \hat{W}_\beta \,:\, (\mathcal{F}_{s_0}y, \mathbb{A}_{\beta,s_0,i}s) \in \hat{W}'_{\beta,s_0,i} \text{ for some } i\}.\end{aligned}$$

The bijective map $\Theta : \tilde{W}_\alpha \to \tilde{W}_\beta$ is defined for a point $(x,t) \in \tilde{W}_\alpha$ such that $(\mathcal{F}_{s_0}x, \mathbb{A}_{\alpha,s_0,i}t) \in \hat{W}'_{\alpha,s_0,i}$ by the rule

$$(y,s) = \Theta(x,t) \iff (\mathcal{F}_{s_0}y, \mathbb{A}_{\beta,s_0,i}s) = \Theta'_{s_0,i}(\mathcal{F}_{s_0}x, \mathbb{A}_{\alpha,s_0,i}t).$$

Because the affine maps and the couplings $\Theta'_{s_0,i}$ are measure preserving, also Θ is measure preserving; the pushforward of the measure $\hat{\mu}_{\mathcal{G}}|_{\tilde{W}_\alpha}$ under Θ is $\hat{\mu}_{\mathcal{E}}|_{\tilde{W}_\beta}$. In particular, the amount of coupled mass equals $\hat{\mu}_{\mathcal{G}}(\tilde{W}_\alpha) = \hat{\mu}_{\mathcal{E}}(\tilde{W}_\beta) = d_0/2$. Finally, the coupling time function $\Upsilon : \tilde{W}_\alpha \to \mathbb{N}$ is defined by

$$\Upsilon(x,t) = s_0.$$

We now have a complete description of how to couple $d_0/2$ units of mass of *any* two standard pairs. Thus, given two standard families $\mathcal{G} = \{(W_\alpha, \nu_\alpha)\}_{\alpha\in\mathfrak{A}}$ and $\mathcal{E} = \{(W_\beta, \nu_\beta)\}_{\beta\in\mathfrak{B}}$, we can couple a subset $\tilde{W}_\alpha \subset \hat{W}_\alpha$ with a subset $\tilde{W}_\beta \subset \hat{W}_\beta$ for any pair $(\alpha,\beta) \in \mathfrak{A}\times\mathfrak{B}$, in which case we have $\nu_\alpha(\tilde{W}_\alpha) = \nu_\beta(\tilde{W}_\beta) = d_0/2$. Recall that the index sets $\mathfrak{A}$ and $\mathfrak{B}$ carry probability factor measures $\lambda_{\mathcal{G}}$ and $\lambda_{\mathcal{E}}$, respectively. They detail how much weight is assigned to standard pairs. Suppose the sets $\mathfrak{A}$ and $\mathfrak{B}$ are finite or countable. Splitting off subrectangles if necessary and stretching them affinely onto complete rectangles as described earlier, we can assume that there exists a bijection $\Delta : \mathfrak{A} \to \mathfrak{B}$ that preserves measure, *i.e.*, $\Delta\lambda_{\mathcal{G}} = \lambda_{\mathcal{E}}$. Hence, the coupling map Θ can be constructed from a subset $\cup_{\alpha\in\mathfrak{A}}\tilde{W}_\alpha \subset \cup_{\alpha\in\mathfrak{A}}\hat{W}_\alpha$ to a subset $\cup_{\beta\in\mathfrak{B}}\tilde{W}_\beta \subset \cup_{\beta\in\mathfrak{B}}\hat{W}_\beta$ so that measure is preserved and in particular so that $\hat{\mu}_{\mathcal{G}}(\cup_{\alpha\in\mathfrak{A}}\tilde{W}_\alpha) = \hat{\mu}_{\mathcal{E}}(\cup_{\beta\in\mathfrak{B}}\tilde{W}_\beta) = d_0/2$. On $\cup_{\alpha\in\mathfrak{A}}\tilde{W}_\alpha$ we set $\Upsilon = s_0$. The map Θ can be constructed also for uncountable families, but we omit the details [7].

4.5. **Recovery step.** Our task is to couple the remaining points of $\hat{W}_\alpha \setminus \tilde{W}_\alpha$ to those of $\hat{W}_\beta \setminus \tilde{W}_\beta$ and to extend Θ and Υ to all of $\hat{W}_\alpha$. We do this recursively. But first we need to prepare the uncoupled parts of the images $\mathcal{F}_{s_0}\hat{W}_\alpha$ and $\mathcal{F}_{s_0}\hat{W}_\beta$ so that they also can undergo the coupling procedure described above.

On the one hand, we have the sets $\mathcal{F}_{s_0}\hat{W}_\alpha \setminus \cup_i \hat{W}_{\alpha,s_0,i}$ and $\mathcal{F}_{s_0}\hat{W}_\beta \setminus \cup_i \hat{W}_{\beta,s_0,i}$ which consist of several rectangles whose base curves do *not* represent proper crossings of the magnet $\mathfrak{M}_{s_0}$. In the worst case, such a base curve is the excess piece of a longer curve that has crossed the magnet $\mathfrak{M}_{s_0}$ properly, but by definition such excess pieces have a uniform lower bound on their length.

On the other hand, we also have the sets $\hat{W}_{\alpha,s_0,i} \setminus \hat{W}'_{\alpha,s_0,i} = \{(x,t) \in \hat{W}_{\alpha,s_0,i} : \tau_\alpha < t \le 1\}$ and $\hat{W}_{\beta,s_0,i} \setminus \hat{W}'_{\beta,s_0,i} = \{(y,s) \in \hat{W}_{\beta,s_0,i} : \tau_{\beta,i}(y) < s \le 1\}$ whose bottom complements were coupled already. We stretch each $\hat{W}_{\alpha,s_0,i} \setminus \hat{W}'_{\alpha,s_0,i}$ affinely onto the complete rectangle $\hat{W}_{\alpha,s_0,i}$, replacing the density ρ_α on it by $(1-\tau_\alpha)\rho_\alpha$. Similarly, we stretch each $\hat{W}_{\beta,s_0,i} \setminus \hat{W}'_{\beta,s_0,i}$ onto $\hat{W}_{\beta,s_0,i}$ so that each vertical fiber $\{s : (y,s) \in \hat{W}_{\beta,s_0,i},\ \tau_{\beta,i}(y) < s \le 1\}$ is mapped affinely onto $[0,1]$, and replace the density ρ_β by the density $(1-\tau_{\beta,i})\rho_\beta$. These transformations are measure preserving; the pushforward of the original measure on the incomplete rectangle is precisely the new measure on the complete rectangle. Notice that the base curves $W_{\bullet,s_0,i}$ are quite short — of the size of the magnet — but nevertheless have a uniform lower bound on their length.

In conclusion, a finite *recovery time* r_0 will be sufficient for the map $\mathcal{T}_{s_0+r_0,s_0+1}$ to stretch the base curves of all the remaining rectangles above to standard length. In fact, some may grow too long but can then be standardized by cutting into shorter pieces, as has been discussed earlier.

We still need to address the issue of regularity of $(1-\tau_\alpha)\rho_\alpha$ and $(1-\tau_{\beta,i})\rho_\beta$ as well as show that $\tau_{\beta,i}$ is actually well defined, *i.e.*, that its values do not exceed 1.

Lemma 22. *Let us take $\eta_\mathrm{r} = \eta_\mathbf{h}$* [6]*. Then $\sup \tau_{\beta,i} \le 1$. Taking the recovery time r_0 sufficiently long, the densities $(1-\tau_\alpha)\rho_\alpha$ and $(1-\tau_{\beta,i})\rho_\beta$ become regular under $\mathcal{T}_{s_0+r_0,s_0+1}$.*

Proof. Throughout the proof we assume that s_0 is sufficiently large to begin with.

By (the proof of) Lemma 9, ρ_α and ρ_β are regular densities on the curves $W_{\alpha,s_0,i}$ and $W_{\beta,s_0,i}$, respectively. Since multiplication by a *constant* preserves the regularity of a density, $(1-\tau_\alpha)\rho_\alpha$ is regular. Showing that $(1-\tau_{\beta,i}(y))\rho_\beta(y)$ is regular requires some analysis.

Recall that the holonomy map $\mathbf{h}$ maps $W_{\alpha,s_0,i}$ onto $W_{\beta,s_0,i}$ by sliding along the connecting leaves of the stable foliation $\mathcal{W}^{s_0}$. Both curves as well as the leaves are inside the magnet. Assuming that the magnet is sufficiently small, $\mathbf{h}$ is as close to the identity as we wish: given any $\delta > 0$, we may assume that $|\mathcal{J}\mathbf{h} - 1| \le \delta$. This follows immediately from Lemma 11, because it can be applied to the k-step pullback of $\mathbf{h}$ that maps $\mathcal{T}^{-1}_{s_0,s_0-k+1}W_{\alpha,s_0,i}$ onto $\mathcal{T}^{-1}_{s_0,s_0-k+1}W_{\beta,s_0,i}$ with k large and the connecting leaves of $\mathcal{W}^{s_0-k}$ sufficiently short (shorter than ℓ_0) for Lemma 11 to apply.

For each $x \in W_{\alpha,s_0,i}$ denote $y = \mathbf{h}x \in W_{\beta,s_0,i}$. As $|W_{\beta,s_0,i}(y_1,y_2)| = \int_{W_\alpha(x_1,x_2)} \mathcal{J}\mathbf{h}\, dm_{W_{\alpha,s_0,i}}$,

$$(1-\delta)|W_{\alpha,s_0,i}(x_1,x_2)| \le |W_{\beta,s_0,i}(y_1,y_2)| \le (1+\delta)|W_{\alpha,s_0,i}(x_1,x_2)|. \tag{18}$$

Observe that from (4) follows easily

$$e^{-C_\mathrm{r}|W_{\bullet,s_0,i}|^{\eta_\mathrm{r}}} \le \frac{|W_{\bullet,s_0,i}|}{\nu_{\bullet,s_0,i}(W_{\bullet,s_0,i})}\,\rho_\bullet \le e^{C_\mathrm{r}|W_{\bullet,s_0,i}|^{\eta_\mathrm{r}}}$$

[6]This fixes the value of η_r.

or

$$\rho_{\bullet} = \frac{\nu_{\bullet,s_0,i}(W_{\bullet,s_0,i})}{|W_{\bullet,s_0,i}|}(1 + \mathcal{O}(|W_{\bullet,s_0,i}|^{\eta_{\mathrm{r}}})).$$

By making the magnet small, the values of $\rho_{\bullet}$ on $W_{\bullet,s_0,i}$ are thus as close to its average as we wish. By (17),

$$\tau_{\beta,i}(y) \le \frac{\tau_\alpha}{1-\delta}\frac{\rho_\alpha(x)}{\rho_\beta(y)} \le \frac{\tau_\alpha}{1-\delta}\frac{1+\mathcal{O}(|W_{\alpha,s_0,i}|^{\eta_{\mathrm{r}}})}{1+\mathcal{O}(|W_{\beta,s_0,i}|^{\eta_{\mathrm{r}}})}\frac{|W_{\beta,s_0,i}|}{|W_{\alpha,s_0,i}|}\frac{Z_{\alpha,s_0}}{Z_{\beta,s_0}} \le \frac{1}{2}\cdot\frac{1+\delta}{1-\delta}\frac{1+\mathcal{O}(|W_{\alpha,s_0,i}|^{\eta_{\mathrm{r}}})}{1+\mathcal{O}(|W_{\beta,s_0,i}|^{\eta_{\mathrm{r}}})},$$

where we have recalled (15), (16), $Z_{\beta,s_0} = \nu_\beta(W_{\beta,s_0,\star}) > d_0$, and (18). The right-hand side can be made arbitrarily close to $\frac{1}{2}$, so that we can take, say, $\sup \tau_{\beta,i} \le \frac{3}{4}$.

From (17) and Lemma 12 it then follows that

$$\begin{aligned}
|\ln(\tau_{\beta,i}(y_1)\rho_\beta(y_1)) - \ln(\tau_{\beta,i}(y_2)\rho_\beta(y_2))| &\le |\ln\rho_\alpha(x_1) - \ln\rho_\alpha(x_2)| + |\ln\mathcal{J}\mathbf{h}(x_2) - \ln\mathcal{J}\mathbf{h}(x_1)| \\
&\le C_{\mathrm{r}}|W_{\alpha,s_0,i}(x_1,x_2)|^{\eta_{\mathrm{r}}} + C_{\mathbf{h}}|W_{\alpha,s_0,i}(x_1,x_2)|^{\eta_{\mathbf{h}}} \\
&\le (C_{\mathrm{r}} + C_{\mathbf{h}})\,(1-\delta)^{-\min(\eta_{\mathrm{r}},\eta_{\mathbf{h}})}|W_{\beta,s_0,i}(y_1,y_2)|^{\min(\eta_{\mathrm{r}},\eta_{\mathbf{h}})}.
\end{aligned}$$

Hence, $\ln(\tau_{\beta,i}\rho_\beta)$ is Hölder and then so is $\ln\tau_{\beta,i} = \ln(\tau_{\beta,i}\rho_\beta) - \ln\rho_\beta$. Using the estimates

$$\min(a,b)|\ln a - \ln b| \le |a-b| \le \max(a,b)|\ln a - \ln b| \qquad a,b>0$$

obtained from the mean-value theorem,

$$\begin{aligned}
&|\ln(1-\tau_{\beta,i}(y_1)) - \ln(1-\tau_{\beta,i}(y_2))| \\
&\quad\le \frac{|\tau_{\beta,i}(y_1) - \tau_{\beta,i}(y_2)|}{1-\sup\tau_{\beta,i}} \le \frac{\sup\tau_{\beta,i}}{1-\sup\tau_{\beta,i}}|\ln\tau_{\beta,i}(y_1) - \ln\tau_{\beta,i}(y_2)| \\
&\quad\le 3(2C_{\mathrm{r}} + C_{\mathbf{h}})(1-\delta)^{-\min(\eta_{\mathrm{r}},\eta_{\mathbf{h}})}|W_{\beta,s_0,i}(y_1,y_2)|^{\min(\eta_{\mathrm{r}},\eta_{\mathbf{h}})}.
\end{aligned}$$

A similar estimate is obtained for $(1-\tau_{\beta,i})\rho_\beta$. The Hölder constant is too large for the density to be regular, but (the proof of) Lemma 9 guarantees that it will become regular after a finite number, r_0, of time steps. □

Finally, at time s_0+r_0, we normalize the measures on all the rectangles to probability measures thereby modifying the factor measures (see Introduction) associated with the rectangle families. As a result, we have two new standard families that can be coupled just as the original ones.

4.6. **Exponential tail bound.** For the standard families $\mathcal{G}$ and $\mathcal{E}$, the first coupling is constructed at time s_0, when enough mass of each family is on the magnet $\mathfrak{M}_{s_0}$:

$$\hat{\mu}_{\mathcal{G}}(\Upsilon = s_0) = \frac{d_0}{2}.$$

After every coupling, there is a recovery period of r_0 steps, during which curves too short can grow to acceptable (*i.e.*, standard) length and densities get regularized sufficiently. After recovery, another s_0 iterations are required to bring enough mass from each standard family on a magnet for the next coupling to be constructed. At the moment of the $(k+1)$st coupling,

$$\hat{\mu}_{\mathcal{G}}(\Upsilon = k(s_0+r_0)+s_0 \,|\, \Upsilon > (k-1)(s_0+r_0)+s_0) = \frac{d_0}{2}$$

is the fraction of the previously uncoupled mass of each standard family $\mathcal{G}$ and $\mathcal{E}$ which lies on the magnet $\mathfrak{M}_{k(s_0+r_0)+s_0}$ and becomes coupled. Hence,

$$\hat{\mu}_{\mathcal{G}}(\Upsilon = k(s_0+r_0)+s_0) = \frac{d_0}{2}\left(1-\frac{d_0}{2}\right)^k.$$

This finishes the proof of the Coupling Lemma. □

Appendix A. Subspace distance

A natural notion of distance between subspaces $A, B \subset \mathbb{R}^M$ is obtained by comparing orthogonal projections to the subspaces in the operator norm:

$$\operatorname{dist}(A,B) = \|P_A - P_B\|, \tag{19}$$

where $P_\cdot$ are the corresponding orthogonal projections. Notice that

$$P_{A\oplus B} = P_A + P_B, \quad \text{if } A \perp B. \tag{20}$$

We will also measure the distance between 1-dimensional subspaces using the metric

$$\operatorname{dist}'(A,B) = \min_{u\in A,\, v\in B\colon \|u\|=\|v\|=1} \|u-v\| = \sqrt{2}\,(1-\langle A,B\rangle)^{1/2}, \tag{21}$$

where

$$\langle A,B\rangle = \max_{u\in A,\, v\in B\colon \|u\|=\|v\|=1} \langle u,v\rangle.$$

Let $u\in A$ and $v \in B$ be unit vectors such that $\langle u,v\rangle \geq 0$. Then

$$\|P_A v - P_B v\| = 1-\langle u,v\rangle^2 \geq 1-\langle u,v\rangle = \frac{1}{2}\|u-v\|^2$$

Hence,

$$\operatorname{dist}'(A,B) \leq \sqrt{2}\operatorname{dist}(A,B)^{1/2}.$$

Appendix B. Uniform Hölder continuity of the (un)stable distribution

Lemma 23. *For all n, the distributions E^n and F^n (see Section 2) are Hölder continuous with the same parameters, and the latter do not depend on the choice of the sequence (T_i).*

Before giving the proof, we need two auxiliary lemmas.

Lemma 24. *For $1 \leq q \leq Q$, there exist constants $k_q \in \mathbb{N}$ and $0 < C_q' < 1$ such that the following holds when ε_q is small enough. If each $T_i \in \mathcal{U}_q$ for a fixed q and if $v \in T_x\mathcal{M}\setminus \mathcal{C}^s_{q,x}$, then $\|D_x\mathcal{T}_n v\| \geq C_q'\Lambda_q^n\|v\|$ for all $n\geq 1$ and $D_x\mathcal{T}_n v \in \mathcal{C}^u_{q,\mathcal{T}_n x}$ for all $n \geq k_q$. These statements are uniform in x.*

Proof. If $v \in T_x\mathcal{M}\setminus \mathcal{C}^s_{q,x}$, we have $\|v^u\| > a_q\|v^s\|$. But $v_n^{u,s} \equiv D_x\widetilde{T}_q^n v^{u,s} \in E^{u,s}_{q,\widetilde{T}_q^n x}$, and $\|v_n^u\| \geq C_q\Lambda_q^n\|v^u\|$ and $\|v^s\| \geq C_q\Lambda_q^n\|v_n^s\|$. We have $\|v_n^u\| \geq C_q^2 a_q \Lambda_q^{2n}\|v_n^s\| \geq 2a_q^{-1}\|v_n^s\|$ for $n\geq k_q$, if k_q is sufficiently large. Because $\|D_x\mathcal{T}_{k_q} - D_x\widetilde{T}_q^{k_q}\| \le C\varepsilon_q$, $\|D_x\mathcal{T}_{k_q}v^s\| \ge c\|v^s\|$, and $\|D_x\widetilde{T}_q^{k_q}v^u\| \ge c\|v^u\|$ hold with some $C = C(k_q)$ and $c = c(k_q)$, we have

$$\begin{aligned}\|D_x\mathcal{T}_{k_q}v^u\| &\ge \|D_x\widetilde{T}_q^{k_q}v^u\| - C\varepsilon_q\|v^u\| \ge \|D_x\widetilde{T}_q^{k_q}v^u\|(1 - Cc^{-1}\varepsilon_q)\\ &\ge 2a_q^{-1}\|D_x\widetilde{T}_q^{k_q}v^s\|(1 - Cc^{-1}\varepsilon_q)\\ &\ge 2a_q^{-1}(\|D_x\mathcal{T}_{k_q}v^s\| - C\varepsilon_q\|v^s\|)(1 - Cc^{-1}\varepsilon_q)\\ &\ge 2a_q^{-1}\|D_x\mathcal{T}_{k_q}v^s\|(1 - Cc^{-1}\varepsilon_q)^2 \ge a_q^{-1}\|D_x\mathcal{T}_{k_q}v^s\|,\end{aligned}$$

provided ε_q is small enough. This estimate shows that $D_x\mathcal{T}_{k_q}v \in \mathcal{C}^u_{q,\mathcal{T}_{k_q}x}$.

The uniform estimate $\|D_x\mathcal{T}_n v\| \ge c_q\|v\|$ holds with some $c_q = c_q(k_q) < 1$ for $1 \le n < k_q$. If $n = k_q + m$, $\|D_x\mathcal{T}_n v\| \ge C_q\Lambda_q^m\|D_x\mathcal{T}_{k_q}v\| \ge c_qC_q\Lambda_q^m\|v\|$. Hence, we can set $C_q' = c_qC_q/\Lambda_q^{k_q}$, so that $\|D_x\mathcal{T}_n v\| \ge C_q'\Lambda_q^n\|v\|$ for all $n \ge 1$. □

The next result on linear maps is cited from [6, Lemma 6.1.1] up to notational changes.

Lemma 25. *Let $\mathcal{L}_n^1$ and $\mathcal{L}_n^2$, $n \in \mathbb{N}$, be two sequences of linear maps $\mathbb{R}^M \to \mathbb{R}^M$. Assume that for some $b > 0$ and $\delta \in (0,1)$,*

$$\|\mathcal{L}_n^1 - \mathcal{L}_n^2\| \le \delta b^n, \quad n \ge 0.$$

Suppose there are two subspaces $\mathcal{E}^1$ and $\mathcal{E}^2$ of $\mathbb{R}^M$ and constants $C_\star > 1$, $0 < \lambda_\star < \mu_\star$ with $\lambda_\star < b$ such that

$$\begin{cases}\|\mathcal{L}_n^i v\| \le C_\star\lambda_\star^n & \text{if } v \in \mathcal{E}^i,\\ \|\mathcal{L}_n^i w\| \ge C_\star^{-1}\mu_\star^n & \text{if } w \perp \mathcal{E}^i.\end{cases}$$

Then

$$\operatorname{dist}(\mathcal{E}^1, \mathcal{E}^2) \le 3C_\star^2\frac{\mu_\star}{\lambda_\star}\delta^{(\ln\mu_\star - \ln\lambda_\star)/(\ln b - \ln\lambda_\star)}.$$

Proof of Lemma 23. We generalize the case of a single map found in [6]. As the orthogonal complements $(T_x\mathcal{M})^\perp$ of $T_x\mathcal{M}$ in $\mathbb{R}^M$ form a smooth distribution $(T\mathcal{M})^\perp$ on $\mathcal{M}$, we first prove that the distribution $E^0 \oplus (T\mathcal{M})^\perp$ is Hölder continuous on $\mathcal{M}$ and then deduce the Hölder continuity of E^0.

Let P_x be the orthogonal projection $\mathbb{R}^M \to T_x\mathcal{M}$. It depends smoothly on x. We define

$$L^{(i)}(x) = D_xT_i \circ P_x.$$

This map extends D_xT_i to a linear map $\mathbb{R}^M \to \mathbb{R}^M$. Let us also set

$$\mathcal{L}_n(x) = L^{(n)}(\mathcal{T}_{n-1}x)\cdots L^{(1)}(x).$$

Recall from above that $E_x^0 \subset \mathcal{C}^s_{1,x}$. Setting $C = \prod_{1\le q\le Q} C_q$ and $\Lambda = \min_{1\le q\le Q}\Lambda_q$, we have

$$\|D_x\mathcal{T}_n v\| \le C^{-1}\Lambda^{-n}\|v\|, \qquad v \in E_x^0,$$

by (A2). This translates to

$$\|\mathcal{L}_n(x)v\| \le C^{-1}\Lambda^{-n}\|v\|, \qquad v \in E_x^0 \oplus (T_x\mathcal{M})^\perp. \tag{22}$$

By Assumption (A4), we can make the cones so narrow that a vector in $T_x\mathcal{M}$ perpendicular to E_x^0 lies in the *complement* of $\mathcal{C}^s_{1,x}$. Setting $C' = \prod_{1\le q\le Q} C_q'$, Lemma 24 thus yields

$$\|D_x\mathcal{T}_n w\| \ge C'\Lambda^n\|w\|, \qquad w \in T_x\mathcal{M} : w \perp E_x^0.$$

In other words,

$$\|\mathcal{L}_n(x)w\| \ge C'\Lambda^n\|w\|, \qquad w \perp E_x^0 \oplus (T_x\mathcal{M})^\perp. \tag{23}$$

Clearly $\|L^{(i)}(x)\| = \|D_xT_i\|$. For brevity, let us write $b_1 = \sup_T \sup_{x\in\mathcal{M}}\|D_xT\|$ and $b_2 = \sup_T \sup_{x\in\mathcal{M}}\|D_x(D_xT\circ P_x)\|$, where T runs over $\mathcal{U}_1\cup\dots\cup\mathcal{U}_Q$. Since $\mathcal{L}_{n+1}(x) = L^{(n+1)}(\mathcal{T}_n x)\mathcal{L}_n(x)$,

$$\begin{aligned}
&\|\mathcal{L}_{n+1}(x) - \mathcal{L}_{n+1}(y)\| \\
&\quad\le \|L^{(n+1)}(\mathcal{T}_n x)\|\,\|\mathcal{L}_n(x) - \mathcal{L}_n(y)\| + \|L^{(n+1)}(\mathcal{T}_n x) - L^{(n+1)}(\mathcal{T}_n y)\|\,\|\mathcal{L}_n(y)\| \\
&\quad\le b_1\|\mathcal{L}_n(x) - \mathcal{L}_n(y)\| + b_2\|\mathcal{T}_n x - \mathcal{T}_n y\|\, b_1^n \le b_1\|\mathcal{L}_n(x) - \mathcal{L}_n(y)\| + b_2 b_1^{2n}\|x-y\|.
\end{aligned}$$

As $\|\mathcal{L}_1(x) - \mathcal{L}_2(y)\| \le b_2\|x-y\|$, we obtain the bound

$$\|\mathcal{L}_n(x) - \mathcal{L}_n(y)\| \le c\|x-y\| b_1^{2n}, \tag{24}$$

with the constant $c = \frac{b_2}{b_1(b_1-1)}$.

The bounds (22), (23), and (24) show that all conditions of Lemma 25 are satisfied, if we take $\mathcal{L}^1_n = \mathcal{L}_n(x)$, $\mathcal{L}^2_n = \mathcal{L}_n(y)$, $\mathcal{E}^1 = E^0_x \oplus (T_x\mathcal{M})^\perp$, and $\mathcal{E}^2 = E^0_y \oplus (T_y\mathcal{M})^\perp$. Writing $\alpha = 2\ln\Lambda/(2\ln b_1 + \ln\Lambda)$ and $K = 3\max(C^{-1}, C'^{-1})^2\Lambda^2 c^\alpha$,

$$\operatorname{dist}\big(E^0_x \oplus (T_x\mathcal{M})^\perp, E^0_y \oplus (T_y\mathcal{M})^\perp\big) \le K\|x-y\|^\alpha$$

provided $\|x-y\| < 1/c$.

By compactness of the manifold and smoothness of the distribution $(T\mathcal{M})^\perp$, we have $\operatorname{dist}((T_x\mathcal{M})^\perp, (T_y\mathcal{M})^\perp) \le L\|x-y\|$ for some L. Hence, by (19) and (20),

$$\operatorname{dist}(E^0_x, E^0_y) \le (K+L)\|x-y\|^\alpha \qquad \text{if } \|x-y\| < 1/c.$$

Notice that this bound does not depend on the sequence $(T_i)_{i\ge1}$. Moreover, the same upper bound is obtained for each distribution E^n by disregarding the first n maps and considering the sequence $(T_i)_{i>n}$ instead. The result for F^n is obtained by reversing time. □

Appendix C. Inclination Lemma type results

Lemma 26. *Fix $1 \le q \le Q$ and let $T_i \in \mathcal{U}_q$ for each i. There exists a constant $C_\# > 0$ such that, for all $w \in \mathcal{C}^u_{q,x}$ and all $\tilde w \in T_x\mathcal{M}$ with $\|w\| = \|\tilde w\| = 1$, the vectors $w_n = D_x\mathcal{T}_n w$ and $\tilde w_n = D_x\mathcal{T}_n\tilde w$ satisfy*

$$\frac{\|\tilde w_n\|}{\|w_n\|} \le C_\# \qquad \forall\, n \ge 0.$$

If also $\tilde w \in \mathcal{C}^u_{q,x}$, the angle between w_n and $\tilde w_n$ tends to zero at a uniform exponential rate:

$$1 - \left|\left\langle \frac{w_n}{\|w_n\|}, \frac{\tilde w_n}{\|\tilde w_n\|}\right\rangle\right| \le \min(c, C\Lambda_q^{-4n}) \tag{25}$$

for some $0 < c < 1$ and $C > 0$.

Proof. We first construct "fake" stable and unstable distributions for finite sequences $T_1, \dots, T_N$. Fix $N \ge 1$. First, choose a distribution $\tilde E^N$ such that $\tilde E^N_x \subset \mathcal{C}^s_{q,x}$ for all x. Then define the distributions $\tilde E^n$, $0 \le n < N$, by pulling back: $\tilde E^n_x = D_{T_{n+1}x}T^{-1}_{n+1}\tilde E^{n+1}_{T_{n+1}x} \subset \mathcal{C}^s_{q,x}$. Next, choose a distribution $\tilde F^0$ such that $\tilde F^0_x \subset \mathcal{C}^u_{q,x}$ for all x. Then set recursively $\tilde F^n_x = D_{T_n^{-1}x}T_n\,\tilde F^{n-1}_{T_n^{-1}x}$ for $n \ge 1$. We will use $\tilde E^n_x$ and $\tilde F^n_x$ as coordinate axes. Let $\mathfrak{e}^n_x \in \tilde E^n_x$ and $\mathfrak{f}^n_x \in \tilde F^n_x$ be unit vectors oriented so that $D_xT_n\mathfrak{e}^n_x$ and $\mathfrak{e}^{n+1}_{T_nx}$ point in the same direction and $D_xT_n\mathfrak{f}^n_x$ and $\mathfrak{f}^{n+1}_{T_nx}$ point in the same direction.

Recall from Section 1.3 that the angle between $E^u_{q,x}$ and $E^s_{q,x}$ is uniformly bounded away from zero. In other words, there exists a $\psi_q > 0$ such that $\langle E^u_{q,x}, E^s_{q,x}\rangle \le 1 - 2\psi_q$ for all $x \in \mathcal{M}$.

By Assumption (A4), the cones can be assumed narrow enough, so that $\langle U, V\rangle \le 1-\psi_q$ for all subspaces $U \subset \mathcal{C}^u_{q,x}$ and $V \subset \mathcal{C}^s_{q,x}$, and for all $x \in \mathcal{M}$. Because of this, $|\langle \mathfrak{e}^n_x, \mathfrak{f}^n_x\rangle| \le 1-\psi_q$ for all x and all $0 \le n \le N$. Thus, we have the uniform bounds

$$\psi_q(\alpha^2+\beta^2) \le \|\alpha \mathfrak{f}^n_x + \beta \mathfrak{e}^n_x\|^2 \le 2(\alpha^2+\beta^2), \qquad \forall\, \alpha, \beta, \quad 0 \le n \le N. \tag{26}$$

Now write $w_n = \alpha_n \mathfrak{f}^n_{\mathcal{T}_n x} + \beta_n \mathfrak{e}^n_{\mathcal{T}_n x}$ and $\tilde{w}_n = \tilde{\alpha}_n \mathfrak{f}^n_{\mathcal{T}_n x} + \tilde{\beta}_n \mathfrak{e}^n_{\mathcal{T}_n x}$ and estimate, for $0 \le n \le N$,

$$\frac{\|\tilde{w}_n\|^2}{\|w_n\|^2} \le \frac{2(\tilde{\alpha}_n^2+\tilde{\beta}_n^2)}{\psi_q^2 \alpha_n^2} = \frac{2\tilde{\alpha}_0^2}{\psi_q^2\alpha_0^2} + \frac{2\tilde{\beta}_0^2\|D_x\mathcal{T}_n \mathfrak{e}^0_x\|^2}{\psi_q^2\alpha_0^2\|D_x\mathcal{T}_n\mathfrak{f}^0_x\|^2} \le \frac{2\tilde{\alpha}_0^2}{\psi_q^2\alpha_0^2} + \frac{2\tilde{\beta}_0^2(C_q\Lambda_q^n)^{-2}}{\psi_q^2\alpha_0^2(C_q\Lambda_q^n)^2}.$$

We have used Assumption (A2) to bound the norms involving $D_x\mathcal{T}_n$. From (26), $\tilde{\alpha}_0^2, \tilde{\beta}_0^2 \le \psi_q^{-1}$. As w is an unstable unit vector, $|\alpha_0|$ is bounded from below by some $A_q > 0$. Thus

$$\frac{\|\tilde{w}_n\|^2}{\|w_n\|^2} \le \frac{2}{\psi_q^3 A_q^2}(1+C_q^{-4}) \equiv C_\#^2 \qquad 0 \le n \le N.$$

But N was arbitrary, so the bound holds for all $n \ge 0$. In particular, $C_\#$ does not depend on the constructed distributions. A computation also shows that, if both $\tilde{w}, w \in \mathcal{C}^u_{q,x}$, then $1 - |\langle w_n, \tilde{w}_n\rangle|/\|w_n\|\|\tilde{w}_n\|$ is of order $\beta_n\tilde{\beta}_n/\alpha_n\tilde{\alpha}_n \le C_q^{-4}\Lambda_q^{-4n}\beta_0\tilde{\beta}_0/\alpha_0\tilde{\alpha}_0 \le C\Lambda_q^{-4n}$. ◻

Lemma 27. *Fix $1 \le q \le Q$, $0 < \lambda < 1$, and $\delta > 0$. Fix $N \ge 1$ and $T_i \in \mathcal{U}_q$ for $1 \le i \le N$. Take two points x_1, x_2 and assume that $d(\mathcal{T}_n x_1, \mathcal{T}_n x_2) < \delta\lambda^n$ if $0 \le n \le N$. Suppose W_1, W_2 are unstable curves with respect to $\{\mathcal{C}^u_{q,x}\}$ and that $x_i \in W_i$. Then*

$$\left|\frac{\mathcal{J}_{\mathcal{T}_n W_1} T_{n+1}(\mathcal{T}_n x_1)}{\mathcal{J}_{\mathcal{T}_n W_2} T_{n+1}(\mathcal{T}_n x_2)} - 1\right| \le C'\mu^n \qquad 0 \le n \le N-1. \tag{27}$$

The constants $C' > 1$ and $0 < \mu < 1$ are independent of N, of the curves W_i, and of the choice of $T_1, \dots, T_N$, as long as the bound on $d(\mathcal{T}_n x_1, \mathcal{T}_n x_2)$ continues to hold.

Proof. Choose $\tilde{F}^0_x = E^u_{q,x}$ and define $\tilde{F}^n_x = D_{T_n^{-1}x}T_n\, \tilde{F}^{n-1}_{T_n^{-1}x}$ for $1 \le n \le N$. By Lemma 23 [7], the distributions $\tilde{F}^n$ belong to a fixed Hölder class, no matter which $T_1, \dots, T_N$ and N are chosen. Notice that $\tilde{F}^n_x \subset \mathcal{C}^u_{q,x}$ for $0 \le n \le N$.

Let u_i^n $(i = 1, 2)$ stand for a unit tangent vector of $\mathcal{T}_n W_i$ at $x_i^n = \mathcal{T}_n x_i$. We also write $L(x) = D_x T_{n+1} \circ P_x$, where P_x is the orthogonal projection $\mathbb{R}^M \to T_x\mathcal{M}$, which is smooth.

$$\begin{aligned}\left|\frac{\mathcal{J}_{\mathcal{T}_n W_1} T_{n+1}(x_1^n)}{\mathcal{J}_{\mathcal{T}_n W_2} T_{n+1}(x_2^n)} - 1\right| &\le \frac{1}{C_q}\left|\|D_{x_1^n}T_{n+1}u_1^n\| - \|D_{x_2^n}T_{n+1}u_2^n\|\right| = \frac{1}{C_q}\left|\|L(x_1^n)u_1^n\| - \|L(x_2^n)u_2^n\|\right| \\ &\le \frac{1}{C_q}\left(\|L(x_1^n) - L(x_2^n)\| + \|L(x_2^n)\| \min_{\sigma=\pm1}\|u_1^n - \sigma u_2^n\|\right) \\ &\le C\left(\lambda^n + \operatorname{dist}'(U_1^n, U_2^n)\right).\end{aligned}$$

We have denoted by U_i^n the linear subspaces of $\mathbb{R}^M$ spanned by u_i^n and recalled the definition in (21). The angle between U_i^n and $\tilde{F}^n_{x_i^n}$ decays exponentially: $\operatorname{dist}'(U_i^n, \tilde{F}^n_{x_i^n}) \le C\Lambda_q^{-2n}$ due to (21) and (25). Hence, it suffices to prove exponential decay of $\operatorname{dist}'(\tilde{F}^n_{x_1^n}, \tilde{F}^n_{x_2^n})$. But this follows from Hölder continuity and the assumption $d(x_1^n, x_2^n) < \delta\lambda^n$. ◻

[7] Lemma 23 has been formulated for F^n as defined in (6). Considering the special case $Q = 1$, we can clearly recover the claimed result for $\tilde{F}^n$ as defined here.

References

[1] Pierre Arnoux and Albert M. Fisher. Anosov families, renormalization and non-stationary subshifts. *Ergodic Theory Dynam. Systems*, 25(3):661–709, 2005.

[2] Arvind Ayyer and Mikko Stenlund. Exponential decay of correlations for randomly chosen hyperbolic toral automorphisms. *Chaos*, 17(4):043116, 7, 2007.

[3] Victor I. Bakhtin. Random processes generated by a hyperbolic sequence of mappings. I. *Izv. Ross. Akad. Nauk Ser. Mat.*, 58(2):40–72, 1994.

[4] Victor I. Bakhtin. Random processes generated by a hyperbolic sequence of mappings. II. *Izv. Ross. Akad. Nauk Ser. Mat.*, 58(3):184–195, 1994.

[5] Xavier Bressaud and Carlangelo Liverani. Anosov diffeomorphisms and coupling. *Ergodic Theory Dynam. Systems*, 22(1):129–152, 2002.

[6] Michael Brin and Garrett Stuck. *Introduction to dynamical systems*. Cambridge University Press, Cambridge, 2002.

[7] Nikolai Chernov. Advanced statistical properties of dispersing billiards. *J. Stat. Phys.*, 122(6):1061–1094, 2006.

[8] Nikolai Chernov and Dmitry Dolgopyat. Brownian Brownian motion. I. *Mem. Amer. Math. Soc.*, 198(927):viii+193, 2009.

[9] Jean-Pierre Conze and Albert Raugi. Limit theorems for sequential expanding dynamical systems on $[0,1]$. In *Ergodic theory and related fields*, volume 430 of *Contemp. Math.*, pages 89–121. Amer. Math. Soc., Providence, RI, 2007.

[10] Anatole Katok and Boris Hasselblatt. *Introduction to the modern theory of dynamical systems*, volume 54 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 1995. With a supplementary chapter by Katok and Leonardo Mendoza.

[11] Nader Masmoudi and Lai-Sang Young. Ergodic theory of infinite dimensional systems with applications to dissipative parabolic PDEs. *Comm. Math. Phys.*, 227(3):461–481, 2002.

[12] William Ott, Mikko Stenlund, and Lai-Sang Young. Memory loss for time-dependent dynamical systems. *Math. Res. Lett.*, 16(3):463–475, 2009.

[13] Lai-Sang Young. Recurrence times and rates of mixing. *Israel J. Math.*, 110:153–188, 1999.

(Mikko Stenlund) Courant Institute of Mathematical Sciences, New York, NY 10012, USA; Department of Mathematics and Statistics, P.O. Box 68, Fin-00014 University of Helsinki, Finland.

E-mail address: `mikko@cims.nyu.edu`

URL: `http://www.math.helsinki.fi/mathphys/mikko.html`