RECURRENCE TIMES AND RATES OF MIXING

Lai-Sang Young*

Department of Mathematics
University of California
Los Angeles, CA 90095-1555
Email: lsy@math.ucla.edu

August 1997

ABSTRACT. The setting of this paper consists of a map making "nice" returns to a reference set. Criteria for the existence of equilibria, speed of convergence to equilibria and for the central limit theorem are given in terms of the tail of the return time function. The abstract setting considered arises naturally in differentiable dynamical systems with some expanding or hyperbolic properties.

This paper is part of an attempt to understand the speed of mixing and related statistical properties for chaotic dynamical systems. More precisely, we are interested in systems that are expanding or hyperbolic on large parts (though not necessarily all) of their phase spaces. A natural approach to this problem is to pick a suitable reference set, and to regard a part of the system as having "renewed" itself when it makes a "full" return to this set. We obtain in this way a representation of the dynamical system in question, described in terms of a reference set and return times. We propose to study this object *abstractly*, that is to say, to set aside the specific characteristics of the original system and to understand its statistical properties purely in terms of these recurrence times. Needless to say, if we are to claim that this approach is valid, we must also show that it is implementable, and that it gives reasonable results in interesting, concrete situations.

The ideas described above were put forth in [**Y**]; they continue to be the underlying theme of the present paper. In [**Y**] we focused on mixing at exponential speeds. One of the aims of this paper is to extend the abstract part of this study to *all speeds of mixing.* Of particular interest is when the recurrence is polynomial, *i.e.* when the probability of not returning in the first n iterates is of order $n^{-\alpha}$. We will show in this case that the speed of mixing is of order $n^{-\alpha+1}$. More generally, let R

*The author is partially supported by a grant from the National Science Foundation and a Guggenheim Fellowship

denote the return time function and m a suitable reference measure on the reference set. We find that the *type* of mixing, meaning whether it is exponential, stretched exponential, or polynomial etc., is determined by the asymptotics of $m\{R > n\}$ as n tends to infinity.

A useful tool for studying decay or correlations is the Perron-Frobenius or transfer operator. Exponential decay corresponds to a gap in the spectrum of this operator, or equivalently, a contraction of some kind with each iteration of the map. Various techniques have been developed for proving the presence of this gap (see e.g. [**R**], [**HK**], [**L1**]), but to my knowledge no systematic way of capturing slower decay rates in chaotic systems have been devised. The method employed in this paper can be summarized as follows. Given two arbitrary initial distributions, we run the system, and as the two measures evolve we try to match up their densities as best we can. Part of this matching process uses coupling ideas from probability. The speeds with which arbitrary initial densities can be matched up give the speed of convergence to equilibrium in the sense of L^1, and that in turn is an upper bound for the speed of correlation decay. This method is, in principle, equally effective for estimating all decay rates.

As for applications, the scheme described in the first paragraph of this introduction has been carried out for several classes of examples, including dispersing billiards and certain logistic and Hénon-type maps [**Y**], [**BY**]. All these have been shown to have exponential decay of correlations. To augment the list above, and to give a quick example of systems that mix polynomially, we will discuss in this paper piecewise expanding 1-dimensional maps with neutral fixed points. To be sure, there are interesting systems in dimensions greater than one that mix slowly. When the derivative of a map is parabolic on an invariant set, even one of measure zero (such as in certain billiards with convex boundaries), the speed of mixing is likely to be at best polynomial. The detailed analyses of these examples, however, are technically quite involved and will not be included here.

This paper is organized as follows. Part I focuses on the abstract dynamical object that, we claim, arises naturally in many dynamical systems with hyperbolic properties. We will not concern ourselves here with how this object is constructed, but accept it as a starting point and study its statistical properties. Part II contains some simple applications. We refer the reader to [**Y**] for a general discussion of the relation between the abstract model and the original system from which it is derived, and for other applications of these "abstract results".

TABLE OF CONTENTS

PART I. STATISTICAL PROPERTIES OF ABSTRACT MODEL

1. Setting and statements of results

1.1. The setup.

The mathematical object described below arises naturally in many dynamical systems with expanding or hyperbolic properties. In the expanding case, it is obtained by looking at "full returns" to an arbitrary disk; in the (invertible) hyperbolic case, it is obtained by considering returns to a set with a hyperbolic product structure and collapsing along stable manifolds. See [**Y**] for a more detailed discussion.

The setting consists of a map F from a space Δ to itself, together with a reference measure m on Δ. We begin with the coarse structure of $F : \Delta \circlearrowleft$. Let Δ_0 be an arbitrary set partitioned into $\{\Delta_{0,i}\}_{i=1,2,\ldots}$ and let $R : \Delta_0 \to \mathbb{Z}^+$ be a return time function that is constant on each $\Delta_{0,i}$. A formal definition of Δ is given by

$$\Delta := \{(z,n) \in \Delta_0 \times \{0,1,2,\ldots\} : n < R(z)\}.$$

We refer to $\Delta_\ell := \Delta \cap \{n = \ell\}$ as the ℓ^{th} level of the tower, and let $\Delta_{\ell,i} = \Delta_\ell \cap \{z \in \Delta_{0,i}\}$. Let $R_i = R \mid \Delta_{0,i}$, so that $\Delta_{R_i-1,i}$ is the top level of the tower directly above $\Delta_{0,i}$. We shall assume for simplicity that $\gcd\{R_i\} = 1$. The map $F : \Delta \circlearrowleft$ sends (z,ℓ) to $(z,\ell+1)$ if $\ell + 1 < R(z)$, and maps each $\Delta_{R_i-1,i}$ bijectively onto Δ_0. We further assume that the partition $\eta := \{\Delta_{\ell,i}\}$ generates in the sense that $\bigvee_{i=0}^{\infty} F^{-i}\eta$ is the trivial partition into points.

For simplicity of notation we will, from here on, refer to points in Δ as x rather than (z,ℓ) with $z \in \Delta_0$. Also, we will identify Δ_0 with the corresponding subset of Δ and let $F^R : \Delta_0 \circlearrowleft$ denote the map defined by $F^R(x) = F^{R(x)}(x)$.

Next we proceed to describe the finer structures of $F : \Delta \circlearrowleft$. Let $\mathcal{B}$ be a σ-algebra of subsets of Δ. We assume that all the sets mentioned above are $\mathcal{B}$-measurable, F and $(F|\Delta_{\ell,i})^{-1}$ are measurable, and that there is a reference measure defined on $(\Delta, \mathcal{B})$ with $m(\Delta_0) < \infty$. We assume that F carries $m|\Delta_{\ell,i}$ to $m|\Delta_{\ell+1,i}$ for $\ell < R_i - 1$. On the top levels, the regularity of F is dictated by the following "Hölder"-type condition we impose on $F^R : \Delta_0 \circlearrowleft$. First we introduce a notion of separation time for $x, y \in \Delta_0$. Let $s(x,y) :=$ the smallest $n \geq 0$ s.t. $(F^R)^n x$, $(F^R)^n y$

lie in distinct $\Delta_{0,i}$'s, so that $s(x,y) \geq 0 \; \forall x,y \in \Delta_0$, $s(x,y) \geq 1 \; \forall x,y \in \Delta_{0,i}$ etc. For each i, we assume that $F^R|\Delta_{0,i} : \Delta_{0,i} \to \Delta_0$ and its inverse are nonsingular with respect to m, so that its Jacobian JF^R wrt m exists and is > 0 m-a.e. We further require that

$$\exists \; C = C_{F,0} > 0 \quad \text{and} \quad \beta \in (0,1) \quad \text{s.t.} \quad \forall \; x,y \in \Delta_{0,i}, \quad \text{any } i,$$

$$(*) \qquad \left| \frac{JF^R(x)}{JF^R(y)} - 1 \right| \leq C \beta^{s(F^R x, F^R y)}.$$

Sometimes it is convenient to have $s(\cdot,\cdot)$ extended to all pairs $x,y \in \Delta$. One way to do this is to let $s(x,y) = 0$ if x,y do not belong in the same $\Delta_{\ell,i}$; and for $x,y \in \Delta_{\ell,i}$, let $s(x,y) = s(x',y')$ where x',y' are the corresponding points in $\Delta_{0,i}$.

Finally we mention some function spaces that are compatible with the structures already introduced. Let $\beta < 1$ be as above, and let

$$\mathcal{C}_\beta(\Delta) := \{\varphi : \Delta \to \mathbb{R} \mid \exists C_\varphi \text{ s.t. } |\varphi(x) - \varphi(y)| \leq C_\varphi \; \beta^{s(x,y)} \; \forall x,y \in \Delta\},$$

$$\mathcal{C}_\beta^+(\Delta) := \{\varphi \in \mathcal{C}_\beta(\Delta) \mid \exists C_\varphi^+ \text{ s.t. on each } \Delta_{\ell,i}, \text{ either } \varphi \equiv 0 \text{ or}$$

$$\varphi > 0 \text{ and } \left| \frac{\varphi(x)}{\varphi(y)} - 1 \right| \leq C_\varphi^+ \; \beta^{s(x,y)} \; \forall x,y \in \Delta_{\ell,i}\}.$$

The test functions to be considered will belong in $\mathcal{C}_\beta$, while the probability measures will have their densities in $\mathcal{C}_\beta^+$.

The setting and notations of 1.1 will be assumed throughout Part I.

1.2. Statements of results.

For a (signed) measure μ on Δ, we let $(F_*^n \mu)(E) := \mu(F^{-n}E)$ and let $|\mu|$ denote the total variation of μ.

We begin with the following very basic result:

Theorem 1. (Existence and properties of equilibrium measures).
Assume $\int R dm < \infty$. Then

(i) $F : \Delta \circlearrowleft$ *admits an invariant probability measure ν that is absolutely continuous wrt m;*

(ii) $\frac{d\nu}{dm} \in \mathcal{C}_\beta^+$ *and is $\geq c_0$ for some $c_0 > 0$;*

(iii) (F,ν) *is exact, hence ergodic and mixing.*

Assume from here on that $\int R dm < \infty$. Let $\hat{R} : \Delta \to \mathbb{Z}$ be the function defined by

$$\hat{R}(x) = \quad \text{the smallest integer } \; n \geq 0 \; \text{ s.t. } \; F^n x \in \Delta_0.$$

Note that $m\{\hat{R} > n\} = \sum_{\ell > n} m(\Delta_\ell)$. The asymptotics of $m\{\hat{R} > n\}$ as $n \to \infty$ will play an extremely important role in the results to follow.

Theorem 2 is the main result of Part I.

Theorem 2. (Speed of convergence to equilibrium).
(I) Lower bounds. *There exist (many) probability measures λ on Δ with $\frac{d\lambda}{dm} \in \mathcal{C}_\beta^+$ s.t.*

$$|F_*^n\lambda - \nu| \geq c\ m\{\hat{R} > n\}$$

for some $c = c(\lambda) > 0$.
(II) Upper bounds. *For arbitrary λ with $\frac{d\lambda}{dm} \in \mathcal{C}_\beta^+$, an upper bound for $|F_*^n\lambda - \nu|$ is determined by the asymptotics of $m\{\hat{R} > n\}$ in conjunction with certain decreasing exponential functions; see 3.5 for the precise relations. Two special cases are:*

(a) *if $m\{\hat{R} > n\} = \mathcal{O}(n^{-\alpha})$ for some $\alpha > 0$, then for all λ as above,*

$$|F_*^n\lambda - \nu| = \mathcal{O}(n^{-\alpha});$$

(b) *if $m\{\hat{R} > n\} = \mathcal{O}(\theta^n)$ for some $\theta < 1$, then $\exists\ \tilde{\theta} < 1$ s.t. for all λ as above,*

$$|F_*^n\lambda - \nu| = \mathcal{O}(\tilde{\theta}^n).$$

Closely related to the speed of convergence to equilibrium is the speed of correlation decay for random variables of the type $\{\varphi \circ F^n\}_{n=0,1,2,\ldots}$ where the underlying probability space is (Δ, ν) and $\varphi : \Delta \to \mathbb{R}$ is an observable. Let $\mathrm{Cov}(\cdot, \cdot)$ denote the covariance of random variables with respect to ν, and recall that

$$\mathrm{Cov}(\varphi \circ F^n, \psi) = \int (\varphi \circ F^n)\psi d\nu - \int \varphi d\nu \int \psi d\nu.$$

The next theorem is really a corollary of the last.

Theorem 3. (Decay of correlations). *The statements in* Part (II) *of* Theorem 2 *continue to be valid if $|F_*^n\lambda - \nu|$ is replaced by $|\mathrm{Cov}(\varphi \circ F^n, \psi)|$ with $\varphi \in L^\infty(\Delta, m)$ and $\psi \in \mathcal{C}_\beta(\Delta)$.*

For $\varphi : \Delta \to \mathbb{R}$ with $\int \varphi d\nu = 0$, we say that the Central Limit Theorem holds for φ (with underlying probability space (Δ, ν)) if $\frac{1}{\sqrt{n}} \sum_{i=0}^{n-1} \varphi \circ F^i$ converges in law to a normal distribution $\mathcal{N}(0, \sigma)$.

Theorem 4. (Central Limit Theorem). *If $m\{\hat{R} > n\} = \mathcal{O}(n^{-\alpha})$ for some $\alpha > 1$, then the Central Limit Theorem holds for all $\varphi \in \mathcal{C}_\beta$ with $\int \varphi d\nu = 0$, with $\sigma > 0$ if and only if $\varphi \circ F \neq \psi \circ F - \psi$ for any ψ.*

Remark. Theorem 1 and Theorem 3 II(b) have been proved in [**Y**]. We will repeat the proof of Theorem 1 for completeness and give a very different proof for Theorem 3 II(b). To my knowledge all the other results are new. For similar results in the Markov setting, see *e.g.* [**Pt**], [**TT**], [**I1**]; for the setting where F is a subshift of finite type and ν is a more general equilibrium measure, see [**R**], [**FL**], [**Po**].

2. Existence and properties of equilibrium

Proof of Theorem 1 [**Y**]. Let $m_0 = m|\Delta_0$. Our first step is to show that there is a finite (F^R)-invariant measure ν_0 on Δ_0 whose density has the desired regularity. Let $\mathcal{P}_0 = \eta|\Delta_0$, *i.e.* $\mathcal{P}_0$ is the partition of Δ_0 into $\{\Delta_{0,j}\}$. Consider $A \in \bigvee\limits_{j=0}^{i-1}(F^R)^{-j}\mathcal{P}_0$ and let $\rho_{i,A} = \frac{d}{dm}(F^R)^i_*(m|A)$. Let $x, y \in \Delta_0$ be arbitrary points, and let $x', y' \in A$ be s.t. $(F^R)^i x' = x$, $(F^R)^i y' = y$. Then for $j \le i$, $s((F^R)^j x',\ (F^R)^j y') = s(x,y) + (i-j)$, so that

$$\log\frac{\rho_{i,A}(y)}{\rho_{i,A}(x)} = \log\frac{J(F^R)^i x'}{J(F^R)^i y'} = \sum_{j=0}^{i-1}\log\frac{JF^R((F^R)^j x')}{JF^R((F^R)^j y')}$$
$$\le \sum_{j=0}^{i-1} C\beta^{s(x,y)+(i-j)-1} \le C'\beta^{s(x,y)}.$$

Let $\rho_n := \frac{d}{dm}\left(\frac{1}{n}\sum\limits_{i=0}^{n-1}(F^R)^i_* m_0\right)$. Since ρ_n is a linear combination of terms of the type $\rho_{i,A}$, our computation above shows that $\log\rho_n(y) \le C'\log\rho_n(x)$ for all $x, y \in \Delta_0$, and $\log\rho_n(y) \le \log\rho_n(x)\cdot C'\beta^k$ for all x, y belonging to the same element of $\bigvee\limits_{i=0}^{k-1}(F^R)^{-i}\mathcal{P}_0$, any $k > 0$. One checks easily that the sequence $\{\rho_n\}$ is relatively compact in $L^\infty(\Delta_0, m)$, and that any measure ν_0 whose density wrt m is a limit point of $\{\rho_n\}$ has the desired properties.

Let $\nu' = \sum\limits_{\ell=0}^{\infty} F^\ell_*(\nu_0|\{R > \ell\})$. Since $\frac{d\nu_0}{dm}$ is uniformly bounded, $\int R d_m < \infty \Rightarrow \nu'(\Delta) < \infty$. Normalize to give the desired probability measure ν. This proves (i). Part (ii) follows from the established regularity of $\frac{d\nu_0}{dm}$ since for $x \in \Delta_\ell$, $\frac{d\nu}{dm}(x) = \frac{d\nu}{dm}(\hat{x})$ where $\hat{x}$ is the point in Δ_0 with $F^\ell\hat{x} = x$.

The exactness of (F, ν) hinges on our assumption that $\gcd\{R_i\} = 1$. We begin with the following preliminary observation: From finite state Markov chain arguments, we know $\exists t_0' \in \mathbb{Z}^+$ s.t. $\Delta_0 \cap F^{-t}\Delta_0 \ne \emptyset\ \forall t \ge t_0'$, so for every $\ell_0 \in \mathbb{Z}^+$, $\exists\ t_0$ s.t. $F^{t_0}\Delta_0 \supset \bigcup\limits_{\ell \le \ell_0}\Delta_\ell$.

Recalling that $\mathcal{B}$ is the σ-algebra on Δ, we let $A \in \bigcap\limits_{n\ge 0} F^{-n}\mathcal{B}$ be s.t. $\nu(A) > 0$. We will show that $\nu(A) > 1 - \varepsilon$ for every pre-assigned $\varepsilon > 0$. Choose $t = t(\varepsilon)$ and $\delta = \delta(\varepsilon, t) > 0$ s.t. for all $B \in \mathcal{B}$ with $m(\Delta_0 - B) < \delta$, we have $m(F^t B) > 1 - \varepsilon$. Suppose for the moment that $m(\Delta_0 - F^n A) < \delta$ for some $n \in \mathbb{Z}^+$. Then, since $A = F^{-(n+t)}A'$ for some $A' \in \mathcal{B}$, we have $\nu(A) = \nu(A') = \nu(F^t(F^n A)) > 1 - \varepsilon$.

To produce an n with the property above, pick $C \in \bigvee\limits_{i=0}^{n-1} F^{-i}\eta$ with $F^n C = \Delta_0$ s.t. $m(A \cap C)/m(C)$ is arbitrarily near 1. Our distortion estimate earlier on then gives

$$\frac{m(F^n(A\cap C))}{m(\Delta_0)} \approx \frac{m(A\cap C)}{m(C)} \approx 1.$$

3. Speed of convergence to equilibrium

We assume throughout that $\int R dm < \infty$ and that Theorem 1 holds.

3.1. Lower bound.

Let λ be a probability measure on Δ with the property that $\frac{d\lambda}{dm} \geq \frac{d\nu}{dm} + c_1$ on $\underset{\ell\geq 1}{\cup} \Delta_\ell$ where $c_1 > 0$ is a small constant. Since $JF \equiv 1$ on $\Delta - F^{-1}\Delta_0$ and $F_*\nu = \nu$, we have, for every n, $\frac{d(F_*^n\lambda)}{dm} \geq \frac{d\nu}{dm} + c_1$ on $\underset{\ell>n}{\cup} \Delta_\ell$. Thus

$$|F_*^n\lambda - \nu| = \int \left| \frac{d(F_*^n\lambda)}{dm} - \frac{d\nu}{dm} \right| dm \geq c_1 \sum_{\ell>n} m(\Delta_\ell) = c_1 m\{\hat{R} > n\}$$

proving Theorem 2(I).

With this observation it is tempting to conjecture that the asymptotics of $m\{\hat{R} > n\}$ alone determine the speed of convergence. This, however, is clearly false. The simplest counterexample is when R is bounded, *i.e.* $m\{\hat{R} > n\} = 0$ for all large n, and $F : \Delta \circlearrowleft$ with $\nu = m$ is isomorphic to a finite state Markov chain for which the speed of convergence to equilibrium is well known to be not faster than exponential. A better guess, then, would be that the speed of convergence is not determined by the asymptotics of $m\{\hat{R} > n\}$ alone, but also by other exponential rates depending on the combinatorics of $m\{R = n\}$ and on the "nonlinearities" of F and $\frac{d\lambda}{dm}$. This in essence is what we are aiming to prove.

3.2. Upper bound: line of approach.

Let λ and λ' be probability measures on Δ with $\frac{d\lambda}{dm}, \frac{d\lambda'}{dm} \in \mathcal{C}_\beta^+$. We wish to estimate $|F_*^n\lambda - F_*^n\lambda'|$, and will do it by trying to match $F_*^n\lambda$ with $F_*^n\lambda'$ in the sense to be described below.

Formally, we consider the product transformation $F \times F : \Delta \times \Delta \circlearrowleft$. Let $P = \lambda \times \lambda'$, and let $\pi, \pi' : \Delta \times \Delta \to \Delta$ be projections onto the first and second coordinates. We will use frequently relations of the type $F^n \circ \pi = \pi \circ (F \times F)^n$. Consider the partition $\eta\times\eta$ on $\Delta\times\Delta$, and note that each element of $\eta\times\eta$ is mapped injectively onto a union of elements of $\eta \times \eta$. Let $(\eta \times \eta)_n := \bigvee_{i=0}^{n-1} (F \times F)^{-i}(\eta \times \eta)$ and let $(\eta \times \eta)_n(x, x')$ denote the element of $(\eta \times \eta)_n$ containing $(x, x') \in \Delta \times \Delta$.

Let $T : \Delta \times \Delta \to \mathbb{Z}^+$ be the first simultaneous return time to Δ_0, i.e. $T(x, x') =$ the smallest $n > 0$ s.t. $F^n x,\ F^n x' \in \Delta_0$. Observe that if $T(x, x') = n$, then $T|(\eta \times \eta)_n(x, x') \equiv n$ and $(F \times F)^n((\eta \times \eta)_n(x, x')) = \Delta_0 \times \Delta_0$.

Suppose for the moment that F is "linear" in the sense that JF is constant on each $\Delta_{\ell,i}$. Assume also that $\frac{d\lambda}{dm}$, $\frac{d\lambda'}{dm}$ are constant on each $\Delta_{\ell,i}$. Under these conditions, if $T(x,x')=n$, then

$$\begin{aligned}\pi_*(F\times F)^n_*(P|(\eta\times\eta)_n(x,x')) &= \frac{P((\eta\times\eta)_n(x,x'))}{m(\Delta_0)}(m|\Delta_0)\\ &= \pi'_*(F\times F)^n_*(P|(\eta\times\eta)_n(x,x')),\end{aligned}$$

and

$$\begin{aligned}|F^n_*\lambda - F^n_*\lambda'| \le\ & |\pi_*(F\times F)^n_*(P|\{T>n\}) - \pi'_*(F\times F)^n_*(P|\{T>n\})|\\ &+ \left|\sum_{i=1}^n F_*^{n-i}\left\{\pi_*(F\times F)^i_*(P|\{T=i\}) - \pi'_*(F\times F)^i_*(P|\{T=i\})\right\}\right|\\ &\le 2P\{T>n\}.\end{aligned}$$

What we have just described is a standard coupling argument for Markov chains said in the language of dynamical systems. Indeed, if F is "linear", (F,ν) is isomorphic to a countable state Markov chain, for which $2P\{T>n\}$ is well known to be an upper bound for the speed of convergence to its equilibrium state.

Returning to the general "nonlinear" situation, we do not have perfect matching at simultaneous returns to Δ_0, i.e. $\pi_*(F\times F)^n_*(P|(\eta\times\eta)_n(x,x')) \ne \pi'_*(F\times F)^n_*(P|(\eta\times\eta)_n(x,x'))$ when $T(x,x')=n$. However, if the initial densities are nice, and we have proper distortion control, then $\frac{d}{d(m\times m)}\left[(F\times F)^n_*(P|(\eta\times\eta)_n(x,x'))\right]$ should be quite regular. Suppose this density lies between c and $2c$ for some $c>0$. We could write $(F\times F)^n_*(P|(\eta\times\eta)_n(x,x'))$ as the sum of a measure of the form $\varepsilon c(m\times m)|(\Delta_0\times\Delta_0)$ for some small $\varepsilon>0$ and another (positive) measure, and think of the first part as having been "matched".

Let us introduce then a sequence of stopping times $T_1<T_2<T_3<\cdots$ defined by $T_1(x,x')=T(x,x')$ where T is as above, and $T_k = T\circ(F\times F)^{T_{k-1}}$ for $k>1$. At each T_k, a small fraction of the measure that reaches $\Delta_0\times\Delta_0$ is matched and is pumped out of the system as described in the last paragraph, and the total measure remaining in the system at time n is an upper bound for $|F^n_*\lambda - F^n_*\lambda'|$. Note that subtracting a constant from a density may cause some deterioration in its distortion estimates, but hopefully all is restored by the next simultaneous return time.

We have described the relation between $|F^n_*\lambda - F^n_*\lambda'|$ and $P\{T>n\}$. In a separate argument it will be shown that $P\{T>n\}$ is quite naturally related to $m\{\hat{R}>n\}$. These two steps will be carried out in 3.3 and 3.4.

3.3. A simultaneous return time and its relation to $\hat{R}$.

The purpose of this subsection is to introduce a stopping time T that is a simultaneous return time of F to Δ_0, or equivalently, a return time of $F\times F$ to $\Delta_0\times\Delta_0$, and to estimate $P\{T > n\}$. It is not necessary that T be the first simultaneous return time as suggested in 3.2; indeed it is probably advantageous to select a T that relates naturally to $m\{\hat{R} > n\}$.

Recall that for $x \in \Delta$, $\hat{R}(x)$ is the smallest $n \geq 0$ such that $F^n x \in \Delta_0$. First we introduce an auxiliary sequence of stopping times $0 \equiv \tau_0 < \tau_1 < \tau_2 < \cdots$ on $\Delta \times \Delta$ defined as follows. Let $n_0 \in \mathbb{Z}^+$ be s.t. $m(F^{-n}\Delta_0 \cap \Delta_0) \geq$ some $\gamma_0 > 0$ for all $n \geq n_0$. The existence of n_0 follows from the mixing property of (F, ν) and the fact that $\frac{d\nu}{dm} \in L^\infty(m)$. We let

$$\begin{aligned}
\tau_1(x, x') &= n_0 + \hat{R}\left(F^{n_0} x\right), \\
\tau_2(x, x') &= \tau_1 + n_0 + \hat{R}\left(F^{\tau_1 + n_0} x'\right), \\
\tau_3(x, x') &= \tau_2 + n_0 + \hat{R}\left(F^{\tau_2 + n_0} x\right), \\
\tau_4(x, x') &= \tau_3 + n_0 + \hat{R}\left(F^{\tau_3 + n_0} x'\right),
\end{aligned}$$

and so on, with the action alternating between x and x'. Notice that had we not put in a time delay n_0, the purpose of which will become clear shortly, $\tau_i - \tau_{i-1}$ would have been the first return time to Δ_0 of $F^{\tau_{i-1}} x$ or $F^{\tau_{i-1}} x'$ depending on whether i is odd or even. Define $T = \tau_i$ where i is the smallest integer ≥ 2 with the property that both $F^{\tau_i} x$ and $F^{\tau_i} x'$ are in Δ_0. Since (F, ν) is mixing, $(F \times F,\ \nu \times \nu)$ is ergodic and T is defined $(m \times m)$-a.e.

Let $\xi_1 < \xi_2 < \xi_3 < \cdots$ be an increasing sequence of partitions on $\Delta \times \Delta$ defined as follows. First, $\xi_1(x, x') = (\bigvee\limits_{j=0}^{\tau_1 - 1} F^{-j}\eta)(x) \times \Delta$; that is to say, the elements of ξ_1 are sets of the form $\Gamma = A \times \Delta$ where τ_1 is constant on Γ and F^{τ_1} maps A injectively onto Δ_0. For $i > 1$, if i is odd (resp. even), define ξ_i to be the refinement of ξ_{i-1} obtained by partitioning each $\Gamma \in \xi_{i-1}$ in the x-direction (resp. x'-direction) into sets $\tilde{\Gamma}$ in such a way that τ_i is constant on each $\tilde{\Gamma}$ and F^{τ_i} maps $\pi\tilde{\Gamma}$ (resp. $\pi'\tilde{\Gamma}$) injectively onto Δ_0. Note that τ_i is measurable wrt ξ_i.

Let us focus more closely on $\Gamma \in \xi_i$, assuming for definiteness that i is even and is ≥ 2. Note that $\tau_1, \tau_2, \cdots, \tau_i$ are constant on Γ. For definiteness assume also that $\Gamma \cap \{T \leq \tau_{i-1}\} = \emptyset$. Observe that Γ is a "rectangle", i.e. $\Gamma = A \times B$ for some $A, B \subset \Delta$. At time τ_{i-1}, $F^{\tau_{i-1}} A = \Delta_0$ and $F^{\tau_{i-1}} B$ is contained in some $\Delta_{\ell,j}$. At time τ_i, $F^{\tau_i} B = \Delta_0$ while $F^{\tau_i} A$ is spread over various parts of $\cup\{\Delta_\ell, \ell \leq \tau_i - \tau_{i-1}\}$. Our definition of T requires that we set $T = \tau_i$ on those parts of Γ whose π-projections at time τ_i lie in Δ_0. Our first lemma will deal with what proportion of Γ this comprises. To define τ_{i+1} at $(x, x') \in \Gamma$, we look at $F^{\tau_i} x$, iterate blindly n_0 times, and let τ_{i+1} be the first return time to Δ_0 after that. Clearly, τ_{i+1} is constant on sets of the form $\Gamma \cap \pi^{-1}\{x\}$ and could be arbitarily large in value. The distribution of $\tau_{i+1} - \tau_i$ on Γ will be the subject of Lemma 2. Observe that $\xi_{i+1}|\Gamma$ partitions Γ into countably many "vertical" strips, and that $\{T = \tau_i\}$ is measurable wrt ξ_{i+1} but not ξ_i.

We now state our two main estimates for $\{\tau_i\}$ and T. Each estimate will come in 2 versions. One holds for all times; its constants depend, unavoidably, on the regularity of λ and λ'. One of the properties of F is that as we iterate, the roughness of the initial data gets washed out. The second version holds only from that point on; its constants are independent of λ or λ'.

Lemma 1. *$\exists \varepsilon_0 = \varepsilon_0(\lambda, \lambda') > 0$ s.t. $\forall i \geq 2$ and $\forall \Gamma \in \xi_i$ with $T|\Gamma > \tau_{i-1}$,*

$$P\{T = \tau_i \mid \Gamma\} \geq \varepsilon_0;$$

the dependence of ε_0 on λ and λ' can be removed if we consider only $i \geq$ some $i_0 = i_0(\lambda, \lambda')$.

Let ξ_0 denote the trivial partition $\{\Delta \times \Delta\}$ and recall that $\tau_0 \equiv 0$.

Lemma 2. *$\exists K_0 = K_0(\lambda, \lambda')$ s.t. $\forall i \geq 0$, $\forall \Gamma \in \xi_i$ and $\forall n \geq 0$,*

$$P\{\tau_{i+1} - \tau_i > n_0 + n \mid \Gamma\} \leq K_0 m\{\hat{R} > n\};$$

the dependence of K_0 on λ and λ' can be removed if we consider only $i \geq$ some $i_0 = i_0(\lambda, \lambda')$.

We begin with some sublemmas. First we record an easy fact already established in the proof of Theorem 1. Recall that $\mathcal{P}_0$ is the partition of Δ_0 into $\{\Delta_{0,i}\}$. Let $\mathcal{P}_n := \bigvee_{i=0}^{n-1} (F^R)^{-i}\mathcal{P}_0$. Then it follows easily from condition (*) in 1.1 that there exists a constant $C_F > 0$ with the property that for all $n \in \mathbb{Z}^+$ and for all x, y belonging in the same element of $\mathcal{P}_n$,

$$\left| \frac{J(F^R)^n(x)}{J(F^R)^n(y)} - 1 \right| \leq C_F \beta^{s((F^R)^n x,\ (F^R)^n y)}.$$

Sublemma 1. *$\exists M_0$ s.t. $\forall n \in \mathbb{Z}^+$,*

$$\frac{dF_*^n m}{dm} \leq M_0.$$

Proof. Let $\mu_n = F_*^n m$. Since $\mu_n(\Delta) \leq m(\Delta) < \infty$, it follows from the distortion estimate above that $\frac{d\mu_n}{dm}|\Delta_0 \leq$ some M_0 $\forall n \geq 0$. The rest follows since $\frac{d\mu_n}{dm}|\Delta_\ell = 1$ for $\ell \geq n$ and comes from $\frac{d\mu_{n-\ell}}{dm}|\Delta_0$ for $\ell < n$. ∎

Recall that η is the partition of Δ into $\{\Delta_{\ell,j}\}$.

Sublemma 2. *For arbitrary $k > 0$, let $\Omega \in \bigvee_{i=0}^{k-1} F^{-i}\eta$ be s.t. $F^k\Omega = \Delta_0$, and let $\mu = F_*^k(\lambda|\Omega)$. Then $\forall x, y \in \Delta_0$, we have*

$$\left| \frac{\frac{d\mu}{dm}(x)}{\frac{d\mu}{dm}(y)} - 1 \right| \leq C_0$$

for some $C_0 = C_0(\lambda)$. The dependence of C_0 on λ can be removed if we assume that the number of $i \le k$ such that $F^i\Omega \subset \Delta_0$ is greater than some $j_0 = j_0(\lambda)$.

Proof. Let $\varphi = \frac{d\lambda}{dm}$, and let $x_0, y_0 \in \Omega$ be s.t. $F^k x_0 = x,\ F^k y_0 = y$. Then

$$\begin{aligned}
\left| \frac{\varphi x_0}{JF^k x_0} \Big/ \frac{\varphi y_0}{JF^k y_0} - 1 \right| &= \frac{JF^k y_0}{\varphi y_0} \left| \frac{\varphi x_0}{JF^k x_0} - \frac{\varphi y_0}{JF^k y_0} \right| \\
&\le \frac{JF^k y_0}{\varphi y_0} \left\{ \varphi x_0 \left| \frac{1}{JF^k x_0} - \frac{1}{JF^k y_0} \right| + \frac{1}{JF^k y_0} \left| \varphi x_0 - \varphi y_0 \right| \right\} \\
&\le \frac{\varphi x_0}{\varphi y_0} \cdot \left| \frac{JF^k y_0}{JF^k x_0} - 1 \right| + \left| \frac{\varphi x_0}{\varphi y_0} - 1 \right| \\
&\le (1 + C\beta^j) C_F + C\beta^j .
\end{aligned}$$

Here C is the "Hölder" constant for φ and j is the number of visits to Δ_0 prior to time k. ■

Proof of Lemma 1. Assume for definiteness that i is even. Let $\Gamma \in \xi_i$ be as in the lemma, and let $\Omega = \pi(\Gamma)$. Since $P = \lambda \times \lambda',\ \pi_*(P|\Gamma) = \text{const} \cdot (\lambda|\Omega)$, so that Sublemma 2 applies to $\mu = F_*^{\tau_{i-1}}(\lambda|\Omega)$. Now

$$P\{T = \tau_i \mid \Gamma\} = \frac{1}{\mu(\Delta_0)} \cdot \mu \left(\Delta_0 \cap F^{-(\tau_i - \tau_{i-1})} \Delta_0 \right),$$

so Lemma 1 with $\varepsilon_0 = \varepsilon_0(\lambda, \lambda')$ follows from our distortion estimate for $\frac{d\mu}{dm}$, our choice of n_0 and the requirement that $\tau_i - \tau_{i-1} \ge n_0$. For $i \ge 2j_0$ where j_0 is as in Sublemma 2, the distortion of $\frac{d\mu}{dm}$ and hence a lower bound on the μ-measure of the part of Δ_0 that returns at time $\tau_i - \tau_{i-1}$ is independent of λ or λ'.

■

Proof of Lemma 2. The cases $i = 0, 1$ are a little different and will be dealt with later. Consider $i \ge 2$ and assume again for definiteness that i is even. Let $\mu = \frac{1}{P(\Gamma)} F_*^{\tau_{i-1}} \pi_*(P|\Gamma)$. Then μ is a probability measure on Δ_0, and

$$\begin{aligned}
P\{\tau_{i+1} - \tau_i > n_0 + n \mid \Gamma\} &= \left(F_*^{(\tau_i - \tau_{i-1}) + n_0} \mu \right) \{\hat{R} > n\} \\
&\le \left| \frac{d}{dm} \left(F_*^{(\tau_i - \tau_{i-1}) + n_0} \mu \right) \right|_\infty m\{\hat{R} > n\} \\
&\le M_0 \left| \frac{d\mu}{dm} \right|_\infty m\{\hat{R} > n\} \qquad \text{by Sublemma 1.}
\end{aligned}$$

Note that by Sublemma 2, $\left| \frac{d\mu}{dm} \right|_\infty$ is bounded above by a constant independent of Γ and possibly depending on λ only for the initial i's. This completes the argument for $i \ge 2$. For $i = 0$, $P\{\tau_1 > n_0 + n\} = (F_*^{n_0} \lambda)\{\hat{R} > n\} \le \left| \frac{d\lambda}{dm} \right|_\infty M_0\, m\{\hat{R} > n\}$; $i = 1$ is treated similarly. ■

3.4. Matching $F_*^n\lambda$ with $F_*^n\lambda'$.

The relevant dynamical system in this second half of the scheme is $\hat F : \Delta\times\Delta \circlearrowleft$ defined by $\hat F = (F\times F)^T$. That is to say, if $\hat\xi_1$ denotes the partition of $\Delta\times\Delta$ into rectangles Γ on which T is constant and $(F\times F)^T$ maps Γ injectively onto $\Delta_0\times\Delta_0$, then $\hat F|\Gamma \stackrel{\text{def}}{=} (F\times F)^T|\Gamma$. Here the reference measure is $m\times m$, and $J\hat F$ refers to the Jacobian of $\hat F$ wrt $m\times m$. Associated with $\hat F$ is a separation time $\hat s(\cdot,\cdot)$ defined as follows: For $w,z\in\Delta\times\Delta$,

$\hat s(w,z) :=$ the smallest $n\ge 0$ s.t. $\hat F^n w$ and $\hat F^n z$ lie in distinct elements of $\hat\xi_1$.

Before proceeding further we verify the following entirely expected relation between $\hat s(\cdot,\cdot)$ and $s(\cdot,\cdot)$. Let $w=(x,x')$ and $z=(y,y')$. We claim that $\hat s(w,z) > n \Rightarrow s(x,y),\ s(x',y')>n$. To see this, observe first that every $\Gamma\in\hat\xi_1|(\Delta_0\times\Delta_0)$ must be contained in $\Delta_{0,j}\times\Delta_{0,j'}$, for some j,j', otherwise $(F\times F)^T$ cannot map Γ injectively onto $\Delta_0\times\Delta_0$. Suppose $\hat s(w,z)>n$, and let k be s.t. $\hat F^n w = (F\times F)^k w$. Let $I=\{i\le k : (F\times F)^i w\in\Delta_0\times\Delta_0\}$. Then $\text{card}(I)\ge n$ and for $i\in I$, $(F\times F)^i z\in\Delta_0\times\Delta_0$ as well. Moreover, $\forall i\in I$, $\exists\, j=j(i),\ j'=j'(i)$ s.t. $(F\times F)^i w,\ (F\times F)^i z\in\Delta_{0,j}\times\Delta_{0,j'}$. This proves that $s(x,y),\ s(x',y')>n$.

Let $\varphi=\frac{d\lambda}{dm}$, $\varphi'=\frac{d\lambda'}{dm}$, and let C_φ and $C_{\varphi'}$ be constants s.t. $\forall x,y\in\Delta$,

$$\left|\log\frac{\varphi x}{\varphi y}\right| \le C_\varphi\beta^{s(x,y)}, \qquad \left|\log\frac{\varphi' x}{\varphi' y}\right| \le C_{\varphi'}\beta^{s(x,y)}.$$

(This of course makes sense only when $\varphi x,\varphi y>0$.) Let $\Phi=\frac{dP}{d(m\times m)}$, i.e. $\Phi(x,x')=\varphi(x)\varphi'(x')$. We record the following easy facts regarding the regularity of $J\hat F$ and Φ.

Sublemma 3. 1. *$\forall w,z\in\Delta\times\Delta$ with $\hat s(w,z)\ge n$, any $n>0$,*

$$\left|\log\frac{J\hat F^n(w)}{J\hat F^n(z)}\right| \le C_{\hat F}\beta^{\hat s(\hat F^n w,\hat F^n z)}$$

where $C_{\hat F}$ can be taken to be $2C_F$;

2. *$\forall w,z\in\Delta\times\Delta$,*

$$\left|\log\frac{\Phi(w)}{\Phi(z)}\right| \le C_\Phi\beta^{\hat s(w,z)}$$

where $C_\Phi = C_\varphi + C_{\varphi'}$.

Proof. Let $w=(x,x'),\ z=(y,y')$, and let k be s.t. $\hat F^n(w)=(F\times F)^k(w)$. Then

$$\begin{aligned}\left|\log\frac{J\hat F^n(x,x')}{J\hat F^n(y,y')}\right| &\le \left|\log\frac{J\hat F^n(x,x')}{J\hat F^n(y,x')}\right| + \left|\log\frac{J\hat F^n(y,x')}{J\hat F^n(y,y')}\right|\\ &= \left|\log\frac{JF^k(x)JF^k(x')}{JF^k(y)JF^k(x')}\right| + \left|\log\frac{JF^k(y)JF^k(x')}{JF^k(y)JF^k(y')}\right|\\ &\le C_F\beta^{s(F^kx,F^ky)} + C_F\beta^{s(F^kx',F^ky')}\\ &\le 2C_F\beta^{\hat s(\hat F^n w,\hat F^n z)}.\end{aligned}$$

The second assertion is proved similarly. ■

We now describe the procedure through which the "matching" is done. Let $T_1 < T_2 < \cdots$ be stopping times on $\Delta \times \Delta$ defined by

$$T_1 = T; \qquad T_n = T_{n-1} + T \circ \hat{F}^{n-1} \qquad \text{for} \qquad n > 1.$$

Note that $\hat{F}^n = (F \times F)^{T_n}$. Let $\hat{\xi}_n := \hat{F}^{-(n-1)}\hat{\xi}_1$, so that $\hat{\xi}_n$ is the partition whose elements Γ have the property that T_n is constant on Γ and $(F \times F)^{T_n}$ maps Γ injectively onto $\Delta_0 \times \Delta_0$. Given $\Phi = \frac{dP}{d(m \times m)}$, we will introduce a decreasing sequence of densities $\hat{\Phi}_0 \geq \hat{\Phi}_1 \geq \hat{\Phi}_2 \geq \cdots$ in such a way that for all i and for all $\Gamma \in \hat{\xi}_i$,

$$\pi_* \hat{F}^i_*((\hat{\Phi}_{i-1} - \hat{\Phi}_i)((m \times m)|\Gamma)) = \pi'_* \hat{F}^i_*((\hat{\Phi}_{i-1} - \hat{\Phi}_i)((m \times m)|\Gamma)). \tag{1}$$

That is to say, $\hat{\Phi}_i|\Gamma$ is the density of the part of $P|\Gamma$ that has not yet been "matched" after time T_i.

The $\hat{\Phi}_i$'s are defined as follows. Let $\varepsilon > 0$ be a small number to be determined later; ε will depend on F (on β, to be precise) but not on Φ. Let $i_1 = i_1(\Phi)$ be s.t. $C_\Phi \beta^{i_1} < C_{\hat{F}}$. For $i < i_1$, let $\hat{\Phi}_i \equiv \Phi$; that is, no attempt is made to match the measures before time T_{i_1}. For $i \geq i_1$, let

$$\hat{\Phi}_i(z) = \left[\frac{\hat{\Phi}_{i-1}(z)}{J\hat{F}^i(z)} \quad - \quad \varepsilon \cdot \min_{w \in \hat{\xi}_i(z)} \frac{\hat{\Phi}_{i-1}(w)}{J\hat{F}^i(w)} \right] \cdot J\hat{F}^i(z).$$

It is easily seen that $\{\hat{\Phi}_i\}$ has property (1) above. The main result of this subsection is

Lemma 3. *For all sufficiently small $\varepsilon > 0$, $\exists \varepsilon_1 > 0$ independent of Φ s.t. for all $i \geq i_1$,*

$$\hat{\Phi}_i \leq (1 - \varepsilon_1)\hat{\Phi}_{i-1} \qquad \text{on all of} \quad \Delta \times \Delta.$$

To prove Lemma 3, it suffices to show that if ε is chosen sufficiently small, then there exists a constant C s.t. for all $\Gamma \in \hat{\xi}_i$,

$$\max_{w \in \Gamma} \frac{\hat{\Phi}_{i-1}(w)}{J\hat{F}^i(w)} \Big/ \min_{w \in \Gamma} \frac{\hat{\Phi}_{i-1}(w)}{J\hat{F}^i(w)} \leq C.$$

To prove this distortion estimate, it is more convenient to work directly with the densities of the pushed forward measures corresponding to the $\hat{\Phi}_i$'s. We introduce some new notations for this purpose: For $z \in \Delta \times \Delta$ let

$$\tilde{\Psi}_{i_1-1,z} = \frac{\Phi(z)}{J\hat{F}^{i_1-1}(z)},$$

and for $i \geq i_1$, let

$$\begin{aligned}
\Psi_{i,z} &= \frac{\tilde{\Psi}_{i-1,z}}{J\hat{F}(\hat{F}^{i-1}z)}, \\
\varepsilon_{i,z} &= \varepsilon \cdot \min_{w \in \xi_i(z)} \Psi_{i,w}, \\
\tilde{\Psi}_{i,z} &= \Psi_{i,z} - \varepsilon_{i,z}.
\end{aligned}$$

Lemma 3 follows immediately from Lemma 3′.

Lemma 3′. *There exists $\hat{C}$ such that the following holds for all sufficiently small ε: $\forall w, z \in \Delta \times \Delta$ with $w \in \hat{\xi}_i(z)$ and $\forall i \geq i_1$,*

$$\left| \log \frac{\tilde{\Psi}_{i,w}}{\tilde{\Psi}_{i,z}} \right| \leq \hat{C} \beta^{\hat{s}(\hat{F}^i w, \hat{F}^i z)}.$$

Proof. We break the argument up into several steps.
(1)

$$\begin{aligned}
\left| \log \frac{\Psi_{i,w}}{\Psi_{i,z}} \right| &\leq \left| \log \frac{\tilde{\Psi}_{i-1,w}}{\tilde{\Psi}_{i-1,z}} \right| + \left| \log \frac{J\hat{F}(\hat{F}^{i-1}z)}{J\hat{F}(\hat{F}^{i-1}w)} \right| \\
&\leq \left| \log \frac{\tilde{\Psi}_{i-1,w}}{\tilde{\Psi}_{i-1,z}} \right| + C_{\hat{F}} \beta^{\hat{s}(\hat{F}^i z, \hat{F}^i w)}.
\end{aligned}$$

(2) Let $\varepsilon' > 0$ be given and fixed. It is obvious that if $\varepsilon > 0$ is sufficiently small and is allowed to depend on i, w and z, then

$$\left| \log \frac{\tilde{\Psi}_{i,w}}{\tilde{\Psi}_{i,z}} \right| \leq (1+\varepsilon') \left| \log \frac{\Psi_{i,w}}{\Psi_{i,z}} \right|.$$

We make the dependence of the various quantities in this relation more transparent for use in a later step. Writing $\varepsilon_i = \varepsilon_{i,z} = \varepsilon_{i,w}$, we have

$$\begin{aligned}
\left| \log \frac{\tilde{\Psi}_{i,w}}{\tilde{\Psi}_{i,z}} - \log \frac{\Psi_{i,w}}{\Psi_{i,z}} \right| &= \left| \log \frac{\Psi_{i,w} - \varepsilon_i}{\Psi_{i,w}} \cdot \frac{\Psi_{i,z}}{\Psi_{i,z} - \varepsilon_i} \right| \\
&= \left| \log \left(1 + \frac{\frac{\varepsilon_i}{\Psi_{i,z}} - \frac{\varepsilon_i}{\Psi_{i,w}}}{1 - \frac{\varepsilon_i}{\Psi_{i,z}}} \right) \right| \\
&\leq C_1 \left| \frac{\frac{\varepsilon_i}{\Psi_{i,z}} - \frac{\varepsilon_i}{\Psi_{i,w}}}{1 - \frac{\varepsilon_i}{\Psi_{i,z}}} \right| \\
&= C_1 \frac{\varepsilon_i}{\Psi_{i,w}} \cdot \left| \frac{\Psi_{i,w}}{\Psi_{i,z}} - 1 \right| \cdot \frac{1}{1 - \frac{\varepsilon_i}{\Psi_{i,z}}} \\
&\leq C_1 \varepsilon \cdot C_2 \left| \log \frac{\Psi_{i,w}}{\Psi_{i,z}} \right| \cdot \frac{1}{1-\varepsilon}.
\end{aligned}$$

Choosing ε small enough so that $C_1 C_2 \frac{\varepsilon}{1-\varepsilon} \le \varepsilon'$, we obtain the desired result. Note the dependences of C_1 and C_2 above. Assuming that $\varepsilon < \frac{1}{4}$, the quantity $*$ in $|\log(1+*)|$ above is $\ge -\frac{1}{3}$, so C_1 does not depend on anything. Observe, however, that C_2 increases as $\Psi_{i,w}/\Psi_{i,z}$ increases; and the larger C_2, the smaller ε will have to be.

(3) Letting ε' be given and assuming that ε is sufficiently small as required, we combine (1) and (2) to obtain the recursive relation

$$\left|\log \frac{\tilde{\Psi}_{i,w}}{\tilde{\Psi}_{i,z}}\right| \le (1+\varepsilon') \left\{ \left|\log \frac{\tilde{\Psi}_{i-1,w}}{\tilde{\Psi}_{i-1,z}}\right| + C_{\hat{F}} \beta^{\hat{s}(\hat{F}^i w, \hat{F}^i z)} \right\}.$$

Also,

$$\begin{aligned} \left|\log \frac{\tilde{\Psi}_{i_1,w}}{\tilde{\Psi}_{i_1,z}}\right| &\le (1+\varepsilon') \left\{ \left|\log \frac{\Phi(w)}{\Phi(z)}\right| + \left|\log \frac{J\hat{F}^{i_1}(z)}{J\hat{F}^{i_1}(w)}\right| \right\} \\ &\le (1+\varepsilon') \left\{ C_\Phi \beta^{\hat{s}(w,z)} + C_{\hat{F}} \beta^{\hat{s}(\hat{F}^{i_1} w, \hat{F}^{i_1} z)} \right\} \\ &\le (1+\varepsilon') \cdot 2 C_{\hat{F}} \beta^{\hat{s}(\hat{F}^{i_1} w, \hat{F}^{i_1} z)} \end{aligned}$$

by our choice of i_1.

(4) It follows from (3) and the relation $\hat{s}(\hat{F}^{i-j} w, \hat{F}^{i-j} z) = \hat{s}(\hat{F}^i w, \hat{F}^i z) + j$ that

$$\begin{aligned} \left|\log \frac{\tilde{\Psi}_{i,w}}{\tilde{\Psi}_{i,z}}\right| &\le (1+\varepsilon') C_{\hat{F}} \beta^{\hat{s}(\hat{F}^i w, \hat{F}^i z)} \ \{1 + (1+\varepsilon')\beta + (1+\varepsilon')^2 \beta^2 + \cdots \\ &\qquad + (1+\varepsilon')^{i-i_1-1} \beta^{i-i_1-1} + 2(1+\varepsilon')^{i-i_1} \beta^{i-i_1}\} \\ &\le \hat{C} \beta^{\hat{s}(\hat{F}^i w, \hat{F}^i z)} \end{aligned}$$

where $\hat{C} := 2(1+\varepsilon') C_{\hat{F}} \sum_{j=0}^{\infty} [(1+\varepsilon')\beta]^j$ provided ε' is chosen small enough that $(1+\varepsilon')\beta < 1$.

(5) In this final step we observe that ε can in fact be chosen independent of i, w or z. To see this, let $\varepsilon > 0$ be small enough that the estimate in (4) holds for all $i < j$ for some j for all w, z with $w = \hat{\xi}_i(z)$. Then by (1),

$$\left|\log \frac{\Psi_{j,w}}{\Psi_{j,z}}\right| \le \hat{C} + C_{\hat{F}},$$

which puts $\Psi_{j,w}/\Psi_{j,z} \in [e^{-(\hat{C}+C_{\hat{F}})}, e^{\hat{C}+C_{\hat{F}}}]$. This in turn imposes an upper bound on C_2 in the last line of the computation in (2). Provided that ε is small enough for $C_1 C_2 \frac{\varepsilon}{1-\varepsilon} \le \varepsilon'$, the estimates in (3), and hence in (4), will hold for $i = j$. ∎

Lemma 4. *For all $n \in \mathbb{Z}^+$,*

$$|F_*^n \lambda - F_*^n \lambda'| \leq 2P\{T_{i_1} > n\} + 2\sum_{i=i_1}^{\infty} (1-\varepsilon_1)^{i-i_1+1} P\{T_i \leq n < T_{i+1}\}$$

where $\varepsilon_1 > 0$ is as in Lemma 3.

Proof. The densities $\hat{\Phi}_i$ are those of the total measures remaining in the system after i iterates of $\hat{F}$. We must now bring these estimates back to "real time". Let $\Phi_0, \Phi_1, \Phi_2, \ldots$ be defined as follows: For $z \in \Delta \times \Delta$, let

$$\Phi_{T_i(z)}(z) = \hat{\Phi}_i(z),$$

$$\Phi_n(z) = \Phi_{T_i(z)}(z) \qquad \text{for} \qquad T_i(z) < n < T_{i+1}(z).$$

Claim: $|F_*^n \lambda - F_*^n \lambda'| \leq 2 \int \Phi_n d(m \times m)$.

To see this, write $\Phi = \Phi_n + \sum_{k=1}^{n} (\Phi_{k-1} - \Phi_k)$, so that

$$\begin{aligned}
&|F_*^n \lambda - F_*^n \lambda'| \\
&= |\pi_*(F \times F)_*^n(\Phi(m \times m)) - \pi'_*(F \times F)_*^n(\Phi(m \times m))| \\
&\leq |\pi_*(F \times F)_*^n(\Phi_n(m \times m)) - \pi'_*(F \times F)_*^n(\Phi_n(m \times m))| \\
&\quad + \sum_{k=1}^{n} |(\pi_* - \pi'_*)\left[(F \times F)_*^n((\Phi_{k-1} - \Phi_k)(m \times m))\right]|.
\end{aligned}$$

The first term is $\leq 2 \int \Phi_n d(m \times m)$. To see that all the other terms vanish, let $A_k = \cup A_{k,i}$ where $A_{k,i} = \{z \in \Delta \times \Delta : k = T_i(z)\}$. Clearly, $A_{k,i}$ is a union of elements of $\hat{\xi}_i$, and for $i \neq i'$, $A_{k,i} \cap A_{k,i'} = \emptyset$. We observe that for $\Gamma \in \hat{\xi}_i | A_{k,i}$, $\Phi_{k-1} - \Phi_k = \hat{\Phi}_{i-1} - \hat{\Phi}_i$; whereas on $(\Delta \times \Delta) - A_k$, $\Phi_{k-1} \equiv \Phi_k$. We therefore have for each k:

$$\begin{aligned}
&\pi_*(F \times F)_*^n((\Phi_{k-1} - \Phi_k)(m \times m)) \\
&= \sum_i \sum_{\Gamma \subset A_{k,i}} F_*^{n-k} \pi_*(F \times F)_*^{T_i}((\hat{\Phi}_{i-1} - \hat{\Phi}_i)((m \times m)|\Gamma)) \\
&= \sum_i \sum_{\Gamma \subset A_{k,i}} F_*^{n-k} \pi'_*(F \times F)_*^{T_i}((\hat{\Phi}_{i-1} - \hat{\Phi}_i)((m \times m)|\Gamma)) \\
&= \pi'_*(F \times F)_*^n((\Phi_{k-1} - \Phi_k)(m \times m)).
\end{aligned}$$

The second equality above uses Equation (1), which along with Lemma 3 are the two main properties of $\hat{\Phi}_i$. This completes the proof of the claim.

To finish the proof of Lemma 4, write

$$\int \Phi_n = \int_{\{n<T_{i_1}\}} \Phi_n + \sum_{i=i_1}^{\infty} \int_{\{T_i \le n < T_{i+1}\}} \Phi_n$$

and observe that

$$\int_{\{n<T_{i_1}\}} \Phi_n \;=\; \int_{\{n<T_{i_1}\}} \Phi \;=\; P\{n < T_{i_1}\}$$

while for $i \ge i_1$,

$$\int_{\{T_i \le n < T_{i+1}\}} \Phi_n = \int_{\{T_i \le n < T_{i+1}\}} \hat{\Phi}_i \le \int_{\{T_i \le n < T_{i+1}\}} (1-\varepsilon_1)^{i-i_1+1}\Phi.$$

■

We finish with the following easy fact which will be used for estimating the right side of the inequality in Lemma 4 in the next section.

Sublemma 4. *$\exists K_1 = K_1(P)$ s.t. $\forall i$ and $\forall \Gamma \in \hat{\xi}_i$,*

$$P\{T_{i+1} - T_i > n \mid \Gamma\} \le K_1 (m \times m)\{T > n\}.$$

The dependence of K_1 on P can be removed if we consider only $i \ge i(P)$.

Proof. The distortion estimate for $\hat{F}^i|\Gamma$ guarantees that $\frac{d}{d(m\times m)} \hat{F}^i_* P \le K_1$ for large enough i . ■

3.5. Summary of discussion.

The goal of Section 3 is to establish a relation between the two sequences $|F^n_*\lambda - F^n_*\lambda'|$ and $m\{\hat{R} > n\}$ without any assumptions on the latter. We do this by considering $F \times F : \Delta \times \Delta \circlearrowleft$ and using as an intermediate object a return time T to $\Delta_0 \times \Delta_0$. Let $P = \lambda \times \lambda'$. Then

(1) T is related to $m\{\hat{R} > n\}$ as follows: There is an auxiliary sequence of stopping times $0 \equiv \tau_0 < \tau_1 < \tau_2 < \cdots$ on $\Delta \times \Delta$ such that $T = \tau_i$ for some $i = i(x, x') \ge 2$ and
   - (a) $P\{\tau_{i+1} - \tau_i > n + n_0 \mid \tau_i\} \le K_0 m\{\hat{R} > n\}$;
   - (b) $P\{T = \tau_{i+1} \mid T > \tau_i\} \ge \varepsilon_0 > 0$;

   n_0 is a constant depending only on F; K_0 and ε_0 also depend on P, but this dependence can be removed if we consider only $i \ge$ some $i_0 = i_0(P)$.

(2) T is related to $|F^n_*\lambda - F^n_*\lambda'|$ as follows: Let $T_1 = T$, and $T_n = T_{n-1} + T \circ (F \times F)^{T_{n-1}}$ for $n > 1$. Then

$$|F^n_*\lambda - F^n_*\lambda'| \le C \sum_{i=1}^{\infty} (1-\varepsilon_1)^i P\{T_i \le n < T_{i+1}\}$$

   for some $\varepsilon_1 > 0$ depending only on F.

4. Some specific convergence rates

The purpose of Section 4 is to apply the results of Section 3 to some special cases. Among the standard decay rates observed or studied in dynamical systems are exponential, stretched exponential and polynomial speeds of decay.

4.1. Polynomial decay: Proof of Theorem 2 II(a).

We assume in this subsection that $m\{\hat{R} > n\} = \mathcal{O}(n^{-\alpha})$ for some $\alpha > 0$ and will show for all λ, λ' satisfying the conditions in Theorem 2 that $|F_*^n \lambda - F_*^n \lambda| = \mathcal{O}(n^{-\alpha})$. Throughout this section we let C denote a generic constant which is allowed to depend on F, λ and λ' but not on n or the iterate in question.

We begin by estimating $P\{T > n\}$. Write $P\{T > n\} = \text{(I)} + \text{(II)}$ where

$$\text{(I)} = \sum_{i \le \frac{1}{2}\left[\frac{n}{n_0}\right]} P\{T > n; \tau_{i-1} \le n < \tau_i\},$$

$$\text{(II)} = P\left\{T > n; \tau_{\frac{1}{2}\left[\frac{n}{n_0}\right]} \le n\right\}.$$

First, we observe that $\text{(II)} \le C(1-\varepsilon_0)^{\frac{1}{2}\left[\frac{n}{n_0}\right]}$ where ε_0 is as in Lemma 1. This is because for $n \ge 4n_0$,

$$\begin{aligned}\text{(II)} &\le P\left\{T > \tau_{\frac{1}{2}\left[\frac{n}{n_0}\right]}\right\} \\ &= P\{T > \tau_2\} P\{T > \tau_3 \mid T > \tau_2\} \cdots P\left\{T > \tau_{\frac{1}{2}\left[\frac{n}{n_0}\right]} \mid T > \tau_{\frac{1}{2}\left[\frac{n}{n_0}\right]-1}\right\}\end{aligned}$$

and each one of these factors is $\le (1-\varepsilon_0)$ by Lemma 1.

Before we begin on (I), observe that for $k \ge 2n_0$, $m\{\hat{R} > k - n_0\} \le \frac{C}{k^\alpha}\left(\frac{k}{k-n_0}\right)^\alpha \le \frac{C}{k^\alpha}$, so that

$$m\left\{\hat{R} > \frac{n}{i} - n_0\right\} \le C\frac{i^\alpha}{n^\alpha} \qquad \forall i \le \frac{1}{2}\left[\frac{n}{n_0}\right].$$

For each fixed i, we write

$$\begin{aligned}P\{T > n; \tau_{i-1} \le n < \tau_i\} &\le P\{T > \tau_{i-1}; n < \tau_i\} \\ &\le \sum_{j=1}^{i} P\left\{T > \tau_{i-1}; \tau_j - \tau_{j-1} > \frac{n}{i}\right\}\end{aligned}$$

and claim that each term in this sum is $\le C(1-\varepsilon_0)^i \frac{i^\alpha}{n^\alpha}$.

Consider first $i, j \ge 3$ (the order of conditioning is slightly different for the "small" terms):

$$P\left\{T > \tau_{i-1}; \tau_j - \tau_{j-1} > \frac{n}{i}\right\} = A \cdot B \cdot C$$

where

$$
\begin{aligned}
A &= P\{T > \tau_2\}P\{T > \tau_3 \mid T > \tau_2\} \cdots P\{T > \tau_{j-2} \mid T > \tau_{j-3}\}, \\
B &= P\left\{T > \tau_{j-1}; \tau_j - \tau_{j-1} > \frac{n}{i} \;\middle|\; T > \tau_{j-2}\right\}, \\
C &= P\left\{T > \tau_j \;\middle|\; T > \tau_{j-1}; \tau_j - \tau_{j-1} > \frac{n}{i}\right\} \cdots \\
&\qquad\qquad P\left\{T > \tau_{i-1} \;\middle|\; T > \tau_{i-2}; \tau_j - \tau_{j-1} > \frac{n}{i}\right\}.
\end{aligned}
$$

Note that A is void when $j \leq 3$, and C is void when $j = i$. Factors in A are each $\leq 1-\varepsilon_0$ by Lemma 1. Each factor in C is of the form $P\{T > \tau_k \mid T > \tau_{k-1}; \tau_j - \tau_{j-1} > \frac{n}{i}\}$ where $k \geq j$. Conditioning on ξ_k, we see that it is also $\leq 1 - \varepsilon_0$. The B-term is $\leq P\left\{\tau_j - \tau_{j-1} > \frac{n}{i} \mid T > \tau_{j-2}\right\}$. Since $\{T > \tau_{j-2}\}$ is ξ_{j-1}-measurable, we have, by Lemma 2, that it is $\leq Cm\left\{\hat{R} > \frac{n}{i} - n_0\right\} \leq C\frac{i^\alpha}{n^\alpha}$.

Observe that the "small" terms are not problematic. For $i < 3$, use the trivial estimate $P\{T > \tau_{i-1}; \tau_{i-1} \leq n < \tau_i\} \leq P\{\tau_i > n\} \leq C\frac{1}{n^\alpha}$. For $i \geq 3$ and, for example, $j = 2$, write

$$
\begin{aligned}
&P\left\{T > \tau_{i-1}; \tau_2 - \tau_1 > \frac{n}{i}\right\} \\
&\leq P\left\{\tau_2 - \tau_1 > \frac{n}{i}\right\} \; P\left\{T > \tau_2 \mid \tau_2 - \tau_1 > \frac{n}{i}\right\} P\{T > \tau_3 \mid \cdots\} \\
&\qquad\qquad \cdots \; P\{T > \tau_{i-1} \mid \cdots\}
\end{aligned}
$$

and argue as before.

Altogether we have shown that

$$
\text{(I)} \;\leq C \sum_{i=1}^{\infty} (1-\varepsilon_0)^i \frac{i^{\alpha+1}}{n^\alpha} \leq \frac{C}{n^\alpha} \;;
$$

hence

$$
P\{T > n\} \leq \frac{C}{n^\alpha} \qquad \text{for all } n.
$$

To complete the argument, we write

$$
\begin{aligned}
|F_*^n\lambda - F_*^n\lambda| &\leq C \sum_{i=0}^{\infty} (1-\varepsilon_0)^i \; P\{T_{i-1} \leq n < T_i\} && \text{by Lemma 4} \\
&\leq C \sum_{i=0}^{\infty} (1-\varepsilon_0)^i \sum_{j=1}^{i} P\left\{T_j - T_{j-1} > \frac{n}{i}\right\} && \text{as above} \\
&\leq C \sum_{i=0}^{\infty} (1-\varepsilon_0)^i \; i \; (m \times m)\left\{T > \frac{n}{i}\right\} && \text{by Sublemma 4.}
\end{aligned}
$$

Using our previous estimate on $P\{T > k\}$ with $P = m \times m$, the last line is $\leq \frac{C}{n^\alpha}$ as claimed.

4.2 Exponential decay: Proof of Theorem 2 II(b).

In this section we assume $m\{\hat{R} > n\} \leq C_1\theta^n$ for some $C_1 > 0$ and $\theta < 1$ and show that $\exists\tilde{\theta} < 1$ s.t. for all λ, λ' satisfying the condition in Theorem 2, $|F_*^n\lambda - F_*^n\lambda'| \leq C\tilde{\theta}^n$. As in the last subsection, C will be used as a generic constant which is allowed to depend only on F, λ and λ'. We emphasize that $\tilde{\theta}$ must be independent of P.

First we prove that $P\{T > n\} \leq C\theta_1^n$ for some $\theta_1 < 1$ independent of P. Let $\delta > 0$ be a small number to be specified later. Then

$$\begin{aligned} P\{T > n\} &= \sum_{i \leq [\delta n]} P\{T > n; \tau_{i-1} \leq n < \tau_i\} + \sum_{i > [\delta n]} P\{T > n; \tau_{i-1} \leq n < \tau_i\} \\ &\leq \sum_{i \leq [\delta n]} P\{\tau_{i-1} \leq n < \tau_i\} + \sum_{i > [\delta n]} P\{T > \tau_{i-1}\}. \end{aligned}$$

The second term is $\leq C(1-\varepsilon_0)^{[\delta n]}$. To estimate the first term, we fix i and write

$$\begin{aligned} & P\{\tau_{i-1} \leq n < \tau_i\} \\ \leq & \sum_{\substack{(k_1, \cdots, k_{i-1}): \\ k_j \geq n_0, \sum k_j \leq n}} P\left\{\tau_j - \tau_{j-1} = k_j,\ j = 1, \cdots, i-1;\ \tau_i - \tau_{i-1} > n - \sum k_j\right\}. \end{aligned}$$

Conditioning as usual, we obtain using Lemma 2 that each term in the sum above is

$$\leq \left(\prod_j K_0 C_1 \theta^{k_j - n_0}\right) \cdot K_0 C_1 \theta^{n - \sum k_j} \ \leq\ (K_0 C_1 \theta^{-n_0})^i \theta^n.$$

Note that K_0 depends on P but can be replaced by K_0^* independent of P if $j \geq$ some $i_0 = i_0(P)$. Thus

$$P\{\tau_{i-1} \leq n < \tau_i\} \leq C \binom{n+i-1}{i-1} \cdot (K_0^* C_1 \theta^{-n_0})^i \theta^n.$$

Now $\binom{n}{[\delta n]} \sim e^{\varepsilon n}$ for some $\varepsilon = \varepsilon(\delta)$ which $\to 0$ as $\delta \to 0$. Choosing $\delta > 0$ sufficiently small that $e^{\varepsilon(\delta)}(K_0^* C_1 \theta^{-n_0})^\delta\ \theta := \theta' < 1$ will ensure that the first term in the estimate of $P\{T > n\}$ above be $\leq [\delta n] \cdot C\theta'^n$ proving the desired estimate for $P\{T > n\}$.

Finally, an upper bound for $|F_*^n\lambda - F_*^n\lambda'|$ is, by Lemma 4,

$$C \sum_{i \leq [\delta_1 n]} P\{T_i \leq n < T_{i+1}\} + C \sum_{i > [\delta_1 n]} (1-\varepsilon_1)^i.$$

We deal with the first term exactly as we dealt with the first term of $P\{T > n\}$ earlier on, but let us check once more that δ_1 can be chosen independent of P:

Sublemma 4 tells us that there exists K_1^* independent of P such that for all $j \geq j_0 = j_0(P)$,

$$P\{T_j - T_{j-1} > k\} \leq K_1^*(m \times m)\{T > k\},$$

and the quantity on the right has been shown to be $\leq K_1^* C_{m\times m}\theta_1^n$ where $C_{m\times m}$ does not depend on P. ■

Remark. Our proof also shows that for all $\alpha \in (0,1)$,

$$m\{\hat{R} > n\} = \mathcal{O}(\theta^{n^\alpha}) \Rightarrow |F_*^n\lambda - F_*^n\lambda'| = \mathcal{O}(\tilde{\theta}^{n^{\alpha'}})$$

for every $\alpha' < \alpha$. This is because $\begin{pmatrix} n \\ [\delta n^\alpha] \end{pmatrix} \lesssim e^{\varepsilon(\delta) n^\alpha \log n}$, forcing us to split our sum into $\sum\limits_{i \leq [\delta n^{\alpha'}]} + \sum\limits_{i > [\delta n^{\alpha'}]}$. Note that the inequality $\theta^{a^\alpha + b^\alpha} \leq \theta^{(a+b)^\alpha}$ goes in the right direction.

5. Decay of Correlations and Central Limit Theorem

The purpose of this section is to prove Theorems 3 and 4. As we shall see, our decay of correlations results are formal consequences of Theorem 2. The Central Limit Theorem also follows quite readily from this and other known results.

5.1. Proof of Theorem 3.

Let $\mathcal{P}$ denote the Perron-Frobenius or transfer operator associated with F, *i.e.* if $\varphi = \frac{d\mu}{dm}$ where μ is a (signed) measure on Δ, then $\mathcal{P}(\varphi) = \frac{d(F_*\mu)}{dm}$.

Let $\varphi \in L^\infty(\Delta, m)$ and $\psi \in \mathcal{C}_\beta(\Delta)$ be as in the statement of Theorem 3, and let $\rho = \frac{d\nu}{dm}$ be the invariant density. We choose $a \geq 0$ and $b > 0$ s.t. $\tilde{\psi} := b(\psi + a)$ is bounded below by a strictly positive constant and $\int \tilde{\psi}\rho dm = 1$. Let λ be the probability measure on Δ with $\frac{d\lambda}{dm} = \tilde{\psi}\rho$. Then

$$\begin{aligned}
\left|\int (\varphi \circ F^n)\psi d\nu - \int \varphi d\nu \int \psi d\nu\right| &= \frac{1}{b}\left|\int (\varphi \circ F^n)\tilde{\psi} d\nu - \int \varphi d\nu \int \tilde{\psi} d\nu\right| \\
&= \frac{1}{b}\left|\int \varphi \mathcal{P}^n(\tilde{\psi}\rho) dm - \int \varphi\rho dm\right| \\
&\leq \frac{1}{b}\int |\varphi| \cdot |\mathcal{P}^n(\tilde{\psi}\rho) - \rho| \; dm \\
&\leq \frac{1}{b}\; |\varphi|_\infty \; |F_*^n\lambda - \nu| \; .
\end{aligned}$$

Since $\rho \in \mathcal{C}_\beta^+$ (Theorem 1), $\tilde{\psi}\rho \in \mathcal{C}_\beta^+$. Hence Theorem 2 applies. ■

5.2. Proof of Theorem 4.

First we recall a general result from [**L2**] which uses an idea in [**KV**]:

Theorem. [**L2**]. *Let $(X, \mathcal{F}, \mu)$ be a probability space, and let $T : X \circlearrowleft$ be a noninvertible ergodic measure-preserving transformation. Let $\varphi \in L^\infty(X, \mu)$ be such that $\int \varphi d\mu = 0$. Assume*

(i) $\sum_{n=1}^{\infty} \left|\int(\varphi \circ T^n)\varphi d\mu\right| < \infty$,

(ii) $\sum_{n=1}^{\infty} \hat{T}^{*n}(\varphi)$ *is absolutely convergent a.e.*

Then the CLT holds for φ, and the variance of the limiting normal distribution $= 0$ iff $\varphi \circ T = \psi \circ T - \psi$ for some measurable ψ.

In the statement above, $\hat{T}^*$ is the dual of the operator $\hat{T} : L^2(X, \mu) \to L^2(X, \mu)$ defined by $\hat{T}(\varphi) = \varphi \circ T$, that is to say, $\hat{T}^*(\varphi)(x) = E(\varphi | T^{-1}\mathcal{F})$ evaluated on $T^{-1}x$. We explain quickly the roles of (i) and (ii). The idea is to reduce the CLT for φ to one for ergodic reverse martingale differences. Observe that $\varphi \circ T^i$ is measurable wrt $T^{-i}\mathcal{F}$, a decreasing sequence of σ-algebras, and that $\{\varphi \circ T^i\}$ is a reverse martingale difference if $\hat{T}^*(\varphi) = 0$. That not being the case in general, one notes that the situation can be "corrected" by adding to $\varphi \circ T$ the function $g - g \circ T$ where g is given by the expression in (ii), assuming that makes sense. This correction, however, creates a new problem: the resulting random variables may not be in L^2 as it is a bit much to expect g to be in L^2 in general. An approximation trick from [**KV**] tells us that all is fine provided that the sum in (i), which is related to σ^2, is finite.

We return now to the setting of Theorem 4 and verify that the theorem cited above can be applied. Let $\varphi \in \mathcal{C}_\beta(\Delta)$ be such that $\int \psi d\nu = 0$. Condition (i) follows immediately from Theorem 3 and our hypothesis that $m\{\hat{R} > n\} = \mathcal{O}(n^{-\alpha})$ for some $\alpha > 1$. To check condition (ii), observe first that

$$\hat{F}^{*n}(\varphi)(x) = \sum_{y \in F^{-n}x} \frac{1}{\rho(x)} \frac{\rho(y)}{JF^n(y)} \cdot \varphi(y) = \frac{1}{\rho(x)} (\mathcal{P}^n(\varphi\rho))(x)$$

where $\rho = \frac{d\nu}{dm}$ and $\mathcal{P}$ is the Perron-Frobenius operator as before. Since $\rho \geq c_0 > 0$ (Theorem 1), it remains only to show that $\sum_{n=1}^{\infty} \mathcal{P}^n(\varphi\rho)$ is absolutely convergent $m-$a.e.

The same manipulations as in the last subsection allow us to write $\varphi\rho = c\left(\frac{d\lambda}{dm} - \frac{d\lambda'}{dm}\right)$ where $c > 0$ is a constant and λ, λ' are probability measures on Δ with $\frac{d\lambda}{dm}, \frac{d\lambda'}{dm} \in \mathcal{C}_\beta^+$. Recall now from 3.4 that there is a sequence of densities Φ_n on $\Delta \times \Delta$ representing the part of $P = \lambda \times \lambda'$ that has not yet been "matched" at time n, *i.e.*

$$F_*^n \lambda - F_*^n \lambda' = \pi_*(F \times F)_*^n(\Phi_n(m \times m)) - \pi'_*(F \times F)_*^n(\Phi_n(m \times m)).$$

Let ψ_n and ψ'_n denote respectively the densities wrt m of the two terms on the right. We then have

$$|\mathcal{P}^n(\varphi\rho)| = c\left|\frac{d}{dm}(F_*^n\lambda) - \frac{d}{dm}(F_*^n\lambda')\right| \leq c(\psi_n + \psi'_n).$$

Our hypothesis together with Lemma 4 and the estimates in 4.1 implies that $\int \psi_n dm = \int \Phi_n d(m\times m) = \mathcal{O}(n^{-\alpha}), \alpha > 1$. It suffices to show that on each Δ_ℓ, $\max\psi_n/\min\psi_n$ is uniformly bounded (independently of n); that would give $\psi_n|\Delta_\ell \leq C\frac{1}{m(\Delta_\ell)}\int\psi_n dm = \mathcal{O}(n^{-\alpha})$. Let $\tilde\eta := \{\Delta_{R_i-1,i}, i = 1,2,\cdots\} \cup \{\Delta_\ell - \cup_i\Delta_{R_i-1,i}, \ell = 1,2,\cdots\}$, and let $(\tilde\eta\times\tilde\eta)_n = \bigvee_{j=0}^{n-1}(F\times F)^{-j}(\tilde\eta\times\tilde\eta)$. The reason for using $\tilde\eta$ (instead of η) here is that for $\Gamma \in (\tilde\eta\times\tilde\eta)_n, (F\times F)^n\Gamma = \Delta_\ell\times\Delta_{\ell'}$ for some ℓ,ℓ'. It suffices therefore to fix ℓ and n, and show that for all $\Gamma\in(\tilde\eta\times\tilde\eta)_n$ with $F^n\pi\Gamma = \Delta_\ell$, the density of $\pi_*(F\times F)^n_*(\Phi_n(m\times m)|\Gamma)$ has the bounded ratio required. Let n_1 be the largest number less than n such that $n_1 = T_k|\Gamma$ for some k. Lemma 3' gives a distortion estimate for the density of $(F\times F)^{n_1}_*(\Phi_{n_1}(m\times m)|\Gamma)$. The measure whose density is of interest to us is simply the push-forward of this by $(F\times F)^{n-n_1}$ followed by π. This completes the verification of the second condition in the theorem cited. Theorem 4 follows.

PART II. APPLICATIONS TO 1-DIMENSIONAL MAPS

6. Expanding circle maps with neutral fixed points

The maps considered in this section are without a doubt the simplest "chaotic" dynamical systems that mix at polynomial speeds.

Notations: "$a_n \approx b_n$" (resp. "$a_n \lesssim b_n$") means there exists a constant $C \geq 1$ such that $C^{-1}b_n \leq a_n \leq Cb_n$ for all n (resp. $a_n \leq Cb_n$ for all n); analogous notations are used for functions; S^1 is identified with $[0,1]/\{0,1\}$, and additive notations are used.

6.1. Statements of results.

Let $f: S^1 \circlearrowleft$ be a degree d map, $d > 1$, with the following properties: There is a distinguished point in S^1, taken to be 0 for convenience, such that

(i) f is C^1 on S^1, and $f' > 1$ on $S^1 - \{0\}$;
(ii) f is C^2 on $S^1 - \{0\}$;
(iii) $f(0) = 0$, $f'(0) = 1$, and for all $x \neq 0$,

$$-xf''(x) \approx |x|^\gamma \qquad \text{for some} \quad \gamma > 0.$$

As $\gamma \downarrow 0$, the interval around 0 on which f' is near 1 shrinks to a point, so in a sense one could think of the limiting case as corresponding to the situation where $f' \geq \lambda$ for some $\lambda > 1$ and f'' is bounded. For convenience, let us agree to refer to this as the "$\gamma = 0$" case.

Let m denote Lebesgue measure on S^1, and let $\mathcal{H}$ denote the set of all Hölder continuous functions on S^1. We abbrebriate "ν absolutely continuous with respect to m" as "$\nu \ll m$". Our next theorem summarizes the mixing properties of f for the various values of γ. In order to present a complete picture, we have included in the statement of the theorem some results that are not new.

Theorem 5. (a) *For* $\gamma \geq 1$: $\frac{1}{n}\sum_{i=0}^{n-1}\delta_{f^i x}$ *converges weakly to the Dirac measure at* 0 *for m-a.e. x; in particular, f admits no finite invariant measure* $\nu \ll m$.

(b) *For* $\gamma < 1$: f *admits an invariant probability measure* $\nu \ll m$ *and* (f, ν) *is mixing.*

(c) *For* $0 < \gamma < 1$: *if* $\mathcal{P}$ *is the Perron-Frobenious operator associated with* f *and* $\rho = \frac{d\nu}{dm}$, *then for all* $\varphi \in \mathcal{H}$ *with* $\int \varphi dm = 1$,

$$\int |\mathcal{P}^n(\varphi) - \rho| dm \ \approx \ n^{1-\frac{1}{\gamma}};$$

and for all $\varphi \in L^\infty(S^1, m)$, $\psi \in \mathcal{H}$,

$$\left| \int (\varphi \circ f^n)\psi d\nu - \int \varphi d\nu \int \psi d\nu \right| = \mathcal{O}(n^{1-\frac{1}{\gamma}}).$$

(d) *For* $\gamma = 0$: *the covariance above is* $\leq C\theta^n$, $\theta < 1$ *depending only on the Hölder exponents of the test functions.*

(e) *For* $0 \leq \gamma < \frac{1}{2}$: *the Central Limit Theorem holds for all* $\varphi \in \mathcal{H}$.

Remark. (b) is a standard result one could find in elementary texts (*e.g.* [**M**]). (a) is also known; see for example [**Pg**] and [**HY**]. (d) is contained in [**HK**]; see also [**Y**]. Results similar to (c) but to my knowledge without the sharp bound have been announced independently during the past year by several authors in addition to myself, including [**H**], [**I2**] and [**LSV**]. (e) is essentially a corollary of (c) and (d) as explained in 5.2.

To illustrate the ideas of this paper we will give in the next few pages complete proofs of all of the assertions above.

6.2. Local analysis is a neighborhood of a neutral fixed point.

The analysis in this subsection is entirely local. For simplicity of notation we will restrict our attention to $f \mid [0, \varepsilon_0]$ where $(0, \varepsilon_0]$ is an interval on which condition (iii) at the beginning of 6.1 holds.

Let $x_0 \in (0, \varepsilon_0]$, and define x_n by $f x_n = x_{n-1}$ for $n = 1, 2, \ldots$. Since $f(x) - x \approx x^{\gamma+1}$, we observe that $\{x_n\}$ has the same asymptotics as $\left\{\frac{1}{n^\alpha}\right\}$ with $\alpha = \frac{1}{\gamma}$. More precisely, let $\Delta x_n := x_n - x_{n+1}$, $\Delta \frac{1}{k^\alpha} := \frac{1}{k^\alpha} - \frac{1}{(k+1)^\alpha}$. Then $x_n \in \left[\frac{1}{(k+1)^\alpha}, \frac{1}{k^\alpha}\right] \Rightarrow \Delta x_n \approx \Delta \frac{1}{k^\alpha}$; this is because $\Delta \frac{1}{k^\alpha} \approx \frac{1}{k^{\alpha+1}} = \left(\frac{1}{k^\alpha}\right)^{\gamma+1}$. In particular, there is a uniform bound on the number of intervals of the form $\left[\frac{1}{(k+1)^\alpha}, \frac{1}{k^\alpha}\right]$ that meet each $[x_{n+1}, x_n]$, and *vice versa*.

Lemma 5. (Distortion estimate). $\exists C_1$ *s.t.* $\forall i, n \in \mathbb{Z}^+$ *with* $i \leq n$ *and* $\forall x, y \in [x_{n+1}, x_n]$,

$$\left|\log \frac{(f^i)'x}{(f^i)'y}\right| \leq C_1 \frac{|f^i x - f^i y|}{\Delta x_{n-i}} \leq C_1.$$

Proof. First we prove a weaker bound than claimed:

$$\begin{aligned}
\left|\log \frac{(f^i)'x}{(f^i)'y}\right| &\leq \sum_{j=0}^{i-1} |\log f'(f^j x) - \log f'(f^j y)| \\
&= \sum_{j=0}^{i-1} \frac{|f''(\xi_j)|}{f'(\xi_j)} \cdot |f^j x - f^j y| \qquad \text{for some} \quad \xi_j \in [f^j x, f^j y] \\
&\lesssim \sum_{j=0}^{i-1} (x_{n-j+1})^{\gamma-1} \cdot (x_{n-j+1})^{\gamma+1} \\
&\lesssim \sum_k \left(\frac{1}{k^\alpha}\right)^{2\gamma} = \sum_k \frac{1}{k^2}.
\end{aligned}$$

Applying the above to all pairs of points in Δ_{n-j}, we obtain that for all $j < i$,

$$\frac{|f^j x - f^j y|}{\Delta x_{n-j}} \approx \frac{|f^i x - f^i y|}{\Delta x_{n-i}}.$$

Substituting this back into the estimate in the first part of the proof, we have

$$\begin{aligned}
\left|\log \frac{(f^i)'x}{(f^i)'y}\right| &\lesssim \sum_{j=0}^{i-1} (x_{n-j+1})^{\gamma-1} \cdot \Delta x_{n-j} \cdot \frac{|f^i x - f^i y|}{\Delta x_{n-i}} \\
&\leq \text{ const. } \frac{|f^i x - f^i y|}{\Delta x_{n-i}}.
\end{aligned}$$

$\blacksquare$

6.3. Invariant measures.

We will gear our exposition toward the $\gamma > 0$ case, pointing out possible simplifications for the $\gamma = 0$ case as we go along.

First we construct a basic partition $\mathcal{A}$ on S^1 with the property that the elements of $\mathcal{A}$ are intervals on which f' can be regarded as roughly constant. To do that we decompose S^1 into $I_1 \cup I_2 \cup \cdots \cup I_d$ where the I_j's are fundamental domains of f (*i.e.* $f(I_j) = S^1$) arranged in a natural order. Assume for definiteness that 0 is the common endpoint of I_1 and I_d. We further partition I_1 and I_d as follows. Let x_0 be the other end point of I_1, construct $x_n, n = 1, 2, \cdots$, as in 6.2, and let $J_n = [x_{n+1}, x_n]$. Likewise we let x'_0 be the end point of I_d other than 0 and decompose I_d into $\cup J'_n$. Let $\mathcal{A} = \{I_2, \cdots, I_{d-1};\ J_n, J'_n,\ n = 0, 1, 2, \cdots\}$.

For purposes of studying invariant measures, we construct a tower similar to that in 1.1 but with one difference, namely that $F^R(\Delta_{0,i})$ is not necessarily all of Δ_0. Let $\Delta_0 := S^1$, and let $\mathcal{A}$ correspond to the partition into $\{\Delta_{0,i}\}$. To define Δ it suffices to specify R. We let $R = 1$ on $I_2 \cup \cdots \cup I_{d-1} \cup J_0 \cup J'_0$, and let $R|J_n = R|J'_n = n+1$ for $n \geq 1$. F is defined as in 1.1, with $F|\Delta_{R_i-1,i}$ determined by $f^R|\Delta_{0,i}$. Note that for $j = 2, \cdots d-1$, we have $f^R(I_j) = S^1$, whereas the f^R-images of all other elements of $\mathcal{A}$ are either $I_2 \cup \cdots \cup I_d$ or $I_1 \cup \cdots \cup I_{d-1}$. Our reference measure on Δ_0 is m; this together with $JF = 1$ on $\Delta - \cup_i \Delta_{R_i-1,i}$ forces a reference measure on the rest of Δ which we will continue to call m. Observe that there exists $\beta < 1$ such that $(f^R)'x \geq \beta^{-1}$ for all $x \in S^1$, so that $|x-y| \leq \beta^n$ whenever $s(x,y) \geq n$. The regularity condition for JF^R now follows from Lemma 5 and the usual distortion property for C^2 expanding maps. Note that $m\{R > n\} = m(\cup_{i \geq n} J_n) + m(\cup_{i \geq n} J'_n)$, which for $\gamma > 0$ is $\approx n^{-\alpha}$ with $\alpha = \gamma^{-1}$.

For $\gamma = 0$, we could do as above and obtain $m\{R > n\} \leq C\theta_0^n$ for some $\theta_0 < 1$, but it is simpler to take $\{\Delta_{0,i}\} := \{I_1, \cdots, I_d\}$ and $R \equiv 1$. Observe that this would not have worked for $\gamma > 0$ for distortion reasons.

Let $\pi : \Delta \to S^1$ be the natural projection satisfying $\pi \circ F = f \circ \pi$.

Existence of finite invaraint measures: A proof identical to that for Theorem 1 shows that F^R admits an invariant probability measure $\bar{\nu}_0 \ll m$ with $c_0 \leq \frac{d\bar{\nu}_0}{dm} \leq c_1$ for some $c_0,\ c_1 > 0$. That $\frac{d\bar{\nu}_0}{dm}$ is bounded follows immediately from its bounded distortion on each I_j; that it is bounded away from 0 follows from the transitive action of F^R on the I_j's. Out of $\bar{\nu}_0$ we construct an F-invariant measure $\bar{\nu}$ which is finite if and only if $\int R dm < \infty$, and the integrability of R corresponds exactly to $\gamma < 1$. Take $\nu = \pi_* \bar{\nu}$. ■

Let $\rho = \frac{d\nu}{dm}$. Note that in the case $\gamma > 0$, we have in fact shown that $\rho|J_k \approx k$. This is because $\nu(J_k) = \bar{\nu}(\pi^{-1} J_k) = \bar{\nu}(\cup_{i \geq k} J_i) \approx k^{-\alpha}$, and it follows using the distortion estimate for ρ that $\rho|J_k \approx \frac{1}{m(J_k)} k^{-\alpha} \approx k$. It is easy to see that ρ is bounded in the $\gamma = 0$ case.

Asymptotic distribution of m-typical points for $\gamma \geq 1$: To prove $\frac{1}{n} \sum_{i=0}^{n-1} \delta_{f^i x} \to \delta_0$, we fix an arbitrarily small neighborhood (x'_N, x_N) of 0, an arbitrary $\epsilon > 0$, and show that for m-a.e. x,

$$\frac{1}{n}\#\{0 \le k < n : f^k x \in (x'_N, x_N)\} > 1 - \epsilon$$

as $n \to \infty$. Choose $N_1 > N$ s.t. $\nu(S^1 - (x'_N, x_N))/\nu(S^1 - (x'_{N_1}, x_{N_1})) < \epsilon$. Let $f^{(N_1)}$ denote the first return map from $S^1 - (x'_{N_1}, x_{N_1})$ to itself. Then $\nu|(S^1 - (x'_{N_1}, x_{N_1}))$ is a finite $f^{(N_1)}$-invariant measure, which is easily seen to be ergodic (its induced map on I_2, for example, is clearly ergodic). Thus for m-a.e. point in $S^1 - (x'_{N_1}, x_{N_1})$, the fraction of time spent in (x'_N, x_N) under $f^{(N_1)}$ is $> 1 - \epsilon$, and that is clearly larger than the corresponding fraction under f. ■

Lower bound for $\int |\mathcal{P}^n(\varphi) - \rho| dm$ *for* $0 < \gamma < 1$: This argument applies to all $\varphi \in L^\infty(S^1, m)$. We may assume $\varphi \ge 0$. Let $\bar\lambda$ be the measure on Δ whose density is equal to φ on Δ_0 and 0 elsewhere. Then $\mathcal{P}^n(\varphi)$ is the density of $\pi_*(F_*^n \bar\lambda)$, and $\frac{d(F_*^n\bar\lambda)}{dm} \le |\varphi|_\infty \frac{d(F_*^n m)}{dm}$ which is uniformly bounded for all n. This together with $(F_*^n\bar\lambda)(\cup_{\ell > n}\Delta_\ell) = 0$ imply that $\mathcal{P}^n(\varphi)|J_k \le C|\varphi|_\infty m(\cup_{j=k}^{k+n} J_j)$. Since $(k+n)^{-\alpha}/k^{-\alpha} \to 1$ uniformly as $k/n \to \infty$, there exists N such that for all $k \ge Nn$, $\mathcal{P}^n(\varphi)|J_k \le \frac{1}{2}\rho|J_k \approx k$. Thus $\int |\mathcal{P}^n(\varphi) - \rho| dm \gtrsim \sum_{k \ge Nn} k m(J_k) \approx n^{-\alpha+1}$. ■

6.4. Decay of correlations.

To study mixing properties it is convenient to work with a setup like that in 1.1. For this purpose we introduce a new stopping time $R^*(x)$ on S^1 defined to be the smallest $n \ge R(x)$ s.t. $f^n x \in I_1$. The new tower, which we denote by $F^* : \Delta^* \circlearrowleft$, is built over I_1 with return time function R^*.

To estimate $m\{R^* > n\}$, we introduce on S^1 an auxiliary sequence of stopping times R_i defined by $R_1 = R$ and $R_i = R_{i-1} + R \circ f^{R_{i-1}}$, so that $R^*(x) = R_i(x)$ where i is the smallest integer ≥ 1 such that $f^{R_i}x \in I_1$. Let $\mathcal{B}_i$ be the σ-algebra on S^1 consisting of intervals $\omega \in \bigvee_{i=0}^{n-1} F^{-i}\mathcal{A}$ (where $\mathcal{A}$ is as in 6.3) with the property that $R_i = n$ on ω. Since f^{R_i} maps each $\omega \in \mathcal{B}_i$ onto a union of I_j's, we have $m\{R_{i+1} - R_i \mid \omega\} < Cm\{R > n\}$. We also claim that for $i > 1$, if $\omega \in \mathcal{B}_i$ is such that $R^* \ne R_j$ on ω for $j = 1, 2, \cdots, i-1$, then $m\{R^* = R_i \mid \omega\} \ge \varepsilon_0$ for some $\varepsilon_0 > 0$. The only worrisome possibility here is for $f^{R_i - 1}\omega$ to be contained in I_1, but this is impossible since R^* would have been equal to the smallest $n \ge R_{i-1}$ when ω enters I_1. The present situation, therefore, is entirely analogous to that in 3.3, with $f : S^1 \circlearrowleft$ instead of $F \times F : \Delta \times \Delta \circlearrowleft$, R_i in the place of τ_i and R^* in the place of T. Mimicking the proofs in 4.1, we conclude that $m\{R^* > n\} = \mathcal{O}(n^{-\alpha})$ for $\gamma > 0$. The $\gamma = 0$ can be dealt with similarly, but with $R \equiv 1$, it is quite easy to see without any of this that $m\{R^* > n\} = \mathcal{O}(\theta_1^n)$ for some $\theta_1 < 1$.

Returning to the tower $F^* : \Delta^* \circlearrowleft$, one sees that f^{R^*} induces a natural partition $\{\Delta^*_{0,i}\}$ on I_1 with the property that f^{R^*} maps each $\{\Delta^*_{0,i}\}$ bijectively onto I_1. The regularity condition for this tower is easily verified as before.

Exactness of (f,ν): For $\gamma<1$, an F^*-invariant probability measure $\bar\nu^*$ exists on Δ^* with $\pi_*\bar\nu^*=\nu$. Since for each j there is an interval $\omega\subset I_1$ with the property that $f^i\omega\subset I_2$ for $i=1,2,\cdots j-1$ and $f^j\omega=I_1$, we have gcd $\{R^*\}=1$. It follows from Theorem 1 that $(F^*,\bar\nu^*)$ is exact. Quotients of exact measure-preserving transformations are exact. ■

Correlation decay and CLT: For $\varphi\in\mathcal{H}$, let φ^* be the function on Δ^* defined by $\varphi^*=\varphi\circ\pi$. Then $\varphi^*\in\mathcal{C}_\beta(\Delta^*)$ where $\beta=(\min(f^{R^*})')^{-\sigma}$ and σ is the Hölder exponent of φ. The assertions on covariance decay in (c) and (d) follow immediately from the discussion above, Theorem 3, and the fact that

$$\int(\varphi\circ f^n)\psi d\nu-\int\varphi d\nu\int\psi d\nu=\int(\varphi^*\circ F^{*n})\psi^* d\bar\nu^*-\int\varphi^* d\bar\nu^*\int\psi^* d\bar\nu^*.$$

The CLT statement follows from Theorem 4 and a similar observation. ■

Upper bound for $\int|\mathcal{P}^n(\varphi)-\rho|dm$: An upper bound is $|F_*^{*n}\bar\lambda^*-\bar\nu^*|$ where $\bar\lambda^*$ is any measure on Δ^* with $\frac{d(\pi_*\bar\lambda^*)}{dm}=\varphi$. (Note that φ^* in the last paragraph is not a candidate for the density of λ^*.) To have the desired estimate on $|F_*^{*n}\bar\lambda^*-\bar\nu^*|$, we must select $\bar\lambda^*$ in such a way that $\frac{d\bar\lambda^*}{dm}\in\mathcal{C}_\beta(\Delta^*)$. One possibility is to identify I_1 with Δ_0^*, J_0 with $\Delta_{0,0}^*$, $I_2\cup\cdots\cup I_d$ with $\Delta_{1,0}^*$, and to "lift" φ accordingly. ■

7. Piecewise expanding maps: the non-Markov case

The purpose of this section is to illustrate how the ideas developed earlier on can be taken one step further to handle 1-dimensional maps that do not have *a priori* Markov structures. The notations "$\approx$" and "$\lesssim$" are as defined in Section 6.

7.1. Setting and results.

Assumptions. Consider $f:[0,1]\circlearrowleft$ with the following properties: $[0,1]=I_1\cup\cdots\cup I_d$ where the I_j's are closed intervals meeting only in their end points. Let $[a,b]$ be one of the I_j's. We assume that

(i) on each $I_j\neq[a,b]$, $|f'|\geq\mu$ for some $\mu>2$ and $|f''|$ is uniformly bounded;

(ii) $f(a)=a$, $f'(a)=1$; $f'(x)\geq\mu$ for $x\in[a,b]$ s.t. $f^ix\notin[a,b]$, $i=1,2$ or 3; and $\exists\gamma$, $0<\gamma<1$, s.t. $\forall x\in(a,b)$, $f''(x-a)\approx(x-a)^{\gamma-1}$.

Theorem 6. *f admits an invariant probability measure $\nu\ll m$. If (f,μ) is mixing, then for all $\varphi\in L^\infty(S^1,m)$ and $\psi\in\mathcal{H}$,*

$$\left|\int(\varphi\circ f^n)\psi d\nu-\int\varphi d\nu\int\psi d\nu\right|=\mathcal{O}(n^{1-\frac{1}{\gamma}}).$$

The Central Limit Theorem holds for all $\varphi\in\mathcal{H}$ if $\gamma<\frac12$.

Remarks. (a) For simplicity we have limited ourselves to one neutral fixed point (and only on one side). The theorem generalizes easily to multiple neutral fixed points and neutral periodic orbits.

(b) We will in fact prove that f admits at most finitely many ergodic probability measures $\nu \ll m$, and that each one is either mixing or is a cyclic permutation of mixing components for some power of f. Our conclusion applies to each of the mixing components.

(c) We require $|f'| \geq \mu$ for some $\mu > 2$ to guarantee that f expands faster than its growth in local complexity. (For uniformly expanding maps, this condition can always be arranged by considering a power of f; it is not automatic for maps with nonuniform expansion.)

As is typically the case, there are two main steps in the implementation of the scheme outlined at the beginning of the introduction. The first estimates the speed with which arbitrarily small sets grow to a fixed size. (If the reference set has a complicated structure, then one needs to consider the statistics of gap sizes etc. but that is irrelevant here.) The outcome of this step depends sensitively on the dynamics in question. The second step relates the growth rates in the first step to the speed of correlation decay. This step tends to be quite generic and not particularly model dependent. These two steps are carried out in 7.2 and 7.3.

7.2. A growth lemma.

Let $\Omega \subset [0,1]$ be an interval and $\delta > 0$ a given number. We are interested in stopping times $S : \Omega \to \mathbb{Z}^+$ with the following properties:

(a) Ω is partitioned into (infinitely many) intervals $\{\omega\}$ on each one of which S is constant;

(b) $f^S(\omega)$ is an interval of length $> 5\delta$;

(c) $|(f^S|\omega)'| \geq \mu$;

(d) $\exists C$ s.t. for all ω and $\forall x, y \in \omega$, $|\log \frac{(f^S)'x}{(f^S)'y}| \leq C|f^S x - f^S y|$.

Let $\alpha = \gamma^{-1}$ be as before.

Lemma 6. *For all sufficiently small $\delta > 0$ there exists a constant $C = C(\delta)$ such that for every interval $\Omega \subset [0,1]$, there is a stopping time S as above with*

$$m\{S > n\} \leq Cn^{-\alpha} \quad \text{for every } n.$$

Proof. First some notations: Let $[a,b] = \cup J_n$ be the partition with $x_0 = b$, $fx_{n+1} = x_n$, and $J_n = [x_{n+1}, x_n]$; and let $\tilde{J}_n = J_{n-1} \cup J_n \cup J_{n+1}$. Two useful partitions are $\mathcal{Q}_0 = \{I_1, \cdots, I_d\}$ and $\mathcal{Q} = \{[0,a], [b,1]; J_n, n = 0, 1, 2, \cdots\}$. If $\mathcal{A}$ and $\mathcal{B}$ are partitions, let $\mathcal{A} \vee \mathcal{B} := \{A \cap B : A \in \mathcal{A}, B \in \mathcal{B}\}$.

We require δ to be small enough that (1) if $\omega \subset I_j$ is any interval with $|\omega| \leq 5\delta$, then $f\omega$ cannot meet more than two I_k's; and (2) $|J_0| > 5\delta$.

We now define S on a given interval Ω which we may assume has length $< 5\delta$. (If not, first subdivide). Let $\mathcal{P}_0 = \mathcal{Q}_0|\Omega$, and consider one $\omega \in \mathcal{P}_0$ at a time. Let $\tilde{\mathcal{P}}_1|\omega$ be essentially $(f^{-1}\mathcal{Q})|\omega$ but modified in the following way: if the leftmost element of $\mathcal{Q}|(f\omega)$ lies in some J_k, adjoin it to its neighbor to the right (if it has a neighbor on the right side) before pulling back by f; simlarly, adjoin the rightmost element of $\mathcal{Q}|(f\omega)$ to its neighbor if it falls on some J_k. Thus the elements $\omega' \in \tilde{\mathcal{P}}_1$ are of three types:

Type 1. $f\omega' \subset [a,b]$ and $J_k \subset f\omega' \subset \tilde{J}_k$ for some k.

Type 2. $\omega' = \omega$ and $f\omega$ is contained in $J_k \cup J_{k+1}$ for some k. We shall refer to ω as a "short component".

Type 3. $f\omega' \not\subset [a,b]$. Note that there is at most one ω' of this type because $f\omega$ cannot meet both $[0,a]$ and $[b,1]$.

For each $\omega' \in \tilde{\mathcal{P}}_1$, we do one of the following: we either declare an S-value on ω' and take it out of consideration forever, or we postpone deciding and put it in a set called Ω_1 which is being created in this procedure. For ω' of Type 1, we let $S|\omega' = k+1$. (Let us verify that this is a legitimate definition: first, $f^{k+1}\omega'$ has only one component and it contains J_0, so $|f^{k+1}\omega'| > 5\delta$; second, since $f^k\omega' \subset \tilde{J}_1$, $f'|(f^k\omega') \geq \mu$; the distortion requirement is also evident.) For ω' of Type 2, let $i_1(\omega')$ be the smallest $i \geq 1$ s.t. $f^i\omega' \not\subset [a,b]$. If $|f^{i_1}\omega'| > 5\delta$, then we declare that $S|\omega' = i_1$. If not, we put it in Ω_1. For Type 3, we let $i_1(\omega') = 1$ and do as in the last case.

It is important to observe that for each $\omega \in \mathcal{P}_0$, we have put at most one $\omega' \in \tilde{\mathcal{P}}_1|\omega$ in Ω_1 (either $\omega' = \omega$, which corresponds to the case where ω is a short component, or ω' is of Type 3) and that the f^{i_1}-image of this ω' is $\leq 5\delta$ in length and it meets at most two of the I_j's. Let $\mathcal{P}_1 = \{(f^{-i_1}\mathcal{Q}_0)|\omega' : \omega' \in \tilde{\mathcal{P}}_1|\Omega_1\}$. Denoting the cardinality of a partition by $card(\cdot)$, we have:

(a) $card(\mathcal{P}_1) \leq 2\, card(\mathcal{P}_0)$;

(b) for all $\omega'' \in \mathcal{P}_1$, $f^{i_1+1}\omega''$ has only one component, and $|(f^{i_1+1})'|\omega''| \geq \mu$.

Next we repeat the procedure above with $\mathcal{P}_1$ in the place of $\mathcal{P}_0$. That is, for each $\omega \in \mathcal{P}_1$, we consider $f^{i_1+1}\omega$, define $\tilde{\mathcal{P}}_2|\omega = (f^{-(i_1+1)}\mathcal{Q})|\omega$ with end segments suitably modified, set $S|\omega' = i_1(\omega') + 1 + k$ if ω' is of Type 1 and $f^{i_1+1}\omega' \supset J_k$, and for Types 2 and 3 define $i_2(\omega')$ to be the smallest $i \geq i_1+1$ s.t. $f^{i_2}(\omega') \not\subset [a,b]$ etc. We create in this process $\Omega_2 \subset \Omega_1$ and $\mathcal{P}_2$ on Ω_2. Step 3 is then carried out for elements of $\mathcal{P}_2$, and so on. One obtains inductively that

(a) $card(\mathcal{P}_k) \leq 2^k\, card(\mathcal{P}_0)$;

(b) for all $\omega'' \in \mathcal{P}_k$, $f^{i_k+1}\omega''$ has only one component, and $|(f^{i_k+1})'|\omega''| \geq \mu^k$.

We now estimate $m\{S > n\}$ where $m\{S > n\}$ is to be interpreted as the set of points determined to have S-value $> n$ together with those not yet assigned an S-value by step n. We write $\{S > n\} \subset B_1 \cup B_2 \cup B_3$ where the B_i's are defined and estimated as follows:

Let $B_1 = \Omega_k$ for some $k \approx \log n$. Since Ω_k contains at most $2^k \cdot card(\mathcal{P}_0)$ intervals of length $< \mu^{-k}$ each, we have $m(B_1) \leq (\frac{2}{\mu})^k card(\mathcal{P}_0) \lesssim n^{-\alpha}$.

Let $B_2 = \{\omega' \in \tilde{\mathcal{P}}_j,\ j < k : \omega'$ is a short component and $f^{i_j+1}\omega' \subset \tilde{J}_p$ for some $p > n^{\frac{\alpha}{\alpha+1}}\}$. Since $p > n^{\frac{\alpha}{\alpha+1}} \Rightarrow |\tilde{J}_p| \lesssim n^{-\alpha}$, we have $m(B_2) \lesssim card(\mathcal{P}_0) \cdot \sum (\frac{2}{\mu})^j \cdot n^{-\alpha}$ which is harmless.

Removing B_1 allows us to consider only those $\omega' \in \tilde{\mathcal{P}}_j,\ j < k \approx \log n$, for which an S-value $> n$ is declared at step j. After removing B_2, we may assume that on such an ω', $i_\ell - i_{\ell-1} \le n^{\frac{\alpha}{\alpha+1}}$ for all $\ell < j$. It suffices therefore to put into B_3 those $\omega' \in \tilde{\mathcal{P}}_j$ with $f^{i_j+1}\omega' \subset \tilde{J}_p$ for $p \ge n - jn^{\frac{\alpha}{\alpha+1}}$. We then have

$$m(B_3) \ \le \ card(\mathcal{P}_0) \sum_{j=0}^{C\log n} \left(\frac{2}{\mu}\right)^j \frac{1}{(n - jn^{\frac{\alpha}{\alpha+1}})^\alpha} \ \lesssim \ \frac{1}{n^\alpha}\ .$$

This completes the proof of Lemma 6. ■

7.3. Invariant measures and decay rates.

We now explain how to derive the desired information from Lemma 6. Let $\{\Lambda_1, \cdots, \Lambda_r\}$ be a partition of $[0,1]$ into intervals of length δ. Our first step is to introduce a suitable return time function R on $[0,1]$ with the properties that (1) the dynamics of $f^R : [0,1] \circlearrowleft$ is Markov-like with respect to the "states" $\{\Lambda_i\}$ (see below) and (2) $m\{R > n\}$ reflects the tail behavior of the stopping times in Lemma 6. In other words, we are going to build a tower over $[0,1]$ with return time function R, but I will omit this language from here on.

We define R on one Λ_j at a time. Let S_0 be a stopping time on Λ_j of the type given by the lemma, and let $\mathcal{A}_0 = \{\omega_0\}$ be its associated partition. For each $\omega_0 \in \mathcal{A}_0$, $f^{S_0}(\omega_0)$ contains at least three Λ_i's (and may intersect two others, one at each end). Let $\Lambda_p, \Lambda_{p+1}, \cdots, \Lambda_{p+q}$ be all the Λ_i's contained in $f^{S_0}(\omega_0)$. We define $R = S_0$ on $(f^{S_0})^{-1}(\Lambda_{p+1} \cup \cdots \cup \Lambda_{p+q-1})$, so that $f^{S_0}(\omega_0 - \{R = S_0\})$ consists of two intervals ω_0^+ and ω_0^- with $\delta \le |f^{S_0}\omega_0^\pm| \le 2\delta$. After doing this for every $\omega_0 \in \mathcal{A}_0$, we have created a partition $\{\omega_0^\pm\}$ of $\Lambda_j - \{R = S_0\}$. For each $\omega_0^\pm$ we consider a stopping time S on $f^{S_0}\omega_0^\pm$ with the properties in Lemma 6 and define $S_1 = S_0 + S \circ f^{S_0}$ on $\omega_0^\pm$. Then S_1 induces on $\Lambda_j - \{R = S_0\}$ a partition $\mathcal{A}_1 = \{\omega_1\}$, and $f^{S_1}\omega_1$ is again an interval containing at least three Λ_i's. As before, we declare that $R = S_1$ on the $(f^{S_1})^{-1}$-image of all but two of these Λ_i's leaving at each end of $f^{S_1}\omega_1$ an interval of length between δ and 2δ. On $\Lambda_j - (\{R = S_0\} \cup \{R = S_1\})$, we define S_2 and so on.

Now on each $\omega_i^\pm$, S_i is constant. Using Lemma 6 and the usual distortion estimates, we have $m\{S_{i+1} - S_i > n \mid \omega_i^\pm\} \le Cn^{-\alpha}$. Moreover, $R > S_i$ on $\omega_i^\pm$, and $m\{R = S_{i+1} \mid \omega_i^\pm\} \ge$ some $\varepsilon_0 = \varepsilon_0(\delta) > 0$. As before we conclude that $m\{R > n\} \le Cn^{-\alpha}$.

Recapitulating, we have partitioned each Λ_j into a countable number of intervals $\{\omega\}$ with the property that $f^R|\omega$ has bounded distortion and the f^R-image of each ω is one of the Λ_k's. This is the finite Markov structure we have alluded to earlier on. Our next step is to use it to obtain information on the invariant measures of f.

Pushing forward $m|\Lambda_j$ by $(f^R)^n$, $n = 1, 2, \cdots$, we see that f^R admits a finite number of ergodic probability measures $\{\nu_i^R\}$ each with a strictly positive density on a union of Λ_k's. Since $\int R dm < \infty$, each ν_i^R gives rise to an f-invariant ergodic measure ν_i. (It is possible, however, to have $\nu_i = \nu_{i'}$ for $i \ne i'$.) We claim that these are the only f-invariant absolutely continuous ergodic measures, for m-a.e. point in $[0,1]$ is eventually mapped into the support of some ν_i^R under f^R.

To study the mixing properties of ν_i, let Λ_j be a state in the support of ν_i^R. Let $R_1 = R$, $R_n = R_{n-1} + R \circ f^{R_{n-1}}$, and let $R^*(x)$ be the smallest R_k s.t. $f^{R_k}(x) \in \Lambda_j$. From Section 2 we see that the tower over Λ_j with return time R^* decomposes into N^* mixing components where $N^* = \gcd \{R^*\}$. These project to the mixing components of ν_i although some may merge.

To prove the assertion on decay rates, it remains only to verify that $m\{R^* > n\} \le Cn^{-\alpha}$. Here we have $m\{R_{k+1} - R_k \mid R_k\} < Cn^{-\alpha}$, and $m\{R^* = R_{k+n},\ 1 \le n < r \mid R^* > R_k\} \ge \varepsilon > 0$ where r is the total number of Λ_ℓ's. This is a slight variation from our usual theme. We leave it to the reader to check that the desired estimate continues to hold.

■

Acknowledgement. The author thanks T. Liggett for information on related questions in probability and S. Isola for interesting discussions on 1-dimensional maps with neutral fixed points.

References

[BY] M. Benedicks and L.-S. Young, *Decay or correlations for certain Henon maps*, 1996 preprint.

[FL] A. Fisher and A. Lopes, *Polynomial decay of correlation and the central limit theorem for the equilibrium state of a non-Hölder potential*, 1997 preprint.

[HK] F. Hofbauer and G. Keller, *Ergodic properties of invariant measures for piecewise monotonic transformations*, Math. Z., **180** (1982), 119-140.

[H] H. Hu, private communication.

[HY] H. Hu and L.-S. Young, *Nonexistence of* SBR *measures for some systems that are "almost Anosov"*, Erg. Th. & Dyn. Sys., **15** (1995), 67-76.

[I1] S. Isola, *On the rate of convergence to equilibrium for countable ergodic Markov chains*, 1997 preprint.

[I2] S. Isola, *Dynamical zeta functions and correlation functions for intermittent interval maps*, preprint.

[KV] C. Kipnis and S.R.S. Varadhan, *Central limit theorem for additive functions of reversible Markov process and applications to simple exclusions*, Commun. Math. Phys. **104** (1986), 1-19.

[L1] C. Liverani, *Decay of correlations*, Annals Math. **142** (1995), 239-301.

[L2] C. Liverani, *Central limit theorem for deterministic systems*, International conference on dynamical systems, Montevideo 1995, Eds. F.Ledrappier, J.Lewowicz, S.Newhouse, Pitman research notes in Math, **362** (1996), 56-75.

[LSV] C. Liverani, Saussol and S. Vaienti, 1997 preprint.

[M] R. Mañé, *Ergodic theory and differentiable dynamics*, springer Verlag, 1983.

[Pi] G. Pianigiani, *First return maps and invariant measures*, Israel J. Math, **35** (1980), 32-48.

[Po] M. Pollicott, *Rates of mixing for potentials of summable variation*, 1997 preprint.

[Pt] J.W. Pitman, *Uniform rates of convergence for Markov chain transition probabilities*, Z. Wahr. verw. Geb. **29** (1974), 193-227.

[R] D. Ruelle, *Thermodynamic formalism*, Addison-Wesley, New York, 1978.

[TT] P. Tuominen and R. Tweedie, *Subgeometric rates of convergence of f-ergodic Markov chains*, Adv. Appl. Prob. **26** (1994), 775-798.

[Y] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, to appear in Annals of Math.