Non-stationary compositions of Anosov diffeomorphisms

Mikko Stenlund
Courant Institute of Mathematical Sciences, New York, NY 10012, USA
and
Department of Mathematics and Statistics, PO Box 68, FIN-00014 University of Helsinki, Finland
E-mail: mikko.stenlund@helsinki.fi

Received 2 February 2011, in final form 19 June 2011
Published 23 September 2011
Online at stacks.iop.org/Non/24/2991

Recommended by L Bunimovich

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 two-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 this kind loses memory of its statistical history rapidly.

Mathematics Subject Classification: 37D20; 60F05, 82C05

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 \( n \geq 0 \), we consider the state at time \( n + 1 \) to be either (i) \( x_{n+1} = T x_n \), where \( T : \mathcal{M} \to \mathcal{M} \) is an \textit{a priori} specified map used at every time step or (ii) \( x_{n+1} = T_{i+1} x_n \), where the maps \( T_i : \mathcal{M} \to \mathcal{M} \) are drawn randomly and independently of each other and of \( x_0, \ldots, x_{n-1} \) from a set of maps \( \mathfrak{T} \) according to a distribution \( \eta \). Now, suppose the initial point \( x_0 \) has random distribution \( \mu \): 
\[ \text{Prob}(x_0 \in E) = \mu(E), \] 
for all measurable sets \( E \subseteq \mathcal{M} \). By stationarity we mean that 
\[ \text{Prob}(x_n \in E) = \mu(E), \] 
for all \( n \geq 0 \), for all measurable sets \( E \subseteq \mathcal{M} \). This condition
translates to \(\mu(T^{-1}E) = \mu(E)\) in case (i) and to \(\int_{\mathcal{E}} \mu(T^{-1}E) \, d\eta(T) = \mu(E)\) in case (ii). In either case the measure \(\mu\) is called invariant. The reader may verify that these definitions result, indeed, in a strictly stationary process in which all finite dimensional distributions are shift-invariant: \(\text{Prob}(x_{i_1+n} \in E_1, \ldots, x_{i_k+n} \in E_m) = \text{Prob}(x_{i_1} \in E_1, \ldots, x_{i_k} \in E_m)\), 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 of random variables.

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

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 modelled 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 \(T_n \circ \ldots \circ T_1\) 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 two-dimensional compact Riemannian manifold \(\mathcal{M}\), which form a prime class of non-trivial 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 \(T_n \circ \ldots \circ T_1 = \text{id}\) 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 \textit{non-stationary} and think of them as descriptions of \textit{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 \(\mu^1\) or \(\mu^2\) and call \(\mu_n^1\) and \(\mu_n^2\), respectively, the corresponding distributions of \(x_n\). Our main result states that if \(\mu^1\) and \(\mu^2\) are sufficiently regular, then the difference \(\int f \, d\mu_n^1 - \int f \, d\mu_n^2\) 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 $\mu_n^1$ and $\mu_n^2$ 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 $\mu_n^1$ and $\mu_n^2$ unless one first integrates against a test function that possesses some regularity along stable manifolds. In spite of the convergence of the difference $\mu_n^1 - \mu_n^2$ for arbitrary initial measures $\mu^1$ and $\mu^2$, in general the limit measures $\lim_{n \to \infty} \mu_n^i$ 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 this 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 this paper concerns quite a different type of question. 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 $1 \leq q \leq Q$, let $\widehat{T}_q : \mathcal{M} \to \mathcal{M}$ be a topologically transitive $C^2$ Anosov diffeomorphism on the two-dimensional compact Riemannian manifold $\mathcal{M}$ with metric $d$ embedded in an ambient space $\mathbb{R}^n$. The Riemannian volume is denoted by $m$. The map $\widehat{T}_q$ admits an invariant Sinai–Ruelle–Bowen (SRB) measure, $\mu_q$, which is mixing; see for example [5]. In general, such a measure is not absolutely continuous with respect to the Riemannian volume. Let $U_q = \mathbb{D}(\widehat{T}_q, \varepsilon_q)$ be disk neighbourhoods of small radii $\varepsilon_q > 0$ in the $C^2$ topology. Now, pick a finite sequence $(T_n)$ of Anosov diffeomorphisms such that

$$T_n \in U_q \quad \forall n \in I_q = (n_{q-1}, n_q],$$

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

$$|I_q| = n_q - n_{q-1} \geq N_q,$$

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

$$T_n = T_n \circ \ldots \circ T_1 \quad n \leq n_Q.$$

(3)

The maps $\widehat{T}_q$ 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 neighbourhood $U_q$ before moving on to $U_{q+1}$. We also write $T_m = T_q \circ \ldots \circ T_m$ for $m \leq n$.

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

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

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

for some constants $\alpha_q > 0$ such that

\begin{enumerate}
  \item[(C1)] $D_x \widehat{T}_q^n C^u_{q,x} \subset \{ 0 \} \cup \text{int} C^u_{q,T^n_q x}$ and $D_x \widehat{T}_q^{-n} C^u_{q,x} \subset \{ 0 \} \cup \text{int} C^u_{q,T^{-n}_q x}$ if $n \geq n_q$,
  \item[(C2)] $\| D_x \widehat{T}_q^n v \| \geq \widehat{C}_q \tilde{N}_q^n \| v \|$ if $v \in C^u_{q,x}$ and $\| D_x \widehat{T}_q^{-n} v \| \geq \widehat{C}_q \tilde{N}_q^{-n} \| v \|$ if $v \in C^s_{q,x}$.
\end{enumerate}

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

We make the following standing assumptions:

\begin{itemize}
  \item[(A0)] $p_q = 1$ in condition (C1) above.
  \item[(A1)] $D_x T C^u_{q,x} \subset \{ 0 \} \cup \text{int} C^u_{q,Tx}$ and $D_x T^{-1} C^u_{q,x} \subset \{ 0 \} \cup \text{int} C^u_{q,T^{-1}x}$ for all $T \in U_q$.
  \item[(A2)] There exist constants $0 < C_q < 1$ and $\Lambda_q > 1$ such that, if each $T_q \in U_q$ for a fixed $q$,
    
    $$\| D_x T_q v \| \geq C_q \Lambda_q^n \| v \| \text{ if } v \in C^u_{q,x} \text{ and } \| D_x T_q^{-1} v \| \geq C_q \Lambda_q^n \| v \| \text{ if } v \in C^s_{q,x};$$

  \item[(A3)] $D_x T C^u_{q,x} \subset \{ 0 \} \cup \text{int} C^u_{q,Tx}$ if $T \in U_q$ and $D_x T^{-1} C^u_{q,x} \subset \{ 0 \} \cup \text{int} C^u_{q,T^{-1}x}$ if $T \in U_q$.
  \item[(A4)] The numbers $\alpha_q$ can be assumed small.
\end{itemize}

1 Such a diffeomorphism is topologically conjugate to an automorphism of the torus.
Convention 1. From now on we will assume that \( Q \) reference Anosov diffeomorphisms \( \widetilde{T}_1, \ldots, \widetilde{T}_Q \) 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 \( (T_i)_{i=1}^n \) of any length \( n \), 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 \( \gamma \)-Hölder continuous observable, if

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

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 \( \rho^i \) are strictly positive and \( \eta \)-Hölder continuous on \( \mathcal{M} \). If \( f \) is continuous, then

\[
\left| \int_{\mathcal{M}} f \circ T_n \, d\mu^1 - \int_{\mathcal{M}} f \circ T_n \, d\mu^2 \right| \leq A_f(n), \quad n \leq n_Q,
\]

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 \( \gamma \)-Hölder, then \( A_f(n) = C_\gamma \theta_\gamma^n \) with \( \theta_\gamma = C(\sup \{ f - \inf \} f) + |f|_\gamma \). 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 \( C_\gamma \) depends on the densities \( \rho^i \), 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 \( \int_{\mathcal{M}} f \circ T_n \, d\mu^1 \) 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 \( \rho^i \).

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 \( \mu^i \) 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 \( \{ \mathcal{C}_{\gamma, \epsilon} \} \) 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 \( \theta_\gamma \) is constant. The latter depends on the reference diffeomorphisms \( \widetilde{T}_q, 1 \leq q \leq Q \). 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 complete the section by discussing assumptions \((A)\) and how they could be relaxed.

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

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

Assumption (A3) guarantees that hyperbolicity prevails when a transition from $U_q$ to $U_{q+1}$ occurs. The first part of (A3) could be relaxed by replacing the map $T \in U_{q+1}$ by sufficiently long compositions $T_q = T_n \circ \cdots \circ T_1$ of maps with each $T_i \in U_{q+1}$: given a sufficiently large $r_q > 0$, $D_x T_n C_{q,x}^u \subset [0, \infty) \cup \text{int} C_{q+1,x}^u$, if $n \geq r_q$ and $T_i \in U_{q+1}$ for $1 \leq i \leq n$. The assumption is then satisfied, for example, if $U_q \cap U_{q+1} \neq \emptyset$ and if $\varepsilon_q$ is small, for all $q$. However, $U_q$ and $U_{q+1}$ need not overlap or even be close to each other, as most vectors in the tangent space $T_q \mathcal{M}$ get eventually mapped by $T'_q = T_n \circ \cdots \circ T_1$ into $C_{q+1,x}^u$ if each $T_i \in U_{q+1}$ and if $\varepsilon_{q+1}$ 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 neighbourhood $U_q$ to the next, as long as the number of steps $|I_q|$ spent in each neighbourhood $U_q$, 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 $U_q$ with a sufficiently small $\varepsilon_q$.

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 $T \in U_q$. By the chain rule $D_x T_{q,x} C_{q,x}^u = D_x T_{q,x} (T_{q,x}^{-1}) D_x T_q C_{q,x}^u$, where $D_x T_{q,x} C_{q,x}^u \subset [0, \infty) \cup \text{int} C_{q,x}^u$. By the continuity of the cones with respect to the base point and the fact that $D_{T_{q,x}} (T_{q,x}^{-1}) = 1 + O(\varepsilon_q)^2$, we have $D_x T_{q,x} C_{q,x}^u \subset [0, \infty) \cup \text{int} C_{q,x}^u$, provided $\varepsilon_q$ is sufficiently small. Compactness guarantees that $\varepsilon_q$ can be chosen independently of $x$.

(A2) For each $i$ and $x$, $D_x T_i = D_x T_q + \varepsilon_{i,x}$, where $\sup_{i,x} \|E_{i,x}\| = O(\varepsilon_q)$. We can bound $\|D_x T_N - D_x T_q\| \leq C(N) \varepsilon_q$. If $\varepsilon_q$ is sufficiently small, we have $\|D_x T_N v\| > \|D_x T_q v\| - C(N) \varepsilon_q \|v\| > \|D_x T_q v\| - \frac{1}{2} C_q \Lambda_q \|v\|$ for $v \in C_{q,x}^u$. Now assume $N = N(q)$ is so large that $\frac{1}{2} C_q \Lambda_q > 1$. Here $N$ depends neither on $x$, on $v$, nor on the choice of the maps $T_i \in U_q$. Let us set $\Lambda_q = \left(\frac{1}{2} C_q \Lambda_q^{1/N}\right)^{1/N}$. The uniform estimate $\|D_x T_N v\| \geq c_q \|v\|$ holds with some $c_q = c_q(N)$ for $1 \leq n < N$. Now, assume $n = kN + l$, $0 \leq l < N$. Then $\|D_x T_N v\| \geq c_q \|D_x T_{N-l} v\| \geq c_q \Lambda_q^{l/N} \|v\|$. Thus, we can take $C_q = c_q / \Lambda_q$.

1.4. Unstable curves with smooth measures

We call a smooth curve $W \subset \mathcal{M}$ unstable with respect to $C_{q,x}^u$ if its tangent space at each point $x \in W$ is contained in the unstable cone $C_{q,x}^u$, i.e., $T_x W \subset C_{q,x}^u$. 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

$$|W| = \int_W dm_w.$$  

2 Here it is understood that $\mathcal{M}$ is embedded in the ambient space $\mathbb{R}^M$ and that $D_x T_{q,x} (T_{q,x}^{-1})$ acts between the linear subspaces $T_{q,x}\mathcal{M}$ and $T_{q,x}\mathcal{M}$ of $\mathbb{R}^M$.  

M Stenlund
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 $\ell \leq |W| \leq L$ and if $\kappa(T_nW) \leq K$ for all $n \geq 0$. If $|W| > L$, we can always “standardize” it by cutting it into shorter subcurves. If an unstable curve is of length less than $\ell$, it will eventually grow under the application of the sequence $(T_i)$, such that $|T_nW| \geq \ell$ for sufficiently large $n$. The dynamics also flattens unstable curves, such that $\kappa(T_nW) \leq K$ for all sufficiently large $n$, even if $\kappa(W) > K$. We will confirm these last two facts in the following. Finally, there 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 $(W, v)$ (w.r.t. $(e^u \alpha)$) consists of a standard unstable curve (w.r.t. $(e^u \alpha)$), $W$, and a probability measure, $\nu$, on $W$. The measure $v$ is assumed absolutely continuous with respect to $m_W$ on $W$ with a density, $\rho$, that is regular in the following sense: for some global constants $C_\ell > 0$ and $\eta_\ell \in (0, 1]$ to be fixed later\(^3\),

$$|\ln \rho(x) - \ln \rho(y)| \leq C_\ell |W(x, y)|^{\eta_\ell}$$

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

$$\sup_\rho \rho \leq e^{C_\ell |W|^\eta_\ell},$$

which implies $\inf \rho \geq (1/|W|)e^{-C_\ell |W|^\eta_\ell} \geq (1/L)e^{-C_\ell |W|^\eta_\ell} > 0$, since $\int_W \rho \, dm_W = 1$. Moreover, $\inf \rho \leq (1/|W|)$. If $W' \subset W$, we obtain by using the previous facts that

$$e^{-C_\ell |W'|} \leq |W'| \nu(W') \leq e^{C_\ell |W'|}.$$ 

Hence, if $D = e^{2C_\ell |W'|}$ and $W', W'' \subset W$,

$$D^{-1} \frac{\nu(W')}{|W'|} \leq \frac{\nu(W')}{|W'|} \leq D \frac{\nu(W'')}{|W''|}.$$  

(5)

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

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

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

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

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

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

\(^3\) $C_\ell$ has to satisfy the condition in lemma 9 and $\eta_\ell$ is determined in lemma 22. Both depend on the reference sequence $\tilde{T}_1, \ldots, \tilde{T}_Q$, but not on the choice of $(T_i)$. 

\[\]
1.5. Main contributions

A technical version of our main result is the following theorem. It states that for reasonable initial distributions $\mu_G$ and $\mu_E$, the images $T_n\mu_G$ and $T_n\mu_E$ converge exponentially in a weak sense.

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

$$\left| \int_M f \circ T_n d\mu_G - \int_M f \circ T_n d\mu_E \right| \leq Bf \theta^n, \quad n \leq n_Q,$$

where

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

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. This paper takes advantage of a version of Young’s coupling method introduced by Dolgopyat and Chernov [7, 8].

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 $(T_1, \ldots, T_{n_Q})$ 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 $(T_1, \ldots, T_{n_Q})$ (as long as $Q \geq 1$ and the maps $\bar{T}_q, 1 \leq q \leq Q$, 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.

---

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 with that particular sequence; see section 2.
2. Distortions and holonomy maps

2.1. Stable foliations $\mathcal{W}^u$

As pointed out above, there is no well-defined sequence of stable foliations associated with the finite sequence $(T_1, \ldots, T_{n_0})$ 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 $T_n = \tilde{T}_Q$ for $n > n_0$—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 future and it is not a priori clear whether they have very much to do with the finite-time dynamics ($1 \leq n \leq n_Q$) 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 $E_{Q,x}$ of $\tilde{T}_Q$. More precisely, let us first define $E^n_x = E_{Q,x}^n$ for $n \geq n_Q$ and then

$$E^n_x = D_{T_{n+1}x}T_{n+1}^{-1}E_{T_{n+1}x}^{n+1}, \quad 0 \leq n \leq n_Q.$$ 

With this definition,

$$D_i T_{n,m}E^n_{x}^{-1} = E^n_{T_{n,m}x}, \quad n \geq m \geq 1.$$ 

Assumptions (A1) and (A3) guarantee that $E^n_{x} \subset \{0\} \cup \text{int} C_{u,x}^n$ for $n + 1 \in I_q$ and $1 \leq q \leq Q$. By assumption (A4), the angle between $E^n_{x}$ and $E^n_{u,x}$ can be assumed uniformly small, for $n \in I_q$ and $1 \leq q \leq Q$.

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

$$\mathcal{W}^u_{x} = T_{n+1}^{-1} \mathcal{W}^u_{T_{n+1}x}, \quad 0 \leq n \leq n_Q.$$ 

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

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

$$F^n_x = D_{T_{n+1}x}T_0 F^{n-1}_{T_{n+1}x}, \quad 1 \leq n \leq n_Q.$$ (6)

Assumptions (A1) and (A3) guarantee that $F^n_{x} \subset \{0\} \cup \text{int} C_{u,x}^n$ for $n \in I_q$ and $1 \leq q \leq Q$. By assumption (A4), the angle between $F^n_{x}$ and $E^n_{u,x}$ can be assumed uniformly small, for $n \in I_q$ and $1 \leq q \leq Q$. The distributions $F^n_{x}$ are in fact the unstable distributions of the sequence $(T_i)$ augmented with the past $T_i = \tilde{T}_Q$ for $i \leq 0$. They serve as Hölder continuous reference distributions that allow us to accurately compare different unstable vectors.

In appendix B we show that the distributions $F^n_{x}$ and $E^n_{x}$ 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 $J_W T$ the Jacobian of the restriction of $T$ to $W$. If $v$ is any non-zero tangent vector of $W$ at $x$, then

$$J_W T(x) = \frac{\|D_T v\|}{\|v\|}.$$ 

Lemma 5 (Growth of unstable curves). Fix $1 \leq q \leq Q$ and let $T_i \in I_q$ for each $i$. If $W$ is an unstable curve with respect to $(C_{u,x}^n)$, setting $\Lambda_q = \sup_{x} \|D_0 \tilde{T}_q\| + \varepsilon_q$,

$$C_q \Lambda_q |W| \leq \|T_0 W\| \leq \tilde{\Lambda}_q |W|.$$ (7)
If \( W, T_1 W, \ldots, T_n W \) are stable curves with respect to \([C_{q,x}]\), then
\[
|T_n W| \leq \frac{|W|}{C_q \Lambda_q^n}.
\] (8)

**Proof.** Since \(|T_n W| = \int_{T_n W} \mathbf{x} \cdot \mathbf{n} \, dx = \int_{T_n W} |\mathbf{v}(y)| \, dx = \int_{T_n W} \|D_x(v(y))u(x)\| \, dx\), where \( v \) is a unit vector tangent to \( W \) at \( x \), it suffices to observe that \( C_q \Lambda_q^n \leq \|D_x T_n v_x\| \leq \Lambda_q^n \) in the ‘unstable case’ and \( \|D_x T_n v_x\| \leq (1/C_q \Lambda_q^n) \) in the ‘stable case’. □

**Lemma 6 (Curvature of unstable curves).** Fix \( 1 \leq q \leq Q \) and let \( T_i \in C_q \) for each \( i \). There exist \( K_1 \) and, for any \( K' > 0, K_2(K') \) and \( n_\epsilon(K') \), such that
\[
\kappa(T_n W) \leq \begin{cases} 
K_1, & n \geq n_\epsilon, \\
K_2, & n \geq 0,
\end{cases}
\]
holds if \( W \) is an unstable curve with respect to \([C_{q,x}]\) and \( \kappa(W) \leq K' \). Note that \( K_1 \leq K_2 \) 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 \( \ell \) and \( L \) and curvature \( \kappa(W) \leq K' \) is a standard curve.

**Proof of lemma 6.** Let \( W \) be an unstable curve and \( y \) its parametrization by arc length, such that \( u(x) = \gamma(t) \in C_{q,x} \) with \( x = \gamma(t) \). Note \( \|u\| = 1 \). The curvature of \( W \) at \( x \) is the length of
\[
\dot{\gamma}(t) = u(x) \cdot \nabla u(x).
\]
Setting \( v(y) = V(x)/\|V(x)\| \) is the unit tangent of \( T_n W \) at \( y = T_n x \). The curvature of \( T_n W \) at \( y \) is thus obtained from
\[
v(y) \cdot \nabla v(y) = Dv(y)u(y) = \|V(x)\|^{-1} Dv(y)T_n(x)u(x) = \|V(x)\|^{-1} D_{xv(y)}u(x).
\]
Here the chain rule \( D_x(v(y)) = Dv(y)DT_n(x) \) was used. Now
\[
D_x(v(y))u(x) = D(\|V(x)\|^{-1} V(x))u(x) = \|V(x)\|^{-1} Dv(x)u(x) + V(x)D(\|V(x)\|^{-1})u(x) = \|V(x)\|^{-1} Dv(x)u(x) + V(x)(-\|V(x)\|^{-3} V(x) \cdot Dv(x))u(x) = \|V(x)\|^{-1} Dv(x)u(x) - \|V(x)\|^{-3} v(y) (v(y) \cdot Dv(x))u(x),
\]
such that
\[
v(y) \cdot \nabla v(y) = \|V(x)\|^{-2} (Dv(x)u(x) - v(y) (v(y) \cdot Dv(x)u(x))),
\]
or compactly
\[
v \cdot \nabla v = \|V\|^2 (Dv - v \cdot Dv).
\] (9)

Note that \( Dv - v \cdot Dv \) is the component of \( Dv \) orthogonal to \( v \) and hence \( \|Dv - v \cdot Dv\| \leq \|Dv\| \). Furthermore, as \( Du = u \cdot \nabla u \), which we recognize to be the curvature of \( W \) at \( x \), we have
\[
Dv = D^2 T_n(u, u) + DT_n(u \cdot \nabla u).
\] (10)
Using lemma 26 and \( \|DT_n u\| \geq C_q \Lambda_q^n \|u\| \), we see from (9) and (10) that
\[
\|v \cdot \nabla v\| \leq \|V\|^2 \|Dv\| \leq \frac{\|D^2 T_n(u, u)\| + \|DT_n(u \cdot \nabla u)\|}{\|DT_n u\|^2} + \frac{\|DT_n u\|^2}{\|DT_n u\|^2} \leq (C_q \Lambda_q^n)^{-2} \|D^2 T_n\|_\infty + (C_q \Lambda_q^n)^{-1} C_q \|u \cdot \nabla u\|.
\] (11)
Fix an $N$ such that $(C_q \Lambda_1^N)^{-1} C_q < 1$. Iterating (11),
\[
\kappa(T_{nN} W) \leq \frac{(C_q \Lambda_1^N)^{-2} \sup_T \|D^2 T_H\|}{1 - (C_q \Lambda_1^N)^{-1} C_q} + ((C_q \Lambda_1^N)^{-1} C_q)^k \kappa(T_I W).
\]
A uniform bound $\max_{0 \leq t < N} \kappa(T_I W) \leq a + b \cdot \kappa(W)$ is also obtained, so we are done. \(\square\)

If $W$ carries a measure $\nu$ with density $\rho$, then $T_n W$ carries the measure $T_n \nu$ whose density, which we denote $T_n \rho$, is
\[
(T_n \rho)(x) = \frac{\rho(T_n^{-1} x)}{J_{T_n} T_n^{-1} x} = J_{T_n W} T_n^{-1}(x) \cdot \rho(T_n^{-1} x).
\]
For controlling the regularity of such densities, we have the following result.

**Lemma 8 (Distortion bound).** Fix $1 \leq q \leq Q$ and let $T_i \in U_q$ for each $i$. If $T_n^{-1} W$ is a standard unstable curve with respect to $(\mathcal{C}_q, x)$ for all $0 \leq n \leq N$ and if $x, y \in W$, then
\[
\ln \frac{J_{T_n^{-1} W}(x)}{J_{T_n^{-1} W}(y)} \leq C_{d,q} |W(x, y)|, \quad n \leq N.
\]
Here $C_{d,q} > 0$ is a constant that does not depend on $W$ or the choice of $(T_i)$.

**Proof.** We first prove that the distortion factor of any map $T \in U_q$ is close to that of $T_I$. To this end, let $W$ be an unstable curve with respect to $(\mathcal{C}_q, x)$, $x \in W$, and $v$ a unit vector tangent to $W$ at $x$. Then
\[
\left| J_{T_I} T_I^{-1}(x) - J_{T_I} T_I^{-1}(x) \right| = \left| D_x T_I^{-1} v - D_x T_I^{-1} v \right| \leq \| (D_x T_I^{-1} - D_x T_I^{-1} v) v \|
\]
\[
= \| D_x T_I^{-1}(D_x T_I^{-1} T_I - T_I^{-1}) D_x T_I^{-1} v \|
\]
\[
\leq C \| D_x T_I^{-1} T_I - T_I^{-1} \| \| J_{T_I} T_I^{-1}(x) \| \leq C_{e_I} J_{T_I} T_I^{-1}(x).
\]

Next, let $y$ parametrize $W(x, y)$ according to arc length. Because
\[
\left| \frac{d}{dt} J_{T_I} T_I^{-1}(y(t)) \right| = \left| \frac{d}{dt} \left( D_{y(t)} T_I^{-1} \dot{y}(t) \right) \right|
\]
\[
\leq \left| \frac{d}{dt} \left( D_{y(t)} T_I^{-1} \dot{y}(t) \right) \right| = \left| D_y^2 T_I^{-1} \| \| \dot{y}(t) \| \| + D_y T_I^{-1} \| \| \dot{y}(t) \| \|
\]
and because the curvature $\| \| \| \dot{y} \| \|$ $\leq K$ for all standard unstable curves,
\[
\ln J_{T_I} T_I^{-1}(x) - \ln J_{T_I} T_I^{-1}(y) \leq \left[ \int_0^{\| \| y(t) \| \|} \frac{d}{dt} \left( \ln J_{T_I} T_I^{-1}(y(t)) \right) \right] \| y(t) \| \| \equiv \tilde{C}_{d,q} |W(x, y)|
\]
where $\tilde{C}_{d,q}$ is independent of the choice of $T$. The desired estimate follows. Indeed, writing $x^{-j} = (T_{n, n-j+1})^{-1} x$, $y^{-j} = (T_{n, n-j+1})^{-1} y$ and $W^{-j} = (T_{n, n-j+1})^{-1} W$ (with $T_{n, n+1} = \text{id}$),
\[
\ln J_{T_n W} T_n^{-1}(x) - \ln J_{T_n W} T_n^{-1}(y) \leq \sum_{j=0}^{n-1} \ln J_{T_n W} T_n^{-1}(x^{-j}) - \ln J_{T_n W} T_n^{-1}(y^{-j}) \leq \sum_{j=0}^{n-1} \tilde{C}_{d,q} |W^{-j}(x^{-j}, y^{-j})|.
\]
Moreover, by (7), $|W^{-j}(x^{-j}, y^{-j})| \leq C_q^{-1} \Lambda_q^{-j} |W(x, y)|$. \(\square\)
2.3. Image of a standard family

\textbf{Lemma 9.} Fix \(1 \leq q \leq Q\) and let \(T_i \in \mathcal{U}_q\) for each \(i\). Let \(\mathcal{G} = (W, \nu)\) be a standard pair with respect to \([C^n_{q,\varepsilon}]\) and assume that \(C_i\) satisfies

\[2C_{d,q}L^{-1-n} \leq C_i.\]

For \(n \geq \ln \frac{2}{\varepsilon} / \ln \Lambda_q\), denote by \(W_i\) the (finitely many) standard pieces of the image \(T_nW\) after it has been standardized by cutting into shorter pieces and split the image measure \(T_n\nu\) into the sum \(\sum c_i \nu_i\), where \(\nu_i\) is a probability measure on \(W_i\) and \(\sum c_i = 1\). Then each \((W_i, \nu_i)\) is a standard pair w.r.t. \([C^n_{q,\varepsilon}]\).

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

\[|\ln \rho_i(x) - \ln \rho_i(y)| \leq |\ln \rho(T_n^{-1}x) - \ln \rho(T_n^{-1}y)| + \frac{\mathcal{J}_W T^{-1}_n(x)}{\mathcal{J}_W T^{-1}_n(y)} \leq C_i |W(T_n^{-1}x, T_n^{-1}y)|^{\eta} + C_{d,q} |W_i(x, y)| \leq (C_i L^{-1-n} + C_{d,q} L^{-1-n}) \ln |W_i(x, y)|^{\eta} \leq C_i |W_i(x, y)|^{\eta}.
\]

We used lemma 8 and also \(W_i(x, y) = T_n(W(T_n^{-1}x, T_n^{-1}y))\) together with lemma 5. \(\square\)

Thus, \(G_n = \{(W_i, \nu_i)\}\) is a standard family equipped with the factor measure \(\lambda_{G_n}(i) = c_i\).

More generally, if \(G\) is a standard family, \(G_n\) 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. \([C^n_{1,\varepsilon}]\)) connected by the stable foliation \(\mathcal{W}\). In other words, for each point \(x \in W_1\) the leaf \(\mathcal{W}^y\) intersects \(W_2\) and conversely for each point \(y \in W_2\) the leaf \(\mathcal{W}^x\) 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 \(h : W_1 \to W_2\) is defined by sliding along the leaf: \(hx = y\).

Since the images \(T_nW_i\) are connected by the stable foliation \(\mathcal{W}\), one can define the holonomy map \(h_n = T_n \circ h \circ T_n^{-1} : T_nW_1 \to T_nW_2\).

\textbf{Remark 10.} Note that if the curves \(W_i\) carry measures \(\nu_i\) that are compatible in the sense that \(\nu_2 = h\nu_1\), then the images \(T_nW_i\) carry compatible measures: \(T_n\nu_2 = h_nT_n\nu_1\). 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 \(h\) is said to be absolutely continuous, if the measure \(h^{-1}m_{W_i}\) is absolutely continuous with respect to the measure \(m_{W_i}\). In this case the Jacobian, which measures distortion under the holonomy map, is defined as the Radon–Nikodym derivative

\[\mathcal{J}(h) = \frac{\text{d}(h^{-1}m_{W_i})}{\text{d}m_{W_i}}.\]

The change of variables formula for integrals is \(dm_{W_i}(y) = \mathcal{J}(h)(x)dm_{W_i}(x)\) with \(y = hx\). For any \(x \in W_1\), we have \(hx = T_n^{-1}h_nT_nx\). It is elementary to check that if \(h_n\) is absolutely continuous, then \(h\) inherits this property via the identity

\[\mathcal{J}(h)(x) = \mathcal{J}(h_nT_n^{-1}h_nT_nx) \cdot \mathcal{J}(h_nT_nx) \cdot \mathcal{J}(h_nT_nx) = \frac{\mathcal{J}_W T_n(x)}{\mathcal{J}_W T_n(hx)} \cdot \mathcal{J}(h_nT_nx).\]

\[\text{Given a Borel set } A \subset W_1, \text{ we have } \int_A m_{W_i} = (h^{-1}m_{W_i})(A) = \int_A (\mathcal{J}(h^{-1}m_{W_i})) dW_i.\]
By reversing the argument, we see that if \( h_m \) is absolutely continuous for some \( m \), then \( h_n \) is absolutely continuous and (12) holds for all values of \( n \geq 0 \).

**Lemma 11 (Absolute continuity of the holonomy map).** Let \( h \) 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 \( (T_i)_i \), such that

\[
|\ln J h_m(T_n x)| \leq c_1 \mu^n
\]

holds for \( x \in W_1 \) and 0 \( \leq n \leq n_Q \). In particular, \( e^{-c_1} \leq J h \leq e^{c_1} \).

As a curiosity, (13) continues to hold for \( n > n_Q \) since \( T_n = \tilde{T}_Q \). In particular, the precise value of \( J h(x) \) could be obtained as the limit \( \lim_{n \to \infty} \frac{J h(T_n x)}{J h(T_n h x)} \). However, we only care about \( n \leq n_Q \). What is important above is that \( c_1 \) and \( \mu \) 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 \( x^n = T_n x \) and \( y = h x \) for \( x \in W_1 \), and \( W^n = T_n W_i \) for \( i = 1, 2 \). We also write \( y^n = T_n y = h_n x^n \). Since \( T_n = \tilde{T}_Q \) and \( V^{n-1} = V^{Q} \), for all \( x \), for all \( n > n_Q \), we know the following: \( h \) inherits absolute continuity from \( h_n \), (12) holds for all \( n \geq 0 \) as explained above, and \( \lim_{n \to \infty} J h_n(x^n) = 1 \). Therefore, with the aid of the chain rule \( J W T_n(x) = J W^{n-1} T_n(x^{n-1}) \cdots J W^{n} T_1(x^1) \), we conclude that

\[
J h_m(x^n) = \prod_{n \geq m} J W^n T_{n+1}(y^n).
\]

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):

\[
|\ln J h_m(x^n)| \leq \sum_{n \geq m} |\ln J W^n T_{n+1}(y^n)| \leq C \frac{\mu^m}{1 - \mu} = c_1 \mu^m.
\]

\( \square \)

**Lemma 12 (Regularity of the holonomy map).** There exist \( 0 < \eta_h < 1 \) and \( C_h > 0 \), such that the following holds. Let \( W_1 \) and \( W_2 \) be standard unstable curves connected by the stable foliation \( V^0 \) as above. For \( x_1, x_2 \in W_1 \) such that \( |W_i(x_1, x_2)| \leq 1 \),

\[
|\ln J h(x_1) - \ln J h(x_2)| \leq C_h |W_i(x_1, x_2)|^m.
\]

**Proof.** Denote \( x^n = T_n x \) for all \( x \) in \( W^n = T_n W_1 \). We also set \( y_i = h x_i \), \( y_i^n = T_n y_i \), and \( W^n = T_n W_i \). By (12),

\[
|\ln J h(x_1) - \ln J h(x_2)| \leq |\ln J W_1 T_m(x_1)| + |\ln J W_1 T_m(y_1)| + \sum_{i=1,2} |\ln J h_m(x^n)|,
\]

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

\[
|\ln J W_1 T_{nQ}(x_1)| \leq \sum_{0 \leq q \leq Q-1} |\ln J W_1 T_{nQ+1}(x_1^n)| \leq \sum_{0 \leq q \leq Q-1} C_d q + 1 |T_{nQ+1}(W_1^{nQ}(x_1^n, x_2^n))| \leq \sum_{0 \leq q \leq Q-1} C_d q + 1 \eta_0 |W_1^{nQ}(x_1^n, x_2^n)|.
\]
Similarly, for any m,

\[ \left| \ln \frac{\mathcal{J}_W T_m(x_1)}{\mathcal{J}_W T_m(x_2)} \right| \leq C \tilde{A}^m |W_1(x_1, x_2)| \quad \text{and} \quad \left| \ln \frac{\mathcal{J}_W T_m(y_1)}{\mathcal{J}_W T_m(y_2)} \right| \leq C \tilde{A}^m |W_2(y_1, y_2)|. \]

Note from the definition of C and \( \tilde{A} \) that they are independent of the curves \( W \) and of the sequence \( (T_i) \). Because \( |W_2(y_1, y_2)| \leq \sup_{W(x_1, x_2)} \mathcal{J} h \cdot |W(x_1, x_2)| \),

\[ \left| \ln \frac{\mathcal{J}_W T_m(x_1)}{\mathcal{J}_W T_m(x_2)} \right| + \ln \frac{\mathcal{J}_W T_m(x_2)}{\mathcal{J}_W T_m(x_1)} \leq C \tilde{A}^m (1 + e^c) |W_1(x_1, x_2)| = c_2 \tilde{A}^m |W_1(x_1, x_2)|. \]

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

\[ \left| \ln \mathcal{J} h(x_1) - \ln \mathcal{J} h(x_2) \right| \leq 2c_1 \mu^m + c_2 \tilde{A}^m |W_1(x_1, x_2)| \leq (2c_1 + c_2) |W_1(x_1, x_2)|^{1 - \ln \tilde{A} / \ln \mu}. \]

\[ \square \]

### 3. Coupling lemma and the proof of theorems 4 and 2

Let \((W, \nu)\) 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 \( \nu \). 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 \( \{(x, t) \in \hat{W} : 0 \leq t \leq \tau(x)\} \) of \( \hat{W} \). Introducing the rectangle facilitates bookkeeping.

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

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

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

1. Let \((x, t) \in \hat{W}_\alpha, \alpha \in \mathcal{A}, \) and \( \Theta(x, t) = (y, s) \in \hat{W}_\beta, \beta \in \mathcal{B} \). Then the points \( x \) and \( y \) lie on the same leaf, say \( W \), of the stable foliation \( W^s \). If \( n \geq \Upsilon(x, t), \) then the distance of the points \( T_n x \) and \( T_n y \) along the leaf \( T_n W \) of the stable foliation \( W^s \) satisfies \( |T_n W(T_n x, T_n y)| < \ell_0 \lambda^{n - \Upsilon(x, t)} \). Here \( \ell_0 > 0 \) has been introduced earlier and

\( \lambda = \max_{1 \leq q \leq Q} A_q^{-1} < 1. \)
(2) The exponential tail bound

\[ \hat{\mu}_G(\Upsilon > n) \leq C_\Upsilon \Theta_\Upsilon^n \]  \tag{14}

holds for uniform constants \( C_\Upsilon > 0 \) and \( \Theta_\Upsilon \in (0, 1) \).

**Proof of theorem 4.** We use the coupling between \( G \) and \( E \) given in the coupling lemma:

\[
\int_M f \circ T_n \, d\mu_G - \int_M f \circ T_n \, d\mu_E = \int_{M \times [0,1]} (f \circ T_n)(x, t) \, d\hat{\mu}_G(x, t) - \int_{M \times [0,1]} (f \circ T_n \circ \Theta)(x, t) \, d\hat{\mu}_G(x, t)
\]

\[
= \int_{M \times [0,1]} (f \circ T_n - f \circ T_n \circ \Theta) \, d\hat{\mu}_G
\]

\[
= \int_{\Upsilon \leq n/2} (f \circ T_n - f \circ T_n \circ \Theta) \, d\hat{\mu}_G + \int_{\Upsilon > n/2} (f \circ T_n - f \circ T_n \circ \Theta) \, d\hat{\mu}_G.
\]

By (14),

\[
\left| \int_{\Upsilon > n/2} (f \circ T_n - f \circ T_n \circ \Theta) \, d\hat{\mu}_G \right| \leq C_\Upsilon (\sup f - \inf f) \Theta_\Upsilon^{n/2}.
\]

On the other hand, assume \( \Upsilon(x, t) \leq n/2 \). Then \( \left| (f \circ T_n - f \circ T_n \circ \Theta)(x, t) \right| \leq |f|_p (\ell_0 \lambda^{n-\Upsilon(x, t)})^{r'} \), by the coupling lemma, such that

\[
\left| \int_{\Upsilon \leq n/2} (f \circ T_n - f \circ T_n \circ \Theta) \, d\hat{\mu}_G \right| \leq \ell_0^p |f|_p \lambda^{rn/2}.
\]

Since \( \ell_0 < 1 \), the proof is complete. \( \square \)

**Proof of theorem 2.** First note that both of the measures \( \mu^i \) can be disintegrated using a suitable measurable partition of \( M \) so that we almost obtain two standard families, with the nuisance that the Hölder constants of the logarithms of the conditional measures possibly exceed \( C_l \) 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 \( \gamma \)-Hölder observables the result then follows immediately from theorem 4, with the above waiting time giving the constant \( C_\gamma(\rho^1, \rho^2) = \Theta^{-N} \). If \( f \) is only continuous, we fix an arbitrarily small \( \varepsilon > 0 \) and, by the Stone–Weierstrass theorem, pick a \( \gamma \)-Hölder \( f_\varepsilon \) such that \( \| f - f_\varepsilon \|_\infty < \varepsilon \). Then \[ \left| \int_M f \circ T_n \, d\mu^1 - \int_M f \circ T_n \, d\mu^2 \right| < C_\gamma(\rho^1, \rho^2) B_\varepsilon \Theta^{n} + 2\varepsilon < 3\varepsilon \]

if \( n > \ln(\varepsilon/C_\gamma(\rho^1, \rho^2) B_\varepsilon)/\ln \Theta_\gamma \). \( \square \)

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%, 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 $\Theta$ and the coupling time function $\Upsilon$ 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 $n \in I_q$, $T_n \in U_q$, and the magnet to be used should reflect the structure of the reference diffeomorphism $\tilde{T}_q$. 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 $W_{s,q}$ 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 $T_n = \tilde{T}_Q$ for times $n > n_Q$ although in reality we only consider the compositions $T_n = T_n \circ \ldots \circ T_1$ for $n \leq n_Q$. 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 $(T_1, \ldots, T_{n_Q})$ as long as the reference automorphisms $(\tilde{T}_1, \ldots, \tilde{T}_Q)$ have been chosen and the earlier assumptions on $(T_1, \ldots, T_{n_Q})$ are being respected.

4.2. Magnets and crossings

In this subsection $1 \leq q \leq Q$ is fixed for good. Unstable curves and standard pairs are to be understood as being defined with respect to the cone family $\{C_{q,x}^w\}$ with $q$ fixed.

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

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

(P1) $W$ crosses $R$ completely, i.e., $W \cap R$ 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 $\ell$ 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 $x \in W \cap R$ divides the curve $W_{s,q,x}^w \cap R$ 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, $\{R^k \setminus R : 1 \leq k \leq k_0\}$, such that each standard unstable curve crosses at least one of the rectangles super-properly.

Proof. Every closed standard curve crosses some rectangle $R$ super-properly. Since crossing a rectangle $R$ super-properly is an open condition in the Hausdorff metric, the set $U_R$ of all closed standard curves crossing $R$ super-properly is an open set. The collection formed by all the sets $U_R$ is an open cover of the space of closed standard curves equipped with the
Lemma 15. Fix $n \geq 1$. By taking $\varepsilon_q$ sufficiently small (depending on $n$) the following holds. If $W$ is an unstable curve and $T^n_q W$ crosses $\mathfrak{R}_q$ super-properly, then $T^n_q W$ crosses $\mathfrak{R}_q$ properly, provided each $T_i \in \mathcal{U}_q$.

**Proof.** For any point $x$, we have the bound $d(T_n x, \tilde{T}^n_q x) \leq C(n)\varepsilon_q$. □

We now pick arbitrarily one of the rectangles $\mathcal{R}^k_i$. 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 16. There exist a subrectangle $\mathfrak{B}_q \subset \mathfrak{R}_q$ and a number $s' \geq 1$ such that the following holds. Assume that $W$ is an unstable curve that crosses a rectangle $\mathcal{R}^k_i$ properly, $n \geq s'$, and $\tilde{T}^n_q \mathcal{R}^k_i \cap \mathfrak{R}_q \neq \emptyset$. Every component of $\tilde{T}^n_q \mathcal{R}^k_i \cap \mathfrak{R}_q$ that intersects $\mathfrak{B}_q$ intersects $W^q_{n,i}$ for precisely one value of the index $i \in J$.

In words, each intersection of $\tilde{T}^n_q \mathcal{R}^k_i$ with $\mathfrak{B}_q$ yields a super-proper crossing of $\tilde{T}^n_q W$, 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 $\tilde{T}^n_q$ 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 $\mathcal{W}^q_{n,i} \cap \mathfrak{R}_q$ in a ratio between $1/5$ and $4/5$. As $\tilde{T}^n_q$ is one-to-one, the components of $\tilde{T}^n_q \mathcal{R}^k_i \cap \mathfrak{R}_q$ 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 $\tilde{T}^n_q W$. The piece $V$ has to extend to a super-proper crossing of $\mathfrak{R}_q$, because $W$ crosses $\mathcal{R}^k_i$ properly and because $n_1$ is large. Thus, $V$ is actually a subcurve of one of the $W^q_{n,i}$. □

Lemma 17. There exist numbers $d'' > 0$ and $s'' \geq 1$ such that if $(W, \nu)$ is a standard pair and $n \geq s''$, then $\nu(\tilde{T}^{-n}(\cup_{i \in J} W^q_{n,i})) \geq d''$.

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 $\mathcal{R}^k_i$ properly. This is possible by lemma 14. By lemma 16, if a component of $\tilde{T}^n_q \mathcal{R}^k_i \cap \mathfrak{R}_q$ intersects the subrectangle $\mathfrak{B}_q$, it is crossed by the curve component $W^q_{n,i}$, for precisely one value of the index $i \in J$. In this case let $\mathfrak{N}^k_{n,i}$ denote the former component of $\tilde{T}^n_q \mathcal{R}^k_i \cap \mathfrak{R}_q$. Thus $\mathfrak{N}^k_{n,i}$ is only defined for a subset $\mathfrak{B}_q \subset \mathfrak{R}_q$ of indices. We have $\nu(\tilde{T}^{-n}(\cup_{i \in J} W^q_{n,i})) \geq \nu(\tilde{T}^{-n}(\cup_{i \in J} \mathfrak{B}_q)) = \nu(\tilde{T}^{-n}(\cup_{i \in J} \mathfrak{B}_q \cap \mathfrak{R}_q)) = c \mu_q(\mathfrak{B}_q \cap \mathfrak{R}_q) \geq c \mu_q(\mathfrak{B}_q) \geq \frac{1}{2} \mu_q(\mathcal{R}^k_i \cap \mathfrak{B}_q) \geq \frac{1}{2} \mu_q(\mathfrak{R}^k_i) \mu_q(\mathfrak{B}_q)$ if $n \geq s''$ and $s''$ is large. The last step in the estimate follows from mixing of the invariant measure $\mu_q$. The third step relies on the absolute continuity with bounded Jacobians of the holonomy maps of $\tilde{T}^n_q$ as well as on the regularity of $\nu$. 

Hausdorff metric. The latter space is compact. We can therefore pick a finite subcover and the corresponding rectangles.
and of the conditional measures of $\mu_q$ on the unstable leaves of $\tilde{T}_q$. To finish, note that $d'' = \mu_q(\mathcal{N}_q^0)\mu_q(\mathcal{B}_q) > 0$, since the interiors of $\mathcal{N}_q^0$ and $\mathcal{B}_q$ are non-empty.

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

**Corollary 18.** There exist numbers $d_0' > 0$ and $s_0' \geq 1$ such that the following holds. Let $T_i \in \mathcal{U}_q$ for each $i$ and $(W, v)$ be a standard pair. If $n \geq s_0'$, then $v(T_n^{-1}(\bigcup_i W_{n,i})) \geq d_0'$.

**Proof.** Fix $m \geq \ln \frac{\tilde{C}_q}{\Lambda_2}$ and $\Delta_q$. By lemma 9, $(T_m W, T_m v)$ can be broken into a finite collection of standard pairs $(W_j, v_j)$ such that $T_m W = \bigcup W_j$ and $T_m v = \sum_j c_j v_j$, where $0 < c_j < 1$ 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 $\tilde{T}_q^s V_{j,k}$ crosses the magnet $\mathcal{R}_q$ super-properly. Moreover, $\sum_k v_j(V_{j,k}) \geq d''$ and, by lemma 15, the images $T_m s_{m+1} V_{j,k}$ cross the magnet $\mathcal{R}_q$ properly. We also have $T_m v(\bigcup_k V_{j,k}) = \sum_j c_j \sum_k v_j(V_{j,k}) \geq d''$. Let us relabel the collection of the subcurves $T_m s_{m+1} V_{j,k} \subset W_j$ by $U_j$. We have so far shown that $v(\bigcup_j U_j) \geq d''$ and that $T_m s_{m+1} U_j$ crosses the magnet $\mathcal{R}_q$ properly.

We are almost done, but we still need to truncate each $U_j$ to a subcurve $U_j$ so that $T_m s_{m+1} U_j = T_m s_{m+1} U_j \cap \mathcal{R}_q$ and argue that $v(\bigcup_j U_j) \geq \alpha v(\bigcup_j U_j)$ for some constant $\alpha > 0$. Such a truncation amounts to choosing the subcurve $\tilde{V}_{j,k} = T_m \tilde{U}_j$ of $V_{j,k} \subset W_j$ so that $T_m s_{m+1} \tilde{V}_{j,k} = T_m s_{m+1} \tilde{V}_{j,k} \cap \mathcal{R}_q$. Now $v_j(\tilde{V}_{j,k}) \geq \alpha v_j(V_{j,k})$ follows from two observation:

- $|\tilde{V}_{j,k}|$ is bounded uniformly away from zero, because $T_m s_{m+1} \tilde{V}_{j,k}$ crosses $\mathcal{R}_q$ completely.
- $|T_m s_{m+1} \tilde{V}_{j,k} \cap \mathcal{R}_q|$ is bounded uniformly from above, because $|\tilde{T}_q^s V_{j,k} \cap \mathcal{R}_q|$ was assumed to be as small as possible (for a super-properly crossing curve). Therefore $|V_{j,k} \setminus \tilde{V}_{j,k}| = |V_{j,k} \setminus T_m s_{m+1} \tilde{V}_{j,k} \cap \mathcal{R}_q|$ is bounded uniformly from above.

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

Note that only $s''$ affected the size of $\varepsilon_q$. This happened when lemma 15 was used. \hfill \Box

### 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 neighbourhood $\mathcal{U}_q$ of any one map $\tilde{T}_q$.

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 $\mathcal{N}_q^0$ 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 already been coupled will always remain coupled. Therefore, we need to introduce the following time-dependent magnets: For every $q \in \{1, \ldots, Q\}$ and every $n \in I_q$, define

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

In other words, $\mathcal{M}_n$ consists of those leaves of $\mathcal{N}_q^0$ that connect the u-sides of $\mathcal{R}_q$ and are entirely inside $\mathcal{R}_q$.

We say that an unstable curve $W$ crosses $\mathcal{M}_n$ properly if $n \in I_q$ and $W$ crosses $\mathcal{R}_q$ properly. Assuming that the cones $\tilde{C}_q^u, v$ are narrow enough, $\mathcal{M}_n$ is close to $\mathcal{R}_q$, and we have $|W \cap \mathcal{M}_n| \geq \frac{1}{4} |W \cap \mathcal{R}_q|$.
Lemma 19. Assume that \( W \) crosses the magnet \( \mathcal{R}_q \) properly and that \( W \) carries the measure \( dv = \rho \, dm_W \), where \( \rho \) satisfies (4). Then, for all \( n \in I_q \),
\[
\nu(W \cap \mathcal{M}_n) \geq \frac{1}{2} \nu(W \cap \mathcal{R}_q).
\]

Proof. By (4), \( \rho(y) \geq \rho(x) e^{-C_1 |W \cap \mathcal{M}_n|^n} \) for any \( x, y \in W \cap \mathcal{R}_q \). In particular, we can average the left side over \( W \cap \mathcal{M}_n \) and the right side over \( W \cap \mathcal{R}_q \), obtaining \( \nu(W \cap \mathcal{M}_n) \geq \frac{|W \cap \mathcal{R}_q|}{|W \cap \mathcal{M}_n|} e^{-C_1 |W \cap \mathcal{M}_n|^n} \nu(W \cap \mathcal{R}_q) \). Since \( \mathcal{R}_q \) has small diameter, the exponential factor is close to 1. For \( n \in I_q \), we also have \( |W \cap \mathcal{M}_n| \geq \frac{1}{2} |W \cap \mathcal{R}_q| \). \( \square \)

For a standard family with respect to \( \{C_{n,q}^u\} \), \( G = \{(W_a, v_a)\}_{a \in \mathbb{A}} \), let \( W_{a,n,i} \) be the connected components of \( T_n W_a \cap \mathcal{M}_n \) that correspond to proper crossings, and introduce the notation
\[
W_{a,n,*} = T_n^{-1}(\cup_i W_{a,n,i}).
\]

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 \( G = \{(W_a, v_a)\}_{a \in \mathbb{A}} \) a standard family with respect to \( \{C_{n,q}^u\} \), then, for \( 1 \leq q \leq Q \) and \( n_{q-1} + s_0 \leq n \leq n_q \),
\[
\mu_G(\cup_a W_{a,n,*}) = \int_{\mathcal{M}_q} v_a(W_{a,n,*}) \, d\lambda_G(\alpha) \geq d_0.
\]

In other words, there are time windows for \( n \), such that a significant fraction of the image under \( T_n \) of the standard family \( G \) lies on the magnet \( \mathcal{M}_n \), 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 \( (T_k)_{k \geq 1} \) and \( G = \{(W_{a,k}, v_{a,k})\}_{a \in \mathbb{A}} \) a standard family with respect to \( \{C_{n,q}^u\} \). 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 \( q' \in \{q + 1, \ldots, Q\} \), at least a \( d_0 \)-fraction of its image under \( T_{n,k} \) lies on the magnet \( \mathcal{M}_n \) 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 \( G \) under \( T_{n,q-1} \) after standardizing the curves becomes a standard family with respect to \( \{C_{n,q}^u\} \). Applying corollary 18 to this standard family and the map \( T_{n,q-1} \) yields a lower bound on the \( T_n \mu_G \)-measure of proper crossings of \( \mathcal{R}_q \). From this we infer a lower bound on the proper crossings of \( \mathcal{M}_n \) by the regularity of densities with the aid of lemma 19. \( \square \)

4.4. Coupling step

Consider first two standard families, \( G = \{(W_a, v_a)\} \) and \( E = \{(W_b, v_b)\} \), consisting of one standard pair each.

A good fraction of the images \( T_n W_a \) and \( T_n W_b \) cross the magnet \( \mathcal{M}_n \) properly, so that \( T_n v_a(\cup_i W_{a,n,i}) = v_a(W_{a,n,*}) > d_0 \) and \( T_n v_b(\cup_j W_{b,n,j}) = v_b(W_{b,n,*}) > d_0 \). Here \( i \) and \( j \) run through some finite index sets and \( T_n v \) is the pushforward of \( v \).

Recall that, for a curve \( W \), \( \hat{W} \) denotes the rectangle \( W \times [0,1] \) with base \( W \), and if \( W \) carries a measure \( dv \) then \( \hat{W} \) carries the measure \( d\hat{v} = dv \otimes dt \). We will construct a coupling from a subset of \( \cup_i W_{a,n,i} \) to a subset of \( \cup_j W_{b,n,j} \) 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 \( W_{a,n,i} \) with precisely one \( W_{b,n,i} \).
Furthermore, we can assume that their relative masses agree:
\[
\frac{\hat{v}_{a,s,i}(\hat{W}_{a,s,i})}{Z_{a,s}} = \frac{\hat{v}_{b,s,i}(\hat{W}_{b,s,i})}{Z_{b,s}} \quad \forall \ i.
\]
(15)

Here \(\hat{v}_{a,s,i}\) is the measure on \(\hat{W}_{a,s,i}\), and \(Z_{a,s} = \sum_i \hat{v}_{a,s,i}(\hat{W}_{a,s,i}) = v(W_{a,s})\). To see that no generality is lost making such assumptions, consider a rectangle \(\hat{W}\) with a measure \(\hat{v}\) 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 \(\hat{v}|_{W \times I_k}\). 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 \(\hat{v}|_{W \times I_k}\) under the affine map \(\hat{A}: W \times I_k \to W \times [0, 1]\) is precisely \(\hat{v}|_{W \times I_k}\). Subdividing the rectangles in the families \(\{\hat{W}_{a,s,i}\}\) and \(\{\hat{W}_{b,s,i}\}\) 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 \(\hat{W}_{a,s,i}\) now comes with an associated affine map \(\hat{A}_{a,s,i}\) (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 \(M\), but this is not a matter of concern.

For each fixed \(i\), we can couple a subset of \(\hat{W}_{a,s,i}\) to a subset of \(\hat{W}_{b,s,i}\) as follows. Choose a number \(\tau_a \in (0, 1/2]\) such that
\[
\tau_a \cdot Z_{a,s} = \frac{d_0}{2}.
\]
(16)

Now, fix \(i\). Omitting some ornaments for the sake of readability, let \(h\) stand for the holonomy map from \(W_{a,s,i}\) to \(W_{b,s,i}\) associated with the stable foliation \(W^h\) and denote by \(\rho\) the density of \(\hat{v}_{a,s,i}\) with respect to \(dW_{a,s,i} \otimes dt\).

The subset \(\hat{W}'_{a,s,i} = \{(x, t) \in \hat{W}_{a,s,i} : 0 \leq t \leq \tau_a\}\) is coupled to a corresponding subset \(\hat{W}'_{b,s,i} = \{(y, s) \in \hat{W}_{b,s,i} : 0 \leq s \leq \tau_{b,i}(y)\}\) via the coupling map \(\Theta'_{s,i}: \hat{W}'_{a,s,i} \to \hat{W}'_{b,s,i}: (x, t) \mapsto (y, s)\) with
\[
y = hx \quad \text{and} \quad s = \frac{\tau_{b,i}(y)}{\tau_a} t.
\]

Note, however, that \(\tau_{b,i}\) is not constant but a function. It is given by the consistency rule
\[
\tau_{b,i}(y) \rho_b(y) = \frac{\tau_a \rho_a(x)}{\hat{h}(x)}.
\]
(17)

The expression on the right-hand side of (17) equals the pushforward of the density \(\tau_a \rho_a\) under the holonomy map, evaluated at \(y = hx\). This guarantees that the coupling is measure preserving: if \(f\) is a measurable function \(\hat{W}_{b,s,i} \to \mathbb{R}\), then
\[
\int_{\hat{W}'_{a,s,i}} (f \circ \Theta'_{s,i})(x, t) \, d\hat{v}_{a,s,i}(x, t) = \int_{W_{a,s,i}} \int_0^{\tau_a} f(\Theta'_{s,i}(x, t)) \, \rho_a(x) \, dW_{a,s,i}(x) \, dt
\]
\[
= \int_{W_{a,s,i}} \left[ \int_0^{\tau_a} f(\Theta'_{s,i}(x, t)) \, dt \right] \rho_a(x) \, dW_{a,s,i}(x)
\]
\[
= \int_{W_{a,s,i}} \left[ \int_0^{\tau_a} f(y, \frac{\tau_{b,i}(y)}{\tau_a} t) \, dt \right] \frac{\rho_a(x)}{\hat{h}(x)} \, dW_{b,s,i}(y)
\]
\[
= \int_{W_{b,s,i}} \left[ \int_0^{\tau_{b,i}(y)} f(y, s) \, ds \right] \frac{\tau_a \rho_a(x)}{\tau_{b,i}(y) \hat{h}(x)} \, dW_{b,s,i}(y)
\]
\[
= \int_{W_{b,s,i}} f(y, s) \, d\hat{v}_{b,s,i}(y, s).
\]
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 \( \{ \hat{W}_{α,0,i} \} \) and \( \{ \hat{W}_{β,0,i} \} \) via a measure preserving map.

We are now in position to describe the desired coupling map \( Θ \) from a subset \( \hat{W}_α \subset \hat{W}_α \) to a subset \( \hat{W}_β \subset \hat{W}_β \). Define

\[
\hat{W}_α = \{(x, t) \in \hat{W}_α : (F_{s_0}x, \tilde{α}_{a,s_0,i}) \in \hat{W}_{α,s_0,i} \text{ for some } i\}, \\
\hat{W}_β = \{(y, s) \in \hat{W}_β : (F_{s_0}y, \tilde{β}_{β,s},s) \in \hat{W}_{β,s},i \text{ for some } i\}.
\]

The bijective map \( Θ : \hat{W}_α \to \hat{W}_β \) is defined for a point \((x, t) \in \hat{W}_α\) such that \((F_{s_0}x, \tilde{α}_{a,s_0,i}) \in \hat{W}_{α,s_0,i}\) by the rule

\[
(y, s) = Θ(x, t) \iff (F_{s_0}y, \tilde{β}_{β,s},s) = Θ'(y, s) = θ_α(x, t).
\]

Because the affine maps and the couplings \( Θ'_{s_0,i} \) are measure preserving, also \( Θ \) is measure preserving; the pushforward of the measure \( \tilde{μ}_β \mid \hat{W}_α \) under \( Θ \) is \( \tilde{μ}_β \mid \hat{W}_β \). In particular, the amount of coupled mass equals \( \tilde{μ}_β(\hat{W}_α) = \tilde{μ}_β(\hat{W}_β) = d_0/2 \). Finally, the coupling time function \( Υ : \hat{W}_α \to \mathbb{N} \) is defined by

\[
Υ(x, t) = s_0.
\]

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 \( G = \{(W_α, ν_α) \}_{α \in \mathfrak{A}} \) and \( E = \{(W_β, ν_β) \}_{β \in \mathfrak{B}} \), we can couple a subset \( \hat{W}_α \subset \hat{W}_α \) with a subset \( \hat{W}_β \subset \hat{W}_β \) for any pair \((α, β) \in \mathfrak{A} \times \mathfrak{B}\), in which case we have \( ν_α(\hat{W}_α) = ν_β(\hat{W}_β) = d_0/2 \). Recall that the index sets \( \mathfrak{A} \) and \( \mathfrak{B} \) carry probability factor measures \( λ_α \) and \( λ_β \), 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 \( Δ : \mathfrak{A} \to \mathfrak{B} \) that preserves measure, i.e. \( Δλ_α = λ_β \). Hence, the coupling map \( Θ \) can be constructed from a subset \( \cup_{α \in \mathfrak{A}} \hat{W}_α \subset \cup_{α \in \mathfrak{A}} W_α \) to a subset \( \cup_{β \in \mathfrak{B}} W_β \subset \cup_{β \in \mathfrak{B}} W_β \) so that measure is preserved and in particular so \( \tilde{μ}_β(\cup_{α \in \mathfrak{A}} \hat{W}_α) = \tilde{μ}_β(\cup_{β \in \mathfrak{B}} W_β) = d_0/2 \). On \( \cup_{α \in \mathfrak{A}} \hat{W}_α \) we set \( Υ = s_0 \). The map \( Θ \) can be constructed also for uncountable families, but we omit the details [8].

### 4.5. Recovery step

Our task is to couple the remaining points of \( \hat{W}_α \setminus \hat{W}_α \) to those of \( \hat{W}_β \setminus \hat{W}_β \) and to extend \( Θ \) and \( Υ \) to all of \( \hat{W}_α \). We do this recursively. But first we need to prepare the uncoupled parts of the images \( F_{s_0} \hat{W}_α \) and \( F_{s_0} \hat{W}_β \) so that they also can undergo the coupling procedure described above.

On the one hand, we have the sets \( F_{s_0} \hat{W}_α \setminus \bigcup_i \hat{W}_{α,s_0,i} \) and \( F_{s_0} \hat{W}_β \setminus \bigcup_i \hat{W}_{β,s_0,i} \) 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 \( \hat{W}_{α,s_0,i} \setminus \hat{W}_{α,s_0,i} = \{(x, t) \in \hat{W}_{α,s_0,i} : τ_α < t \leq 1\} \) and \( \hat{W}_{β,s_0,i} \setminus \hat{W}_{β,s_0,i} = \{(y, s) \in \hat{W}_{β,s_0,i} : τ_β(y) < s \leq 1\} \) whose bottom complements were coupled already. We stretch each \( \hat{W}_{α,s_0,i} \setminus \hat{W}_{α,s_0,i} \) affinely onto the complete rectangle \( \hat{W}_{α,s_0,i} \), replacing the density \( ρ_α \) on it by \((1 - τ_α)ρ_α\). Similarly, we stretch each \( \hat{W}_{β,s_0,i} \setminus \hat{W}_{β,s_0,i} \) onto \( \hat{W}_{β,s_0,i} \) so that each vertical fibre \( \{s : (y, s) \in \hat{W}_{β,s_0,i}, τ_β(y) < s \leq 1\} \) is mapped affinely.
onto $[0, 1]$, and replace the density $\rho_\beta$ 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. Note that the base curves $W_{\cdot,s_0,i}$ 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 $T_{r_0+t_0,s_0+1}$ 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 $\eta_\ell = \eta_0^6$. Then $\sup \tau_{\beta,i} \leq 1$. 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 $T_{r_0+t_0,s_0+1}$.

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

By (the proof of) lemma 9, $\rho_\alpha$ and $\rho_\beta$ are regular densities on the curves $W_{\cdot,s_0,i}$ 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 $h$ maps $W_{\cdot,s_0,i}$ onto $W_{\beta,s_0,i}$ by sliding along the connecting leaves of the stable foliation $W^m$. Both curves as well as the leaves are inside the magnet. Assuming that the magnet is sufficiently small, $h$ is as close to the identity as we wish: given any $\delta > 0$, we may assume that $|Jh - 1| \leq \delta$. This follows immediately from lemma 11, because it can be applied to the $k$-step pullback of $h$ that maps $T_{r_0+t_0-k+1} W_{\cdot,s_0,i}$ onto $T_{r_0-t_0-k+1} W_{\beta,s_0,i}$ with $k$ large and the connecting leaves of $W^{m-k}$ sufficiently short (shorter than $\ell_0$) for lemma 11 to apply.

For each $x \in W_{\cdot,s_0,i}$ denote $y = hx \in W_{\beta,s_0,i}$. As $|W_{\beta,s_0,i}(y_1, y_2)| = \int_{W_{\cdot,s_0,i}} |W_{\beta,s_0,i}(x_1, x_2)| = (1 - \delta)|W_{\cdot,s_0,i}(x_1, x_2)| \leq |W_{\beta,s_0,i}(y_1, y_2)| \leq (1 + \delta)|W_{\cdot,s_0,i}(x_1, x_2)|$. (18)

Observe that from (4) follows easily

$$e^{-C_1|W_{\cdot,s_0,i}|^\eta} \leq \frac{|W_{\cdot,s_0,i}|}{v_{\cdot,s_0,i}(W_{\cdot,s_0,i})} \rho_\cdot \leq e^{C_1|W_{\cdot,s_0,i}|^\eta}$$

or

$$\rho_\cdot = \frac{v_{\cdot,s_0,i}(W_{\cdot,s_0,i})}{|W_{\cdot,s_0,i}|} (1 + O(|W_{\cdot,s_0,i}|^\eta)).$$

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

$$\tau_{\beta,i}(y) \leq \frac{\tau_\alpha}{1 - \delta} \rho_\alpha(x) \leq \frac{\tau_\alpha}{1 - \delta} \rho_\alpha(y) \leq \frac{\tau_\alpha}{1 - \delta} 1 + O(|W_{\cdot,s_0,i}|^\eta) \frac{|W_{\beta,s_0,i}|}{|Z_{\cdot,s_0,i}|} Z_{\cdot,s_0,i}$$

$$\leq \frac{1}{2} \left( 1 + \delta 1 + O(|W_{\cdot,s_0,i}|^\eta) \right)$$

where we have recalled (15), (16), $Z_{\cdot,s_0} = v_\beta(W_{\cdot,s_0,i}) > d_{s_0}$, and (18). The right-hand side can be made arbitrarily close to $\frac{1}{2}$, so that we can take, say, $\sup \tau_{\beta,i} \leq \frac{3}{4}$. This fixes the value of $\eta_\ell$. 

From (17) and lemma 12 it then follows that

\[
| \ln(\tau_{\beta,i}(y_1)\rho_{\beta}(y_1)) - \ln(\tau_{\beta,i}(y_2)\rho_{\beta}(y_2)) | \\
\lesssim | \ln \rho_{\alpha}(x_1) - \ln \rho_{\alpha}(x_2) | + | \ln Jh(x_2) - \ln Jh(x_1) | \\
\lesssim C_t [W_{a,\nu_0}(x_1, x_2)]^n + C_h [W_{a,\nu_0}(x_1, x_2)]^n \\
\lesssim (C_t + C_h) (1 - \delta)^{-\min(\eta_\gamma, \eta_h)} |W_{\beta,\nu_0,i}(y_1, y_2)|^{\min(\eta_\gamma, \eta_h)}.
\]

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

\[
\min(a, b) | \ln a - \ln b | \leq | a - b | \leq \max(a, b) | \ln a - \ln b | \quad a, b > 0
\]

obtained from the mean-value theorem,

\[
| \ln(1 - \tau_{\beta,i}(y_1)) - \ln(1 - \tau_{\beta,i}(y_2)) | \\
\lesssim \frac{| \tau_{\beta,i}(y_1) - \tau_{\beta,i}(y_2) |}{1 - \sup \tau_{\beta,i}} \\
\lesssim \frac{1}{1 - \sup \tau_{\beta,i}} | \ln \tau_{\beta,i}(y_1) - \ln \tau_{\beta,i}(y_2) | \\
\lesssim 3 (2C_t + C_h)(1 - \delta)^{-\min(\eta_\gamma, \eta_h)} |W_{\beta,\nu_0,i}(y_1, y_2)|^{\min(\eta_\gamma, \eta_h)}.
\]

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 \( G \) and \( E \), the first coupling is constructed at time \( s_0 \), when enough mass of each family is on the magnet \( M_{s_0} \):

\[
\hat{\mu}_G(\Upsilon = s_0) = \frac{d_0}{2}.
\]

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,

\[
\hat{\mu}_G(\Upsilon = k(s_0 + r_0) + s_0 \mid \Upsilon > (k - 1)(s_0 + r_0) + s_0) = \frac{d_0}{2}
\]

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

\[
\hat{\mu}_G(\Upsilon = k(s_0 + r_0) + s_0) = \frac{d_0}{2} \left(1 - \frac{d_0}{2}\right)^k.
\]

This completes 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:

$$\text{dist}(A, B) = \| P_A - P_B \|,$$

where $P$ are the corresponding orthogonal projections. Note that

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

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

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

where

$$\langle A, B \rangle = \max_{u \in A, v \in B} 1 - |u, v|. \quad (22)$$

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

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

Hence,

$$\text{dist}'(A, B) \leq \sqrt{2} \text{dist}(A, B)^{1/2}. \quad (23)$$

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 $\varepsilon_q$ is small enough. If each $T_i \in \mathcal{U}_q$ for a fixed $q$ and if $v \in T_i \mathcal{M} \backslash C^2_{q,n}$, then $\| D_i T_n v \| \geq C'_q \Lambda_q^n \| v \|$ for all $n \geq 1$ and $D_i T_n v \in C^0_{q,T_n}$ for all $n \geq k_q$. These statements are uniform in $x$.

Proof. If $v \in T_i \mathcal{M} \backslash C^2_{q,n}$, we have $\| v'' \| > a_q \| v' \|$. But $v''_n \equiv D_i \tilde{T}_q^n v''_n \in E^{u,s}_{q,T_n}$, and $\| v''_n \| \geq C_q \Lambda_q^n \| v'' \|$ and $\| v'' \| \geq C_q \Lambda_q^n \| v_n'' \|$. We have $\| v_n'' \| \geq C_q \Lambda_q^n \| v_n'' \|$ for $n \geq k_q$, if $k_q$ is sufficiently large. Because $\| D_i T_n v \| \geq C \varepsilon_q$, $\| D_i T_n v'' \| \geq C \| v'' \|$, and $\| D_i \tilde{T}_q^n v'' \| \geq c \| v'' \|$ hold with some $C = C(k_q)$ and $c = c(k_q)$, we have

$$\| D_i T_n v'' \| \geq \| D_i \tilde{T}_q^n v'' \| - C \varepsilon_q \| v'' \| \geq \| D_i \tilde{T}_q^n v'' \|(1 - C^{-1} \varepsilon_q)$$

$$\geq 2^{-1} \| D_i \tilde{T}_q^n v'' \|(1 - C^{-1} \varepsilon_q)$$

$$\geq 2^{a_q^{-1} \| D_i T_n v'' \|} (1 - C^{-1} \varepsilon_q)$$

$$\geq 2^{a_q^{-1} \| D_i T_n v'' \|} (1 - C^{-1} \varepsilon_q)^2 \geq a_q^{-1} \| D_i T_n v'' \|,$$

provided $\varepsilon_q$ is small enough. This estimate shows that $D_i T_n v \in C^0_{q,T_n}$.

The uniform estimate $\| D_i T_n v \| \geq c_q \| v \|$ holds with some $c_q = c_q(k_q) < 1$ for $1 \leq n < k_q$. If $n = k_q + m$, $\| D_i T_n v \| \geq C_q \Lambda_q^m \| D_i T_n v \| \geq c_q C_q \Lambda_q^m \| v \|$. Hence, we can set $C'_q = c_q C_q / \Lambda_q^{k_q}$, so that $\| D_i T_n v \| \geq C'_q \Lambda_q^n \| v \|$ for all $n \geq 1$. \hfill $\Box$

The next result on linear maps is cited from [6, lemma 6.1.1] up to notational changes.
Lemma 25. Let $L_n^b$ and $L_n^c$, $n \in \mathbb{N}$, be two sequences of linear maps $\mathbb{R}^M \to \mathbb{R}^M$. Assume that for some $b > 0$ and $\delta \in (0, 1)$,
\[
\|L_n^b - L_n^c\| \leq \delta b^n, \quad n \geq 0.
\]
Suppose there are two subspaces $E^1$ and $E^2$ of $\mathbb{R}^M$ and constants $C_\ast > 1$, $0 < \lambda_\ast < \mu_\ast$ with $\lambda_\ast < b$ such that
\[
\|L_n^b v\| \leq C_\ast \lambda_\ast^n \quad \text{if } v \in E^1,
\]
\[
\|L_n^b w\| \geq C_\ast^{-1} \mu_\ast^n \quad \text{if } w \perp E^1.
\]
Then
\[
dist(E^1, E^2) \leq 3C_\ast \frac{H}{\lambda_\ast} \sqrt{\frac{\ln \mu_\ast - \ln \lambda_\ast}{\ln b - \ln \lambda_\ast}}.
\]

Proof of lemma 23. We generalize the case of a single map found in [6]. As the orthogonal complements $(T_i, M)^\perp$ of $T_iM$ in $\mathbb{R}^M$ form a smooth distribution $(T_iM)^\perp$ on $M$, we first prove that the distribution $E_0^0 \oplus (T_iM)^\perp$ is Hölder continuous on $M$ and then deduce the Hölder continuity of $E_0^0$.

Let $P_s$ be the orthogonal projection $\mathbb{R}^M \to T_sM$. It depends smoothly on $x$. We define
\[
L^{(i)}(x) = D_i T_i \circ P_s.
\]
This map extends $D_i, T_i$ to a linear map $\mathbb{R}^M \to \mathbb{R}^M$. Let us also set
\[
L_n(x) = L^{(i_1)}(T_{i_1} x) \cdots L^{(i_1)}(x).
\]
Recall from above that $E_0^0 \subset C_{1,s}$. Setting $C = \prod_{1 \leq q \leq Q} C_q$ and $\Lambda = \min_{1 \leq q \leq Q} \Lambda_q$, we have
\[
\|D_i T_n v\| \leq C^{-1} \Lambda^{-n} \|v\|, \quad v \in E_0^0,
\]
by (A2). This translates to
\[
\|L_n(x) v\| \leq C^{-1} \Lambda^{-n} \|v\|, \quad v \in E_0^0 \oplus (T_iM)^\perp.
\]
By assumption (A4), we can make the cones so narrow that a vector in $T_iM$ perpendicular to $E_0^0$ lies in the complement of $C_{1,s}$. Setting $C' = \prod_{1 \leq q \leq Q} C'_q$, lemma 24 thus yields
\[
\|D_i T_n w\| \geq C' \Lambda^n \|w\|, \quad w \in T_iM : w \perp E_0^0.
\]
In other words,
\[
\|L_n(x) w\| \geq C' \Lambda^n \|w\|, \quad w \perp E_0^0 \oplus (T_iM)^\perp.
\]
Clearly $\|L^{(i)}(x)\| = \|D_i T_i\|$. For brevity, let us write $b_1 = \sup_{T_i \in M} \|D_i T\|$ and $b_2 = \sup_{T \in M} \|D_i (T_i T \circ P_s)\|$, where $T$ runs over $\cup_i \cup_i \cup_{Q} \cup_{Q}$. Since $L_{n+1}(x) = L^{(i+1)}(T_{i+1}) L_n(x)$,
\[
\|L_{n+1}(x) - L_{n+1}(y)\|
\leq \|L^{(i+1)}(T_{i+1})\| \|L_n(x) - L_n(y)\| + \|L^{(i+1)}(T_{i+1})\| \|L_n(y)\|
\leq b_1 \|L_n(x) - L_n(y)\| + b_2 \|T_{i+1} x - T_{i+1} y\| b^n_1
\leq b_1 \|L_n(x) - L_n(y)\| + b_2 b^n_1 \|x - y\|.
\]
As $\|L_1(x) - L_2(y)\| \leq b_2 \|x - y\|$, we obtain the bound
\[
\|L_n(x) - L_n(y)\| \leq c \|x - y\| b^n_1,
\]
with the constant $c = \frac{b_1}{b_2(b_1 - 1)}$. 

The bounds (22), (23) and (24) show that all conditions of lemma 25 are satisfied, if we take \( \mathcal{L}_n = \mathcal{L}_n(x) \), \( \mathcal{L}_n = \mathcal{L}_n(y) \), \( \epsilon^1 = \mathcal{E}_x \oplus (T_x M)^\perp \) and \( \mathcal{E}^2 = \mathcal{E}_y \oplus (T_y M)^\perp \). Writing \( \alpha = 2 \ln \Lambda/(2 \ln b_1 + \ln \Lambda) \) and \( K = 3 \max(C^{-1}, C^{-1})^2 \Lambda^2 e^\alpha \),
\[
\text{dist}(\mathcal{E}^1 \oplus (T_x M)^\perp, \mathcal{E}^2 \oplus (T_y M)^\perp) \leq K \|x - y\|^\alpha
\]
provided \( \|x - y\| < 1/c \).

By compactness of the manifold and smoothness of the distribution \( (T_x M)^\perp \), we have \( \text{dist}(\mathcal{E}^1 \oplus (T_x M)^\perp, \mathcal{E}^2 \oplus (T_y M)^\perp) \leq L \|x - y\|^\alpha \) for some \( L \). Hence, by (19) and (20),
\[
\text{dist}(\mathcal{E}^1 \oplus (T_x M)^\perp, \mathcal{E}^2 \oplus (T_y M)^\perp) \leq (K + L) \|x - y\|^\alpha \quad \text{if} \quad \|x - y\| < 1/c.
\]

Note that this bound does not depend on the sequence \((T_i)_{i \geq 1}\). Moreover, the same upper bound is obtained for each distribution \( \mathcal{E}^n \) by disregarding the first \( n \) maps and considering the sequence \((T_i)_{i \geq n}\) instead. The result for \( F^n \) is obtained by reversing time. \( \square \)

**Appendix C. Inclination lemma type results**

**Lemma 26.** Fix \( 1 \leq q \leq Q \) and let \( T_i \in \mathcal{U}_q \) for each \( i \). There exists a constant \( C_q > 0 \) such that, for all \( w \in C^a_{q, x} \) and all \( \tilde{w} \in T_x M \) with \( \|w\| = \|	ilde{w}\| = 1 \), the vectors \( w_n = D_n T_n w \) and \( \tilde{w}_n = D_n T_n \tilde{w} \) satisfy
\[
\|\tilde{w}_n\|/\|w_n\| \leq C_q \quad \forall n \geq 0.
\]
If also \( \tilde{w} \in C^a_{q, x} \), the angle between \( w_n \) and \( \tilde{w}_n \) tends to zero at a uniform exponential rate:
\[
1 - \left| \left( \frac{w_n}{\|w_n\|} - \frac{\tilde{w}_n}{\|	ilde{w}_n\|} \right) \right| \leq \min(c, C \Lambda^{-dn}) \tag{25}
\]
for some \( 0 < c < 1 < C > 0 \).

**Proof.** We first construct ‘fake’ stable and unstable distributions for finite sequences \( T_1, \ldots, T_N \). Fix \( N \geq 1 \). First, choose a distribution \( E^N \) such that \( E^N \subset C^a_{q, x} \) for all \( x \). Then define the distributions \( E^0 = D_n T_n \mathcal{E}^0 \), \( 0 < n < N \), by pulling back: \( E^0 = D_n T_n E^0 \subset C^a_{q, x} \).

Next, choose a distribution \( F^0 \) such that \( F^0 \subset C^a_{q, x} \). Then set recursively \( F^n = D_n T_n F^{n-1} \) for \( n \geq 1 \). We will use \( E^0_n \) and \( F^n \) as coordinate axes. Let \( e_n^a \in E^a_n \) and \( p_n^a \in F^n \) be unit vectors oriented so that \( D_n T_n e_n^a \) and \( e_{n+1}^a \) point in the same direction and \( D_n T_n P_{n+1}^a \) and \( F^n \) point in the same direction.

Recall from section 1.3 that the angle between \( E^a_{q, x} \) and \( E^r_{q, x} \) is uniformly bounded away from zero. In other words, there exists a \( \psi_q > 0 \) such that \( (E^a_{q, x}, E^r_{q, x}) \leq 1 - 2 \psi_q \) for all \( x \in \mathcal{M} \). By assumption (A4), the cones can be assumed narrow enough, so that \( (U, V) \leq 1 - \psi_q \) for all subspaces \( U \subset C^a_{q, x} \) and \( V \subset C^r_{q, x} \), and for all \( x \in \mathcal{M} \). Because of this, \( |(e_n^a, p_n^a)| \leq 1 - \psi_q \) for all \( x \) and all \( 0 \leq n \leq N \). Thus, we have the uniform bounds
\[
\psi_q (\alpha^2 + \beta^2) \leq \|\alpha^a n + \beta^a e_n^a\|^2 \leq 2(\alpha^2 + \beta^2), \quad \forall \alpha, \beta, \quad 0 \leq n \leq N. \tag{26}
\]
Now write \( w_n = n_{P_{n}^a} + n_{e_n^a} \) and \( \tilde{w}_n = n_{P_{n}^a} + \tilde{n}_{e_n^a} \) and estimate, for \( 0 \leq n \leq N \),
\[
\|\tilde{w}_n\|^2/\|w_n\|^2 \leq \frac{2(\alpha^2 + \beta^2)}{\psi_q^2 \alpha^2} = \frac{2\alpha^2}{\psi_q^2 \alpha^2} + \frac{2\beta^2}{\psi_q^2 \beta^2} \leq \frac{2\alpha^2}{\psi_q^2 \alpha^2} + \frac{2\beta^2}{\psi_q^2 \beta^2} (C_q \Lambda^n)^2 \leq \frac{2\alpha^2}{\psi_q^2 \alpha^2} + \frac{2\beta^2}{\psi_q^2 \beta^2} (C_q \Lambda^n)^2.
\]
We have used assumption (A2) to bound the norms involving \( D_n T_n \). From (26), \( \tilde{\alpha}^2 \), \( \tilde{\beta}^2 \) \( \tilde{\beta}_0 \) is bounded from below by some \( A_q > 0 \). Thus
\[
\frac{\|\tilde{u}_n\|^2}{\|u_n\|^2} \leq \frac{2}{\psi_q^2 A_q^3} (1 + C_q^{-4}) \equiv C_q^2 \quad 0 \leq n \leq N.
\]
But \( N \) was arbitrary, so the bound holds for all \( n \geq 0 \). In particular, \( C_q \) does not depend on the constructed distributions. A computation also shows that, if both \( \tilde{w}, \tilde{w}_n \) are of order \( \beta_n \beta_n'/|\alpha_0\alpha_0'| \), then
\[
1 - \frac{\|u_n, \tilde{u}_n\|}{\|u_n\|\|\tilde{u}_n\|} \text{ is of order } C_q^{-4} \beta_n \beta_n'/|\alpha_0\alpha_0'| \leq C_q^{-4} A_q^{-\delta n}. \]
\( \square \)

**Lemma 27.** Fix \( 1 \leq q \leq Q \), \( 0 < \lambda < 1 \), and \( \delta > 0 \). Fix \( N \geq 1 \) and \( T_1 \in \mathcal{U}_q \) for \( 1 \leq i \leq N \). Take two points \( x_1, x_2 \) and assume that \( d(T_n x_1, T_n x_2) < \delta \lambda^n \) if \( 0 \leq n \leq N \). Suppose \( W_1, W_2 \) are unstable curves with respect to \( \{C_q^0\} \) and that \( x_i \in W_i \). Then
\[
\frac{|\gamma_{T_n W_i T_{n+1}(T_n x_1)} - 1|}{|\gamma_{T_n W_i T_{n+1}(T_n x_2)} - 1|} \leq C' \mu^n \quad 0 \leq n \leq N - 1. \tag{27}
\]
The constants \( C' > 1 \) and \( 0 < \mu < 1 \) are independent of \( N \), of the curves \( W_i \), and of the choice of \( T_1, \ldots, T_N \), as long as the bound on \( d(T_n x_1, T_n x_2) \) continues to hold.

**Proof.** Choose \( \tilde{F}_0^q = E_{\tilde{F}_0}^q \) and define \( \tilde{F}_1^q = D_{T_{n+1}T_n} \tilde{F}_{T_n}^{-1} \) for \( 1 \leq n \leq N \). By lemma 23\(^7\), the distributions \( \tilde{F}_q^q \) belong to a fixed H"older class, no matter which \( T_1, \ldots, T_N \) and \( N \) are chosen. Note that \( \tilde{F}_q^q \subset C_{\tilde{F}_q}^0 \) for \( 0 \leq n \leq N \).

Let \( u^n_i \) \( (i = 1, 2) \) stand for a unit tangent vector of \( T_n W_i \) at \( x^n_i = T_n x_i \). We also write \( L(x) = D_1 T_{n+1} \circ P_x \), where \( P_x \) is the orthogonal projection \( \mathbb{R}^M \to T_x \mathcal{M} \), which is smooth.
\[
\frac{|\gamma_{T_n W_i T_{n+1}(x^n_1)} - 1|}{|\gamma_{T_n W_i T_{n+1}(x^n_2)} - 1|} \leq \frac{1}{C_q} \frac{||D_{x^n_1} u^n_1 - D_{x^n_2} u^n_2||}{||L(x^n_1)u^n_1|| + ||L(x^n_2)u^n_2||} \leq \frac{1}{C_q} \left( \|L(x^n_1) - L(x^n_2)\| + \|L(x^n_2)\| \min_{\sigma = \pm 1} \|u^n_\sigma - \sigma u^n_2\| \right) \leq C \left( \lambda^n + \text{dist}(U^n_i, U^n_j) \right).
\]
We have denoted by \( U^n_i \) the linear subspaces of \( \mathbb{R}^M \) spanned by \( u^n_i \) and recalled the definition in (21). The angle between \( U^n_i \) and \( \tilde{F}_q^q \) decays exponentially: \( \text{dist}(U^n_i, \tilde{F}_q^q) \leq C \Lambda_q^{-2n} \) due to (21) and (25). Hence, it suffices to prove exponential decay of \( \text{dist}(\tilde{F}_q^q, \tilde{F}_q^q) \). But this follows from Hölder continuity and the assumption \( d(x^n_1, x^n_2) < \delta \lambda^n \). \( \square \)

**Acknowledgments**

The author has received financial support from the Academy of Finland and the V"ais"al"a Fund. He wishes to thank Lai-Sang Young for many discussions.

**References**

[1] Arnoux P and Fisher A M 2005 Anosov families, renormalization and non-stationary subshifts Ergod. Theory Dyn. Syst. 25 661–709

[2] Ayer A and Stenlund M 2007 Exponential decay of correlations for randomly chosen hyperbolic toral automorphisms Chaos 17 043116

\(^7\) 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.
[3] Bakhtin V I 1994 Random processes generated by a hyperbolic sequence of mappings: I Izv. Ross. Akad. Nauk Ser. Mat. 58 40–72
[4] Bakhtin V I 1994 Random processes generated by a hyperbolic sequence of mappings: II Izv. Ross. Akad. Nauk Ser. Mat. 58 184–95
[5] Bressaud X and Liverani C 2002 Anosov diffeomorphisms and coupling Ergod. Theory Dyn. Syst. 22 129–52
[6] Brin M and Stuck G 2002 Introduction to Dynamical Systems (Cambridge: Cambridge University Press)
[7] Chernov N and Dolgopyat D 2009 Brownian motion: I Mem. Am. Math. Soc. 198 viii+193
[8] Chernov N 2006 Advanced statistical properties of dispersing billiards J. Stat. Phys. 122 1061–94
[9] Conze J-P and Raugi A 2007 Limit theorems for sequential expanding dynamical systems on [0, 1] Ergodic Theory and Related Fields (Contemporary Mathematics vol 430) (Providence, RI: American Mathematics Society) pp 89–121
[10] Katok A and Hasselblatt B 1995 Introduction to the Modern Theory of Dynamical Systems (Encyclopedia of Mathematics and its Applications vol 54) (Cambridge: Cambridge University Press) (With a supplementary chapter by A Katok and L Mendoza)
[11] Masmoudi N and Young L-S 2002 Ergodic theory of infinite dimensional systems with applications to dissipative parabolic PDEs Commun. Math. Phys. 227 461–81
[12] Ott W, Stenlund M and Young L-S 2009 Memory loss for time-dependent dynamical systems Math. Res. Lett. 16 463–75
[13] Young L-S 1999 Recurrence times and rates of mixing Israel J. Math. 110 153–88