



# Markov structures and decay of correlations for non-uniformly expanding dynamical systems <sup>☆</sup>

José F. Alves <sup>a,\*</sup>, Stefano Luzzatto <sup>b</sup>, Vilton Pinheiro <sup>c</sup>

<sup>a</sup> *Departamento de Matemática Pura, Faculdade de Ciências do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal*

<sup>b</sup> *Mathematics Department, Imperial College, 180 Queen's Gate, London SW7, UK*

<sup>c</sup> *Departamento de Matemática, Universidade Federal da Bahia, Av. Ademar de Barros s/n, 40170-110 Salvador, Brazil*

Received 15 May 2002; received in revised form 18 February 2004; accepted 15 December 2004

Available online 20 April 2005

---

## Abstract

We consider non-uniformly expanding maps on compact Riemannian manifolds of arbitrary dimension, possibly having discontinuities and/or critical sets, and show that under some general conditions they admit an induced Markov tower structure. Moreover, the decay of the return time function can be controlled in terms of the time generic points need to achieve some uniform expanding behavior. As a consequence we obtain some rates for the decay of correlations of those maps and conditions for the validity of the Central Limit Theorem.

© 2005 L'Association Publications de l'Institut Henri Poincaré. Published by Elsevier B.V. All rights reserved

MSC: 37D25; 37A25; 37D50

---

## 1. Introduction and statement of results

The purpose of this paper is to study the geometrical structure and statistical properties of piecewise smooth dynamical systems which satisfy some asymptotic expansion properties almost everywhere. We begin with a discussion of the statistical properties we are interested in, and the precise statement of our assumptions and results concerning these properties. We then state our main result on the existence of an induced Markov map and present our main application to class of two-dimensional non-uniformly expanding *Viana maps*.

---

<sup>☆</sup> Work carried out at the Federal University of Bahia, University of Porto and Imperial College, London. Partially supported by CMUP, PROLYN, SAPIENS and UFBA.

\* Corresponding author.

*E-mail addresses:* [jfalves@fc.up.pt](mailto:jfalves@fc.up.pt) (J.F. Alves), [stefano.luzzatto@ic.ac.uk](mailto:stefano.luzzatto@ic.ac.uk) (S. Luzzatto), [viltonj@ufba.br](mailto:viltonj@ufba.br) (V. Pinheiro).

*URLs:* <http://www.fc.up.pt/cmup/home/jfalves>, <http://www.ma.ic.ac.uk/~luzzatto>.

1.1. Statistical properties

One of the most powerful ways of describing the dynamical features of systems, specially those having a very complicated geometrical and topological structure of individual orbits, is through invariant probability measures. Any such measure can be decomposed into ergodic components and, by a simple application of Birkhoff’s Ergodic Theorem, almost every initial condition in each ergodic component has the same statistical distribution in space. On such a component, a map  $f$  is said to be *mixing* if

$$|\mu(f^{-n}(A) \cap B) - \mu(A)\mu(B)| \rightarrow 0, \quad \text{when } n \rightarrow \infty,$$

for any measurable sets  $A, B$ . Standard counterexamples show that in general there is no specific rate at which this loss of memory occurs: it is always possible to choose sets  $A$  and  $B$  for which mixing is arbitrarily slow. It is sometimes possible however, to define the *correlation function*

$$C_n(\varphi, \psi) = \left| \int (\varphi \circ f^n)\psi \, d\mu - \int \varphi \, d\mu \int \psi \, d\mu \right|,$$

and to obtain specific rates of decay which depend only on the map  $f$  (up to a multiplicative constant which is allowed to depend on  $\varphi, \psi$ ) as long as the *observables*  $\varphi, \psi$  belong to some appropriate functional space. Notice that choosing these observables to be characteristic functions this gives exactly the original definition of mixing.

The precise dynamical features which cause mixing, and in particular the dynamical features which cause different rates of decay of the correlation function, are still far from understood. Exponential mixing for uniformly expanding and uniformly hyperbolic systems has been known since the work of Sinai, Ruelle and Bowen [27,11,12,10] and may not seem surprising in view of the fact that all quantities involved are exponential. However the subtlety of the question is becoming more apparent in the light of recent examples which satisfy asymptotic exponential expansion estimates but only subexponential decay of correlations. The simplest case is that of one-dimensional maps which are expanding everywhere except at some fixed point  $p$  for which  $f'(p) = 1$ . In certain cases (essentially depending on the second derivative  $f''(p)$ ) there is an absolutely continuous mixing invariant measure with positive Lyapunov exponent but strictly subexponential [22,20,30,24] (and in some cases even subpolynomial [19]) decay of correlations. In this case the indifferent fixed point is *slowing down* the mixing process since nearby points are moving away (and thus “mixing”) at a slower, subexponential, rate rather than the exponential rate at which they move away from other fixed or periodic point.

A more subtle slowing down effect occurs in smooth one-dimensional maps with critical points where the rate of mixing is essentially determined by the rate of growth of the derivative along the critical orbit [9]. Here, points close to the critical point shadow its orbit for a certain amount of time slowing down the mixing process like in the case of an indifferent fixed point if the derivative growth along the critical orbit is subexponential. In this paper we identify for the first time a general feature which plays an important role in determining the rate of decay of correlation for the system. This is the *degree of non-uniformity* of the expansivity which measures how close the system is to being uniformly expanding by quantifying the initial time one has to wait for typical points to start behaving as though the system were uniformly expanding. The precise definition will be given below.

We also obtain conditions for the validity of the Central Limit Theorem, which states that the probability of a given deviation of the average values of an observable along an orbit from the asymptotic average is essentially given by a Normal Distribution: given a Hölder continuous function  $\phi$  which is not a coboundary ( $\phi \neq \psi \circ f - \psi$  for any  $\psi$ ) there exists  $\sigma > 0$  such that for every interval  $J \subset \mathbb{R}$ ,

$$\mu \left\{ x \in X: \frac{1}{\sqrt{n}} \sum_{j=0}^{n-1} \left( \phi(f^j(x)) - \int \phi \, d\mu \right) \in J \right\} \rightarrow \frac{1}{\sigma \sqrt{2\pi}} \int_J e^{-t^2/2\sigma^2} \, dt.$$

We present our results first of all in the case of local diffeomorphisms and then in the case in which the map might contain discontinuities and/or critical points.

*Non-uniformly expanding local diffeomorphisms.* Let  $M$  be a compact Riemannian manifold of dimension  $d \geq 1$  and  $\text{Leb}$  a normalized Riemannian volume form on  $M$  that we call *Lebesgue measure*. Let  $f : M \rightarrow M$  be a  $C^2$  local diffeomorphism and suppose that there exists a constant  $\lambda > 0$  such that for Lebesgue almost all points  $x \in M$  the following *non-uniform expansivity* condition is satisfied:

$$\liminf_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \log \|Df_{f^i(x)}^{-1}\|^{-1} \geq \lambda > 0. \tag{*}$$

This formulation is motivated by the fact that we want to make an assumption about the *average expansion in every direction*. Indeed for a linear map  $A : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , the condition  $\|A\| > 1$  only provides information about the existence of *some* expanded direction, whereas the condition  $\|A^{-1}\|^{-1} > 1$  (i.e.  $\log \|A^{-1}\|^{-1} > 0$ ) is exactly equivalent to saying that *every* direction is expanded by  $A$ . Condition (\*) implies that the *expansion time* function

$$\mathcal{E}(x) = \min \left\{ N : \frac{1}{n} \sum_{i=0}^{n-1} \log \|Df_{f^i(x)}^{-1}\|^{-1} \geq \frac{\lambda}{2} \quad \forall n \geq N \right\}$$

is defined and finite almost everywhere in  $M$ . We think of this as the *waiting time* before the exponential derivative growth kicks in. Our results indicate that a main factor influencing the rate of decay of correlation is rate of decay of the tail of this function, i.e. the rate of decay of the measure of the set of points which have not yet started expanding uniformly by time  $n$ . We remark that the choice of  $\lambda/2$  in the definition of  $\mathcal{E}$  is just for convenience, any other positive constant smaller than  $\lambda$  would work and would yield the same results.

**Theorem 1.** *Let  $f : M \rightarrow M$  be a transitive  $C^2$  local diffeomorphism satisfying condition (\*) and suppose that there exists  $\gamma > 1$  such that*

$$\text{Leb}(\{\mathcal{E}(x) > n\}) \leq \mathcal{O}(n^{-\gamma}).$$

*Then there exists an absolutely continuous,  $f$ -invariant, probability measure  $\mu$ . Some finite power of  $f$  is mixing with respect to  $\mu$  and the correlation function  $C_n$  for Hölder continuous observable on  $M$  satisfies*

$$C_n \leq \mathcal{O}(n^{-\gamma+1}).$$

*Moreover, if  $\gamma > 2$  then the Central Limit Theorem holds.*

The existence and ergodicity of the measure  $\mu$  was proved in [3]. Our construction gives an alternative proof of the absolute continuity of  $\mu$  and allows us to obtain the estimates on the rate of Decay of Correlation and on the validity of the Central Limit Theorem. We remark that the questions concerning existence and ergodicity of an absolutely continuous invariant measure are quite distinct from the questions of the statistical properties with respect to the measure. Our results apply and are of interest even if an absolutely continuous, ergodic,  $f$ -invariant, probability measure  $\mu$  on  $M$  is already known to exist. In fact, in this case condition (\*) admits a very natural formulation simply in terms of the average

$$\int \log \|Df^{-1}\|^{-1} d\mu > 0.$$

Indeed Birkhoff’s Ergodic Theorem then implies that the limit

$$\lambda = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=0}^{n-1} \log \|Df_{f^i(x)}^{-1}\|^{-1} = \int \log \|Df^{-1}\|^{-1} d\mu > 0$$

exists for  $\mu$ -almost every  $x \in M$ . In particular the expansion time function  $\mathcal{E}(x)$  is also defined and finite almost everywhere and the conclusions of the Theorem hold under the given conditions on the rate of decay of the measure of  $\{\mathcal{E}(x) > n\}$ .

*Maps with critical points and discontinuities.* We can generalize our results to the case in which  $f$  is a local diffeomorphism outside a *critical/singular set*  $\mathcal{S} \subset M$  satisfying the following geometrical non-degeneracy conditions which essentially say that  $f$  behaves like a power of the distance to  $\mathcal{S}$ : there are constants  $B > 1$  and  $\beta > 0$  such that for every  $x \in M \setminus \mathcal{S}$

$$\frac{1}{B} \text{dist}(x, \mathcal{S})^\beta \leq \frac{\|Df(x)v\|}{\|v\|} \leq B \text{dist}(x, \mathcal{S})^{-\beta} \quad \text{for all } v \in T_x M; \tag{S1}$$

Moreover the functions  $\log \det Df$  and  $\log \|Df^{-1}\|$  are *locally Lipschitz* at points  $x \in M \setminus \mathcal{S}$  with Lipschitz constant depending on  $\text{dist}(x, \mathcal{S})$ : for every  $x, y \in M \setminus \mathcal{S}$  with  $\text{dist}(x, y) < \text{dist}(x, \mathcal{S})/2$  we have

$$|\log \|Df(x)^{-1}\| - \log \|Df(y)^{-1}\|| \leq B \frac{\text{dist}(x, y)}{\text{dist}(x, \mathcal{S})^\beta}; \tag{S2}$$

$$|\log |\det Df(x)^{-1}| - \log |\det Df(y)^{-1}|| \leq B \frac{\text{dist}(x, y)}{\text{dist}(x, \mathcal{S})^\beta}; \tag{S3}$$

We assume that orbits have *slow approximation* or *subexponential recurrence* to the critical set in the following sense. Let  $d_\delta(x, \mathcal{S})$  denote the  $\delta$ -truncated distance from  $x$  to  $\mathcal{S}$  defined as  $d_\delta(x, \mathcal{S}) = d(x, \mathcal{S})$  if  $d(x, \mathcal{S}) \leq \delta$  and  $d_\delta(x, \mathcal{S}) = 1$  otherwise. Then, given any  $\epsilon > 0$  there exists  $\delta > 0$  such that for Lebesgue almost every  $x \in M$

$$\limsup_{n \rightarrow +\infty} \frac{1}{n} \sum_{j=0}^{n-1} -\log \text{dist}_\delta(f^j(x), \mathcal{S}) \leq \epsilon. \tag{**}$$

Again this is an asymptotic statement and we have no a-priori knowledge about how fast this limit is approached or with what degree of uniformity for different points  $x$ . Since some control of the recurrence at finite times is important for our construction we introduce the *recurrence time* function

$$\mathcal{R}(x) = \min \left\{ N \geq 1: \frac{1}{n} \sum_{i=0}^{n-1} -\log \text{dist}_\delta(f^i(x), \mathcal{S}) \leq 2\epsilon, \forall n \geq N \right\}.$$

Condition (\*\*) implies that the *recurrence time* function is defined and finite almost everywhere in  $M$ . Before we state our results in this case, it will be useful to introduce for each  $n \geq 1$  the set

$$\Gamma_n = \{x: \mathcal{E}(x) > n \text{ or } \mathcal{R}(x) > n\}.$$

This is the set of points which at time  $n$  have not yet achieved either the uniform exponential growth or the uniform subexponential recurrence given by conditions (\*) and (\*\*). To be rigorous one should explicit the dependence of  $\mathcal{R}(x)$  and  $\Gamma_n$  on  $\epsilon$  and  $\delta$ . However, condition (\*\*) is not needed in all strength. Actually, we only need that condition in Lemma 2.2 to assure the existence of hyperbolic times. For this, it is enough to fix some  $\epsilon$  and  $\delta$  conveniently chosen; see [3, Lemma 5.4] for details.

**Theorem 2.** *Let  $f : M \rightarrow M$  be a transitive  $C^2$  local diffeomorphism outside a critical/singular set  $\mathcal{S}$  satisfying the non-degeneracy conditions stated above. Suppose that  $f$  satisfies the non-uniform expansivity condition (\*) and the slow approximation condition (\*\*) to the critical set and suppose that there exists  $\gamma > 1$  such that*

$$\text{Leb}(\Gamma_n) \leq \mathcal{O}(n^{-\gamma}).$$

*Then there exists an absolutely continuous,  $f$ -invariant, probability measure  $\mu$ . Some finite power of  $f$  is mixing with respect to  $\mu$  and the correlation function  $C_n$  for Hölder continuous observables on  $M$  satisfies*

$$C_n \leq \mathcal{O}(n^{-\gamma+1}).$$

*Moreover, if  $\gamma > 2$  then the Central Limit Theorem holds.*

Notice that the assumptions of Theorem 2 contain those of Theorem 1 as a special case where  $\mathcal{S} = \emptyset$ . We have stated the two results separately because the local diffeomorphism case is sufficiently interesting on its own and to emphasize the fact that the recurrence condition only applies to the case in which a critical and/or singular set exists. Both theorems extend to arbitrary dimension the results of [4] in which similar results were obtained for one-dimensional maps.

We remark also that even though condition (\*\*) is not needed in all its strength for the proof (it is sufficient that the statement holds for some  $\varepsilon$  sufficiently small depending on expansivity rate  $\lambda$  and on the constants  $B, \beta$  in the non-degeneracy conditions for the critical set), it is nevertheless more natural than it might appear at first sight. For example, if an ergodic,  $f$ -invariant, absolutely continuous probability measure  $\mu$  is given, then this condition just amounts to supposing that this invariant measure does not give too much weight to neighbourhoods of  $\mathcal{S}$  in the sense that

$$\int |\log \text{dist}(x, \mathcal{S})| d\mu < \infty.$$

Indeed, as for the expansivity condition, this immediately implies (\*\*) by Birkhoff’s Ergodic Theorem. Notice moreover that this integrability condition is satisfied if the singular set  $\mathcal{S}$  and the Radon–Nykodim derivative of  $\mu$  with respect to Lebesgue satisfy some mild regularity conditions.

### 1.2. Markov structure

Our strategy for proving the results stated above is to establish the existence of a *Markov tower* structure: a ball  $\Delta \subset M$  and a countable partition  $\mathcal{P}$  of  $\Delta$  into topological balls with smooth boundaries with the property that each element  $U$  of  $\mathcal{P}$  has an associated return time  $R(U)$  so that  $f^{R(U)}(U) = \Delta$  with some uniform bounds on the volume distortion between one return and the next. Moreover we set up a combinatorial and probabilistic argument which allows us to obtain estimates for the tail  $\text{Leb}(\{R > n\})$  of the return time function in terms of the tail of the expansivity and recurrence functions defined above.

**Main Theorem 1.** *Let  $f : M \rightarrow M$  be a transitive  $C^2$  local diffeomorphism outside a critical/singular set  $\mathcal{S}$  satisfying the non-degeneracy conditions stated above. Suppose that  $f$  satisfies the non-uniform expansivity condition (\*) and the slow approximation condition (\*\*). Then there exists a ball  $\Delta \subset M \setminus \mathcal{S}$ , a countable partition  $\mathcal{P}$  (mod 0) of  $\Delta$  into topological balls  $U$  with smooth boundaries, and a return time function  $R : \Delta \rightarrow \mathbb{N}$  piecewise constant on elements of  $\mathcal{P}$  satisfying the following properties:*

- (1) Markov: for each  $U \in \mathcal{P}$  and  $R = R(U)$ ,  $f^R : U \rightarrow \Delta$  is a  $C^2$  diffeomorphism (and in particular a bijection). Thus the induced map

$$F : \Delta \rightarrow \Delta \quad \text{given by } F(x) = f^{R(x)}(x)$$

is defined almost everywhere and satisfies the classical Markov property.

- (2) Uniform expansivity: There exists  $\hat{\lambda} > 1$  such that for almost all  $x \in \Delta$  we have

$$\|DF(x)^{-1}\|^{-1} \geq \hat{\lambda}.$$

In particular the separation time  $s(x, y)$  given by the maximum integer such that  $F^i(x)$  and  $F^i(y)$  belong to the same element of the partition  $\mathcal{P}$  for all  $i \leq s(x, y)$ , is defined and finite for almost every pair of points  $x, y \in \Delta$ .

- (3) Bounded volume distortion: There exist a constant  $K > 0$  such that for any pair of points  $x, y \in \Delta$  with  $\infty > s(x, y) \geq 1$  we have

$$\left| \frac{\det DF(x)}{\det DF(y)} - 1 \right| \leq K \hat{\lambda}^{-s(F(x), F(y))}.$$

Moreover, assuming polynomial decay of  $\text{Leb}(\Gamma_n)$  we have the same decay of return times:

(4) if there exists  $\gamma > 0$  such that  $\text{Leb}(\Gamma_n) \leq \mathcal{O}(n^{-\gamma})$ , then  $\text{Leb}(\{R > n\}) \leq \mathcal{O}(n^{-\gamma})$ .

In the particular case that  $\text{Leb}(\Gamma_n)$  decays faster than any polynomial we obtain *super-polynomial decay* for the tail of return times:  $\text{Leb}(\{R > n\}) \leq \mathcal{O}(n^{-\gamma})$  for every  $\gamma > 0$ .

We remark that the significance of the existence of a Markov Tower structures goes well beyond the consequences this has for the statistical properties of the map. It can be thought of as a partial generalization, to the framework of non-uniformly expanding maps, of the remarkable (and classical) Theorem of Bowen that any uniformly hyperbolic compact invariant set for a  $C^2$  diffeomorphisms admits a finite Markov partition ([11] see also [6,7,27]). Besides the intrinsic interest of such a statement, this fact has been used innumerable times in relation to all kinds of results concerning uniformly hyperbolic systems. There has been some success in the direct generalization of this result, for example to systems with discontinuities [14,21]. However the constructions always give rise to countable partitions and any conclusions about the invariant measures and other statistical properties then depends on a corresponding ergodic theory for countable subshifts which is much less developed than the finite case, although some results exist, see for example [13,23,25,26].

A significant break-through was achieved recently by Young in [29,30] where the general problem of proving the existence of Markov partitions was essentially reformulated in terms of proving the existence of *Markov Towers* or *induced Markov maps* as defined above. One important advantage of these structures is that statistical information about the system is deduced from statistical information about the return times and not encoded in some kind of transition matrix which would in general be very hard to determine. Moreover the actual construction of Markov Towers has at least two significant advantages. Firstly, one can choose conveniently some small region of the dynamical phase space, instead of having to construct a partition of the entire space, and use *approximate* information about the remaining part of the space to construct a return map. Secondly, one does not need a single iterate of the map to have special (Markov) properties, but is allowed to wait a certain amount of time until this property is obtained. Most importantly, only some approximate (statistical) information is required concerning the length of this waiting time.

### 1.3. Viana maps

An important class of non-uniform expanding dynamical systems (with critical sets) in dimension greater than one was introduced by Viana in [28]. This has served as a model for some relevant results on the ergodic properties of non-uniformly expanding maps in higher dimensions; see [1–3,5].

This class of maps can be described as follows. Let  $a_0 \in (1, 2)$  be such that the critical point  $x = 0$  is pre-periodic for the quadratic map  $Q(x) = a_0 - x^2$ . Let  $S^1 = \mathbb{R}/\mathbb{Z}$  and  $b: S^1 \rightarrow \mathbb{R}$  be a Morse function, for instance,  $b(s) = \sin(2\pi s)$ . For fixed small  $\alpha > 0$ , consider the map

$$\begin{aligned} \hat{f}: S^1 \times \mathbb{R} &\longrightarrow S^1 \times \mathbb{R}, \\ (s, x) &\longmapsto (\hat{g}(s), \hat{q}(s, x)) \end{aligned}$$

where  $\hat{q}(s, x) = a(s) - x^2$  with  $a(s) = a_0 + \alpha b(s)$ , and  $\hat{g}$  is the uniformly expanding map of the circle defined by  $\hat{g}(s) = ds \pmod{\mathbb{Z}}$  for some large integer  $d$ . In fact,  $d$  was chosen greater or equal to 16 in [28], but recent results in [18] showed that some estimates in [28] can be improved and  $d = 2$  is enough. It is easy to check that for  $\alpha > 0$  small enough there is an interval  $I \subset (-2, 2)$  for which  $\hat{f}(S^1 \times I)$  is contained in the interior of  $S^1 \times I$ . Thus, any map  $f$  sufficiently close to  $\hat{f}$  in the  $C^0$  topology has  $S^1 \times I$  as a forward invariant region. We consider from here on these maps restricted to  $S^1 \times I$ .

Taking into account the expression of  $\hat{f}$  it is not difficult to check that it behaves like a power of the distance close to the critical set  $\{x = 0\}$ . Moreover, there is a small neighbourhood  $\mathcal{N}$  of  $\hat{f}$  in the  $C^3$  topology of maps from

$S^1 \times I$  into itself, such that any  $f \in \mathcal{N}$  also behaves like a power of the distance close to its critical set, which is close to  $\{x = 0\}$ . The most important results for  $f \in \mathcal{N}$  are summarized below:

- (1)  $f$  is non-uniformly expanding and its orbits have slow approximation to the critical set [28,2];
- (2) there are constants  $C, c > 0$  such that  $\text{Leb}(\Gamma_n) \leq Ce^{-c\sqrt{n}}$  for every  $n \geq 1$  [28,2];
- (3)  $f$  is topologically mixing and has a unique ergodic absolutely continuous invariant (thus SRB) measure [1,5];
- (4) the density of the SRB measure varies continuously in the  $L^1$  norm with  $f$  [5];
- (5)  $f$  is stochastically stable [2].

As a consequence of our theorems, we obtain the following result:

**Theorem 3.** *Any  $f \in \mathcal{N}$  has super-polynomial decay of correlations and Central Limit Theorem holds for  $f$ .*

The decay of correlations for Viana maps has remained unknown for several years. Some results have been obtained in [8] for the (stretched exponential) decay of correlations of a skew product system closely related to Viana maps, but these results do not apply immediately to Viana maps themselves. One of the initial purposes of the present work was precisely to attempt to make this additional step. In some sense our results go much further as they apply to non-uniformly expanding systems in much greater generality. We have paid a price however, as we do not succeed in obtaining stretched exponential estimates in this case. An optimal estimate for the decay of correlations of Viana maps continues to be elusive, perhaps indicating that there is some additional subtlety which we do not yet understand.

#### 1.4. Remarks

Before starting the proof of the Main Theorem we discuss our basic strategy and the main technical issues involved in the construction. We also mention some recent work related to the research presented here.

*Strategy.* We start by choosing essentially arbitrarily a point  $p$  with dense pre-images and some sufficiently small ball  $\Delta_0$  around this point. This will be the domain of definition of our induced map. We then attempt to implement the naive strategy of iterating  $\Delta_0$  until we find some good return iterate  $n_0$  such that  $f^{n_0}(\Delta_0)$  completely covers  $\Delta_0$  and some bounded distortion property is satisfied. There exists then some topological ball  $U \subset \Delta_0$  such that  $f^{n_0}(U) = \Delta_0$ . This ball is then by definition an element of the final partition of  $\Delta_0$  for the induced Markov map and has an associated return time  $n_0$ . We then continue iterating the complement  $\Delta_0 \setminus U$  until more good returns occur. Most of the paper is dedicated to showing that this strategy can indeed be implemented in a successful way, yielding a partition (mod 0) of  $\Delta_0$  into piecewise disjoint subsets, and an associated return time function which is Lebesgue integrable. The construction also yields substantial information about the tail of the return time function, i.e. the decay of the measure of the set of points whose return time is larger than  $n$ . Indeed the main motivation for this paper is to show that the rate of decay of this tail is closely related to the rate at which the derivative along orbits approaches the asymptotic expansion rate.

*Technical issues.* There are two main technical difficulties, distinct but related to each other, in carrying out the plan suggested above. The first has to do with the geometry of the returns to  $\Delta_0$ , and in particular of the geometry of the set of points which does not return at a given time. Such a set can be visualized as a ball  $\Delta_0$  containing an increasing number of smaller topological balls corresponding to the elements of the final partition which have return times smaller than  $n$ . The exact location and shape of these smaller disks is quite difficult to control, as is the location and shape of their images at time  $n$ . Therefore some care is required, as well as the introduction of some auxiliary partitions and waiting times, to make sure that the set of points returning at time  $n$  is disjoint from the set of points which have already returned at some earlier time. These geometrical issues are essentially related

to the higher dimensional nature of the dynamics and arise also in the uniformly expanding case. This case has been treated in [29] and we follow essentially the same strategy and notation here. We still give all the details, for completeness and to make sure that any further problems associated to the non-uniformity of the expansion are dealt with as well.

The second technical problem, on the other hand, is precisely due to the strictly non-uniform nature of expansion in our situation. The process of defining the set  $\{R \leq n\}$  of points which have an associated return time less than or equal to  $n$ , as a union of disjoint topological disks in  $\Delta_0$ , gives rise to very “small” regions in the complement  $\{R > n\}$ , i.e. regions which are squeezed into strange shapes by the geometry of the previous returns. It is important to control the extent to which this can happen and to show that even these small regions eventually grow large enough so that they can cover  $\Delta_0$  and thus contain an element of the partition for the induced map. In the uniformly hyperbolic case, once the suitable definitions and notation have been introduced, a relatively straightforward calculation shows this to be the case and shows that in fact this growth of small regions to uniformly large scale occurs uniformly exponentially fast. In our context we only have much more abstract information about the eventual expansion at almost every point and therefore this part of the argument is more subtle.

We shall use the idea of *Hyperbolic Times* to show that our assumptions imply that almost every point has a basis of (arbitrarily small) neighbourhoods which at some time are mapped to uniformly large scale with bounded distortion. It follows that the speed at which this large scale is achieved is not uniform but rather depends on the distribution of hyperbolic times associated to points in the regions in question, which can be arbitrarily large. We conclude that the final return time function for the Markov induced map is related to the statistics of hyperbolic times. Since hyperbolic times are naturally related to speed at which some uniform expansion estimates begin to hold, this yields our desired conclusions.

One of the key issues we have to address is the relation between the statistics of hyperbolic times, the spatial distribution of points having hyperbolic time at some given time, and the geometrical structure of sets arising from the construction of the partition described above; see Corollary 2.3, Eq. (6) and Proposition 6.1. We are able to implement a partially successful strategy in this respect: in the polynomial case we establish an essentially optimal link between the rate of decay of the expansion/recurrence function and the rate of decay of correlations. The nature of the argument does not immediately extend to the exponential case.

*Related work.* In addition to the references given above we mention some related work of Buzzi and Maume-Deschamps [15–17] in which the decay of correlations on towers associated to some uniformly and non-uniformly expanding maps are also studied. There is an empty intersection between the results presented there and the ones of the present paper, as the assumptions there are quite different from ours and provide a different point of view on causes for various rates of decay of correlations. Huyi also addresses specifically the question of why certain maps should have slow decay of correlations in the context of multidimensional maps with an indifferent fixed point [31].

*Overview of the paper.* The paper is completely dedicated to the proof of the Main Theorem on the existence of the Markov tower and the associated tail estimates. By recent results of Young [29,30] the rate of decay of the tail of the return time function in this framework has direct implications for the rates of decay of correlations and the Central Limit Theorem and therefore Theorems 1 and 2 follow by an application of her results.

In Section 2 we give several estimates related to the time it takes for small domains to grow to some fixed size while preserving some bounded distortion properties. In Section 3 we give the precise algorithm for constructing the Markov Tower and describe the associated combinatorial information. The final three Sections 4, 5 and 6 are dedicated to proving that this algorithm effectively results in a countable partition (mod 0) with the required properties.



## 2. Growing to large scale

In this section we give the basic growth estimates on which the algorithm for the construction of the Markov Tower is based. First of all we define the notion of Hyperbolic Time and show that almost all points have an infinite basis of neighbourhoods which grow to some fixed size with bounded distortion for some corresponding infinite sequence of hyperbolic times. The set of hyperbolic times depends on the point and the first hyperbolic time for a given point can be arbitrarily large in general, although we do have some degree of control since it is related to the values of the expansivity and recurrence functions  $\mathcal{E}$  and  $\mathcal{R}$  at that point. Next we prove a useful and non-obvious consequence of our assumptions, namely that if we fix some  $\varepsilon > 0$  then there exist some  $N_\varepsilon$  depending only on  $\varepsilon$  such that any ball of radius  $\varepsilon$  has some subset which grows to a fixed size with bounded distortion within  $N_\varepsilon$  iterates. Finally we show that our “base”  $\Delta_0$  can be chosen in such a way that any other sufficiently large ball contains a subset which is mapped bijectively to  $\Delta_0$  with bounded distortion and within some fixed number of iterates. A combination of these estimate will play a crucial role in obtaining control of the tail of the return times to  $\Delta_0$ .

*Hyperbolic times: growing to uniform scale in variable time.* Let  $B > 1$  and  $\beta > 0$  be as in the hypotheses (S1)–(S3). In what follows  $b$  is any fixed constant satisfying  $0 < b < \min\{1/2, 1/(4\beta)\}$ . Given  $\sigma < 1$  and  $\delta > 0$ , we say that  $n$  is a  $(\sigma, \delta)$ -hyperbolic time for a point  $x \in M$  if for all  $1 \leq k \leq n$ ,

$$\prod_{j=n-k}^{n-1} \|Df(f^j(x))^{-1}\| \leq \sigma^k \quad \text{and} \quad \text{dist}_\delta(f^{n-k}(x), S) \geq \sigma^{bk}.$$

For each  $n \geq 1$  we define

$$H_n = H_n(\sigma, \delta) = \{x \in M: n \text{ is a } (\sigma, \delta)\text{-hyperbolic time for } x\}.$$

We give two well-established results which show that (i) if  $n$  is a hyperbolic time for  $x$ , the map  $f^n$  is a diffeomorphism with uniformly bounded volume distortion on a neighborhood of  $x$  which is mapped to a disk of uniform radius; (ii) almost every point has lots of hyperbolic times. We say that  $f^n$  has *volume distortion* bounded by  $D$  on a set  $V$  if, for every  $x, y \in V$ ,

$$\frac{1}{D} \leq \frac{|\det Df^n(x)|}{|\det Df^n(y)|} \leq D.$$

**Lemma 2.1.** *Given  $\sigma < 1$  and  $\delta > 0$ , there exist  $\delta_1, D_1, \kappa > 0$ , depending only on  $\sigma, \delta$  and on the map  $f$ , such that for any  $x \in M$  and  $n \geq 1$  a  $(\sigma, \delta)$ -hyperbolic time for  $x$ , there exists a neighborhood  $V_n(x)$  of  $x$  with the following properties:*

- (1)  $f^n$  maps  $V_n(x)$  diffeomorphically onto the ball  $B(f^n(x), \delta_1)$ ;
- (2) for  $1 \leq k < n$  and  $y, z \in V_n(x)$ ,  $\text{dist}(f^{n-k}(y), f^{n-k}(z)) \leq \sigma^{k/2} \text{dist}(f^n(y), f^n(z))$ ;
- (3)  $f^n$  has volume distortion bounded by  $D_1$  on  $V_n(x)$ ;
- (4)  $V_n(x) \subset B(x, \kappa^{-n})$ .

**Proof.** For the proofs of items (1)–(3) see Lemma 5.2 and Corollary 5.3 in [3]. Item (4) is an immediate consequence of item (2).  $\square$

We shall often refer to the sets  $V_n(x)$  as *hyperbolic pre-balls* and to their images  $f^n(V_n(x))$  as *hyperbolic balls*. Notice that the latter are indeed balls of radius  $\delta_1$ .

**Lemma 2.2.** *There exists  $\theta > 0$  and  $\delta > 0$  depending only on  $f$  and  $\lambda$  such that for Lebesgue almost every  $x \in M$  and  $n \geq \mathcal{E}(x)$  there exist  $(\sigma, \delta)$ -hyperbolic times  $1 \leq n_1 < \dots < n_l \leq n$  for  $x$  with  $l \geq \theta n$ .*

**Proof.** See Lemma 5.4 of [3]. Let us remark for the sake of completeness that the proof of the lemma gives  $\sigma = e^{-\lambda/4}$ .  $\square$

**Corollary 2.3.** For every  $n \geq 1$  and every  $A \subset M \setminus \Gamma_n$  with positive Lebesgue measure we have

$$\frac{1}{n} \sum_{j=1}^n \frac{\text{Leb}(A \cap H_j)}{\text{Leb}(A)} \geq \theta.$$

**Proof.** Take  $n \geq 1$  and  $A \subset M \setminus \Gamma_n$  with positive Lebesgue measure. Observe that  $n \geq \mathcal{E}(x)$  for all  $x \in A$ , by definition of  $\mathcal{E}(x)$ . Let  $\xi_n$  be the measure in  $\{1, \dots, n\}$  defined by  $\xi_n(J) = \#J/n$ , for each subset  $J$ . Then, using Fubini’s theorem

$$\frac{1}{n} \sum_{j=1}^n \text{Leb}(A \cap H_j) = \int \left( \int_A \chi(x, i) \, d\text{Leb}(x) \right) d\xi_n(i) = \int \left( \int_A \chi(x, i) \, d\xi_n(i) \right) d\text{Leb}(x),$$

where  $\chi(x, i) = 1$  if  $x \in H_i$  and  $\chi(x, i) = 0$  otherwise. Now, Lemma 2.2 means that the integral with respect to  $d\xi_n$  is larger than  $\theta > 0$ . So, the last expression above is bounded from below by  $\theta \text{Leb}(A)$ .  $\square$

*Growing to uniform scale in uniform time.* Now we show a simple (albeit slightly counterintuitive) fact that any  $\varepsilon$  ball has a subset which grows to fixed size within some uniformly bounded maximum number of iterates.

**Lemma 2.4.** For each  $\varepsilon > 0$  there exists  $N_\varepsilon > 0$  such that any ball  $B \subset M$  of radius  $\varepsilon > 0$  contains a hyperbolic pre-ball  $V_n \subset B$  with  $n \leq N_\varepsilon$ .

**Proof.** Given  $\varepsilon > 0$  and a ball  $B(z, \varepsilon)$ , choose  $N'_\varepsilon$  large enough so that any hyperbolic pre-ball  $V_n$  associated to a hyperbolic time  $n \geq N'_\varepsilon$  will be contained in a ball of radius  $\varepsilon/10$  ( $N'_\varepsilon \sim \kappa^{-1} \log(10\varepsilon^{-1})$ ). Now notice that each point has an infinite number of hyperbolic times and therefore we have that

$$\text{Leb} \left( M \setminus \bigcup_{j=N'_\varepsilon}^n H_j \right) \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

Therefore it is possible to choose

$$N_\varepsilon = \min \left\{ n \geq N'_\varepsilon : \text{Leb} \left( M \setminus \bigcup_{j=N'_\varepsilon}^n H_j \right) \leq \varepsilon^d / 10 \right\}$$

where  $d$  is the dimension of  $M$ . This ensures that there is a point  $\hat{x} \in B(z, \varepsilon/2)$  with a hyperbolic time  $n \leq N_\varepsilon$  and associated hyperbolic pre-ball  $V_n(x) \subset B(z, \varepsilon)$ .  $\square$

*Returning to a given domain.* Now we derive an useful consequence of the transitivity of  $f$ . Given  $\delta > 0$ , we say that a subset  $A$  of  $M$  is  $\delta$ -dense if any point in  $M$  is at a distance smaller than  $\delta$  from  $A$ .

**Lemma 2.5.** Given  $\delta > 0$  there is  $p \in M$  and  $N_0 \in \mathbb{N}$  such that  $\bigcup_{j=0}^{N_0} f^{-j}(\{p\})$  is  $\delta$  dense in  $M$  and disjoint from  $S$ .

**Proof.** Observe that the properties of  $f$  imply that the images and preimages of sets with zero Lebesgue measure still have zero Lebesgue measure. Hence, the set

$$\mathcal{B} = \bigcup_{n \geq 0} f^{-n} \left( \bigcup_{m \geq 0} f^{-m}(S) \right)$$

has Lebesgue measure equal to zero. On the other hand, since  $f$  is transitive, we have by [3] that there is a unique SRB measure for  $\mu$ , which is an ergodic and absolutely continuous with respect to Lebesgue measure, and whose support is the whole manifold  $M$ . Moreover, the ergodicity of  $\mu$  implies that  $\mu$  almost every point in  $M$  has a dense orbit. Since  $\mu$  is absolutely continuous with respect to Leb, then there is a positive Lebesgue measure subset of points in  $M$  with dense orbit. Thus there must be some point  $q \in M \setminus \mathcal{B}$  with dense orbit. Take  $N_0 \in \mathbb{N}$  for which  $q, f(q), \dots, f^{N_0}(q)$  is  $\delta$ -dense. The point  $p = f^{N_0}(q)$  satisfies the conclusions of the lemma.  $\square$

We fix once and for all  $p \in M$  and  $N_0 \in \mathbb{N}$  for which  $\bigcup_{j=0}^{N_0} f^{-j}(\{p\})$  is  $\delta_1/3$  dense in  $M$  and disjoint from the critical set  $\mathcal{S}$ . Recall that  $\delta_1 > 0$  is the radius of hyperbolic balls given by Lemma 2.1. Take constants  $\varepsilon > 0$  and  $\delta_0 > 0$  so that

$$\sqrt{\delta_0} \ll \delta_1/2 \quad \text{and} \quad 0 < \varepsilon \ll \delta_0.$$

**Lemma 2.6.** *There exist a constant  $D_0 > 0$  depending only on  $f, \sigma, \delta_1$  and the point  $p$ , such that for any ball  $B \subset M$  of radius  $\delta_1$  (in particular for any hyperbolic ball), there exist an open set  $V \subset B$  and an integer  $0 \leq m \leq N_0$  for which*

- (1)  $f^m$  maps  $V$  diffeomorphically onto  $B(p, 2\sqrt{\delta_0})$ ;
- (2)  $f^m|_V$  has volume distortion bounded by  $D_0$ .

**Proof.** Since  $\bigcup_{j=0}^{N_0} f^{-j}(\{p\})$  is  $\delta_1/3$  dense in  $M$  and disjoint from  $\mathcal{S}$ , choosing  $\delta_0 > 0$  sufficiently small we have that each connected component of the preimages of  $B(p, 2\sqrt{\delta_0})$  up to time  $N_0$  are bounded away from the critical set  $\mathcal{S}$  and are contained in a ball of radius  $\delta_1/3$ .

This immediately implies that any ball  $B \subset M$  of radius  $\delta_1$  contains a preimage  $V$  of  $B(p, 2\sqrt{\delta_0})$  which is mapped diffeomorphically onto  $B(p, 2\sqrt{\delta_0})$  in at most  $N_0$  iterates. Moreover, since the number of iterations and the distance to the critical region are uniformly bounded, the volume distortion is uniformly bounded.  $\square$

**Remark 2.7.** It will be useful to emphasize that  $\delta_0$  and  $N_0$  have been chosen in such a way that all the connected component of the preimages of  $B(p, 2\sqrt{\delta_0})$  up to time  $N_0$  satisfy the conclusions of the lemma. In particular, they are uniformly bounded away from the critical set  $\mathcal{S}$ , and so there is some constant  $C_0 > 1$  depending only on  $f$  and  $\delta_1$  such that

$$\frac{1}{C_0} \leq \|Df^m(x)\|, \|(Df^m(x))^{-1}\| \leq C_0$$

for all  $1 \leq m \leq N_0$  and  $x$  belonging to an  $m$ -preimage of  $B(p, 2\sqrt{\delta_0})$ .

### 3. The partitioning algorithm

We now describe the construction of the (mod 0) partition of  $\Delta_0 = B(p, \delta_0)$ . The basic intuition is that we wait for some iterate  $f^k(\Delta_0)$  to cover  $\Delta_0$  completely, and then define the subset  $U \subset \Delta_0$  such that  $f^k : U \rightarrow \Delta_0$  is a diffeomorphism, as an element of the partition with return time  $k$ . We then continue to iterate the complement  $\Delta_0 \setminus U$  until this complement covers again  $\Delta_0$  and repeat the same procedure to define more elements of the final partition with higher return times. Using the fact that small regions eventually become large due to the expansivity condition (and the lemmas given above), it follows that this process can be continued and that Lebesgue almost every point eventually belongs to some element of the partition and that the return time function depends on the time that it takes small regions to become large on average and this turns out to depend precisely on the tail of the expansivity condition function.

The formalization of this argument requires several technical constructions which we explain below. The construction is inductive and we give precisely the general step of the induction below. For the sake of a better visualization of the process, and to motivate the definitions, we start with the first step.

*First step of the induction.* First of all we introduce neighborhoods of  $p$

$$\Delta_0^0 = \Delta_0 = B(p, \delta_0), \quad \Delta_0^1 = B(p, 2\delta_0), \quad \Delta_0^2 = B(p, \sqrt{\delta_0}) \quad \text{and} \quad \Delta_0^3 = B(p, 2\sqrt{\delta_0}).$$

For  $0 < \sigma < 1$  given by Lemma 2.1, let

$$I_k = \{x \in \Delta_0^1: \delta_0(1 + \sigma^{k/2}) < \text{dist}(x, p) < \delta_0(1 + \sigma^{(k-1)/2})\}, \quad k \geq 1,$$

be a partition (mod 0) into countably many rings of  $\Delta_0^1 \setminus \Delta_0$ . Take  $R_0$  some large integer to be determined below; we ignore any dynamics occurring up to time  $R_0$ . Let  $k \geq R_0 + 1$  be the first time that  $\Delta_0 \cap H_k \neq \emptyset$ . For  $j < k$  we define formally the objects  $\Delta_j, A_j, A_j^\varepsilon$  whose meaning will become clear in the next paragraph, by  $A_j = A_j^\varepsilon = \Delta_j = \Delta_0$ . Let  $(U_{k,j}^3)_j$  be the connected components of  $f^{-k}(\Delta_0^3) \cap A_{k-1}^\varepsilon$  contained in hyperbolic pre-balls  $V_m$  with  $k - N_0 \leq m \leq k$  which are mapped diffeomorphically onto  $\Delta_0^3$  by  $f^k$ . Now let

$$U_{k,j}^i = U_{k,j}^3 \cap f^{-k} \Delta_0^i, \quad i = 0, 1, 2,$$

and set  $R(x) = k$  for  $x \in U_{k,j}^0$ . Now take

$$\Delta_k = \Delta_{k-1} \setminus \{R = k\}.$$

We define also a function  $t_k: \Delta_k \rightarrow \mathbb{N}$  by

$$t_k(x) = \begin{cases} s & \text{if } x \in U_{k,j}^1 \text{ and } f^k(x) \in I_s \text{ for some } j; \\ 0 & \text{otherwise.} \end{cases}$$

Finally let

$$A_k = \{x \in \Delta_k: t_k(x) = 0\}, \quad B_k = \{x \in \Delta_k: t_k(x) > 0\}$$

and

$$A_k^\varepsilon = \Delta_k \cap \bigcup_{x \in A_k \cap H_{k+1}} (f|_{V_{k+1}(x)})^{-1} B(f^{k+1}(x), \varepsilon).$$

Note that this is a small neighborhood of  $A_k \cap H_{k+1}$  whose iterates at time  $k + 1$  are at a distance smaller than  $\varepsilon$  from  $f^{k+1}(A_k \cap H_{k+1})$ .

*General step of the induction.* The general inductive step of the construction now follows by repeating the arguments above with minor modifications. More precisely we assume that sets  $\Delta_i, A_i, A_i^\varepsilon B_i, \{R = i\}$  and functions  $t_i: \Delta_i \rightarrow \mathbb{N}$  are defined for all  $i \leq n - 1$ . For  $i \leq R_0$  we just let  $A_i = A_i^\varepsilon = \Delta_i = \Delta_0, B_i = \{R = i\} = \emptyset$  and  $t_i \equiv 0$ . Now let  $(U_{n,j}^3)_j$  be the connected components of  $f^{-n}(\Delta_0^3) \cap A_{n-1}^\varepsilon$  contained in hyperbolic pre-balls  $V_m$ , with  $n - N_0 \leq m \leq n$ , which are mapped onto  $\Delta_0^3$  by  $f^n$ . Take

$$U_{n,j}^i = U_{n,j}^3 \cap f^{-n} \Delta_0^i, \quad i = 0, 1, 2,$$

and set  $R(x) = n$  for  $x \in U_{n,j}^0$ . Take also

$$\Delta_n = \Delta_{n-1} \setminus \{R = n\}.$$

The definition of the function  $t_n: \Delta_n \rightarrow \mathbb{N}$  is slightly different in the general case.

$$t_n(x) = \begin{cases} s & \text{if } x \in U_{n,j}^1 \setminus U_{n,j}^0 \text{ and } f^n(x) \in I_s \text{ for some } j, \\ 0 & \text{if } x \in A_{n-1} \setminus \bigcup_j U_{n,j}^1, \\ t_{n-1}(x) - 1 & \text{if } x \in B_{n-1} \setminus \bigcup_j U_{n,j}^1. \end{cases}$$

Finally let

$$A_n = \{x \in \Delta_n : t_n(x) = 0\}, \quad B_n = \{x \in \Delta_n : t_n(x) > 0\}$$

and

$$A_n^\varepsilon = \Delta_n \cap \bigcup_{x \in A_n \cap H_{n+1}} (f|_{V_{n+1}(x)})^{-1} B(f^{n+1}(x), \varepsilon).$$

At this point we have completely described the inductive construction of the sets  $A_n$ ,  $A_n^\varepsilon$ ,  $B_n$  and  $\{R = n\}$ .

We conclude this section with a remark concerning the role of the sets  $B_n$  as a kind of shield protecting the sets of the partition constructed up to time  $n$ , and some observations to motivate the last two sections.

*A remark on the construction.* Associated to each component  $U_{n-k}^0$  of  $\{R = n - k\}$ , for some  $k > 0$ , we have a collar  $U_{n-k}^1 \setminus U_{n-k}^0$  around it; knowing that the new components of  $\{R = n\}$  do not “intersect too much”  $U_{n-k}^1 \setminus U_{n-k}^0$  is important for preventing overlaps on sets of the partition. We will see that this is indeed the case as long as  $\varepsilon > 0$  is taken small enough.

**Lemma 3.1.** *If  $\varepsilon > 0$  is sufficiently small, then  $U_n^1 \cap \{t_{n-1} > 1\} = \emptyset$  for each  $U_n^1$ .*

**Proof.** Take some  $k > 0$  and let  $U_{n-k}^0$  be a component of  $\{R = n - k\}$ . Let  $Q_k$  be the part of  $U_{n-k}^1$  that is mapped by  $f^{n-k}$  onto  $I_k$  and assume that  $Q_k$  intersects some  $U_n^3$ . Recall that, by construction,  $Q_k$  is precisely the part of  $U_{n-k}^1$  on which  $t_{n-1}$  takes the value 1, and  $U_n^3$  is contained in a hyperbolic pre-ball  $V_m$  with  $k - N_0 \leq m \leq k$ .

Let  $q_1$  and  $q_2$  be any two points in distinct components (inner and outer, respectively) of the boundary of  $Q_k$ . Assuming that  $q_1, q_2 \in U_n^3$ , then  $q_1, q_2 \in V_m$ , and so by Lemma 2.1 and Remark 2.7

$$\text{dist}(f^{n-k}(q_1), f^{n-k}(q_2)) \leq C_0 \sigma^{(k-N_0)/2} \text{dist}(f^n(q_1), f^n(q_2)). \tag{1}$$

We also have

$$\text{dist}(f^{n-k}(q_1), f^{n-k}(q_2)) \geq \delta_0(1 + \sigma^{(k-1)/2}) - \delta_0(1 + \sigma^{k/2}) = \delta_0 \sigma^{k/2} (\sigma^{-1/2} - 1),$$

which combined with (1) gives

$$\text{dist}(f^n(q_1), f^n(q_2)) \geq C_0^{-1} \sigma^{N_0/2} \delta_0 (\sigma^{-1/2} - 1).$$

On the other hand, since  $U_n^3 \subset A_{n-1}^\varepsilon$  by construction of  $U_n^3$ , taking

$$\varepsilon < C_0^{-1} \sigma^{N_0/2} \delta_0 (\sigma^{-1/2} - 1)$$

we have  $U_n^3 \cap \{t_{n-1} > 1\} = \emptyset$ . This implies  $U_n^1 \cap \{t_{n-1} \geq 1\} = \emptyset$ .  $\square$

#### 4. The induced map

In this section we briefly discuss the first and fourth items and prove the second and third items in the statement of the Main Theorem.

*The Markov property.* The construction detailed in Section 3 provides an algorithm for the definition of a family of topological balls contained in  $\Delta$  and satisfying the Markov property as required. In the next two sections we show that this algorithm does indeed produce a partition mod 0 of  $\Delta$  and obtain estimates for the rate of decay of the tail of the return times.

*Uniform expansivity.* Recall that by construction, the return time  $R(U)$  for  $U$  an element of the partition  $\mathcal{P}$  of  $\Delta$ , is formed by a certain number  $n$  of iterations given by the hyperbolic time of a hyperbolic pre-ball  $V_n \supset U$ , and a certain number  $m \leq N_0$  of additional iterates which is the time it takes to go from  $f^n(V_n)$  which could be anywhere in  $M$ , to  $f^{n+m}(V_n)$  which covers  $\Delta$  completely. By choosing  $R_0$  sufficiently large it then follows from Remark 2.7 that there exists a constant  $\hat{\lambda} > 1$  and a time  $n_0$  such that for any hyperbolic time  $n \geq n_0$  and any point  $x \in V_n$  and  $1 \leq m \leq N_0$ , we have

$$\|(Df^{n+m}(x))^{-1}\|^{-1} \geq \hat{\lambda} > 1.$$

We immediately have the uniform expansivity property of the Main Theorem

$$\|(DF_x)^{-1}\|^{-1} = \|(Df_x^{R(x)})^{-1}\|^{-1} \geq \hat{\lambda} > 1.$$

In particular, this implies that for any  $x, y \in \Delta$  which have the same combinatorics, i.e. which remain in the same elements of the partition  $\mathcal{P}$  for some number  $s(x, y)$  of iterates of the induced map  $F$ , we have

$$\text{dist}(x, y) \leq \hat{\lambda}^{-s(x,y)}. \tag{2}$$

*Distortion estimates.* The distortion estimate required for our Main Theorem follows immediately from (2) above and the following more classical formulation of the bounded distortion property:

**Lemma 4.1.** *There exists a constant  $\tilde{B} > 0$  such that for any  $x, y$  belonging to the same element  $U \in \mathcal{P}$  with return time  $R$ , we have*

$$\log \left| \frac{\det DF(x)}{\det DF(y)} \right| = \log \left| \frac{\det Df^R(x)}{\det Df^R(y)} \right| \leq \tilde{B} \text{dist}(f^R(x), f^R(y)).$$

**Proof.** Recall that by construction, the return time  $R(U)$  for  $U$  an element of the partition  $\mathcal{P}$  of  $\Delta$ , is formed by a certain number  $n$  of iterations given by the hyperbolic time of a hyperbolic pre-ball  $V_n \supset U$ , and a certain number  $m \leq N_0$  of additional iterates which is the time it takes to go from  $f^n(V_n)$  which could be anywhere in  $M$ , to  $f^{n+m}(V_n)$  which covers  $\Delta$  completely. Some standard formal manipulation based on the chain rule gives

$$\log \left| \frac{\det Df^R(x)}{\det Df^R(y)} \right| = \log \left| \frac{\det Df^{R-n}(f^n(x))}{\det Df^{R-n}(f^n(y))} \right| + \log \left| \frac{\det Df^n(x)}{\det Df^n(y)} \right|.$$

Since  $f^i(x)$  and  $f^i(y)$  are uniformly bounded away from  $\mathcal{S}$  for  $n \leq i \leq R$  (recall Remark 2.7), we may write

$$\log \left| \frac{\det Df^{R-n}(f^n(x))}{\det Df^{R-n}(f^n(y))} \right| \leq B_1 \text{dist}(f^R(x), f^R(y))$$

where  $B_1$  is some constant not depending on  $x, y$  or  $R$ . On the other hand, by construction of  $V_n$  (see the proof of Lemma 5.2 in [3]), there must be some  $z \in V_n$  for which  $n$  is a hyperbolic time and such that, for  $0 \leq j < n$ , the distance from  $f^j(z)$  to either  $f^j(x)$  or  $f^j(y)$  is smaller than  $\text{dist}(f^n(x), f^n(y))\sigma^{(n-j)/2}$ , which is much smaller than  $\sigma^{b(n-j)} \leq \text{dist}(f^j(z), \mathcal{S})$ . Thus, by (S3) we have

$$\log \left| \frac{\det Df^n(x)}{\det Df^n(y)} \right| \leq \sum_{j=0}^{n-1} \log \left| \frac{\det Df(f^j(x))}{\det Df(f^j(y))} \right| \leq \text{dist}(f^n(x), f^n(y)) \sum_{j=0}^{n-1} 2B \frac{\sigma^{(n-j)/2}}{\sigma^{b\beta(n-j)}}.$$

Since  $b\beta < 1/2$ , there must be some  $B_2 > 0$  such that

$$\log \left| \frac{\det Df^n(x)}{\det Df^n(y)} \right| \leq B_2 \text{dist}(f^n(x), f^n(y)).$$

Using again that  $f^i(y)$  and  $f^i(x)$  are uniformly bounded away from  $S$  (for  $n \leq i \leq R$  (cf. Remark 2.7)) it follows that

$$\text{dist}(f^n(x), f^n(y)) \leq B_2 \text{dist}(f^R(x), f^R(y)),$$

where  $B_2$  is some constant not depending on  $x, y$  or  $R$ . This completes the proof of the lemma.  $\square$

*Looking ahead: probabilistic estimates.* For proving the Main Theorem we only need to study the decay of  $\text{Leb}(\Delta_n)$  in terms of  $\text{Leb}(\Gamma_n)$ . That is our purpose in the next two sections. We will show in Proposition 5.4 that there is a constant  $a_0 > 0$  such that

$$\text{Leb}(B_n) \leq a_0 \text{Leb}(A_n). \tag{3}$$

We will also show in Proposition 5.1 that there are  $N = N(\varepsilon) \geq 1$  and a constant  $c_0 > 0$  for which

$$\text{Leb}\left(\bigcup_{i=0}^N \{R = n + i\}\right) \geq c_0 \text{Leb}(A_{n-1} \cap H_n). \tag{4}$$

Taking into account that  $\Delta_n = A_n \cup B_n$ , it easily follows from (3) and (4) that there is a constant  $b_0 > 0$  such that

$$\text{Leb}\left(\bigcup_{i=0}^N \{R = n + i\}\right) \geq b_0 \frac{\text{Leb}(A_{n-1} \cap H_n)}{\text{Leb}(A_{n-1})} \text{Leb}(\Delta_{n-1}).$$

This immediately implies that

$$\text{Leb}(\Delta_{n+N}) \leq \left(1 - b_0 \frac{\text{Leb}(A_{n-1} \cap H_n)}{\text{Leb}(A_{n-1})}\right) \text{Leb}(\Delta_{n-1}). \tag{5}$$

It is no restriction to assume that  $R_0 > 2(N + 1)$  and we do it. Take any large  $n$  and let  $k_0 \geq 1$  be the smallest integer for which  $n - 1 - k_0(N + 1) \leq R_0$ . The above assumption on  $R_0$  and  $N$  implies that  $n - (k_0 + 1)(N + 1) \geq 1$ . Now we consider the partition of  $\{n - (k_0 + 1)(N + 1), \dots, n - 1\}$  into the sets

$$\begin{aligned} J_N &= \{n - 1, n - 1 - (N + 1), \dots, n - 1 - k_0(N + 1)\}, \\ &\vdots \\ J_1 &= \{n - N, n - N - (N + 1), \dots, n - N - k_0(N + 1)\}, \\ J_0 &= \{n - (N + 1), n - 2(N + 1), \dots, n - (k_0 + 1)(N + 1)\}. \end{aligned}$$

Applying (5) repeatedly we arrive at the following set of  $N + 1$  inequations:

$$\begin{aligned} \text{Leb}(\Delta_{n+N}) &\leq \prod_{j \in J_N} \left(1 - b_0 \frac{\text{Leb}(A_j \cap H_{j+1})}{\text{Leb}(A_j)}\right) \text{Leb}(\Delta_0), \\ &\vdots \\ \text{Leb}(\Delta_n) &\leq \prod_{j \in J_0} \left(1 - b_0 \frac{\text{Leb}(A_j \cap H_{j+1})}{\text{Leb}(A_j)}\right) \text{Leb}(\Delta_0). \end{aligned}$$

Multiplying the terms in the inequations above and ignoring factors from  $n - (k_0 + 1)(N + 1)$  to  $R_0 - 1$  on the right hand side (observe that those factors are smaller than 1), we obtain

$$\prod_{j=0}^N \text{Leb}(\Delta_{n+j}) \leq \prod_{j=R_0}^{n-1} \left(1 - b_0 \frac{\text{Leb}(A_{j-1} \cap H_j)}{\text{Leb}(A_{j-1})}\right) \text{Leb}(\Delta_0)^{N+1}.$$

Taking into account that  $(\Delta_n)_n$  forms a decreasing sequence of sets we finally have

$$\text{Leb}(\Delta_{n+N}) \leq \exp\left(-\frac{b_0}{N+1} \sum_{j=R_0}^n \frac{\text{Leb}(A_{j-1} \cap H_j)}{\text{Leb}(A_{j-1})}\right) \text{Leb}(\Delta_0). \tag{6}$$

In Section 6 we will prove the Main Theorem by considering several different cases, according to the behavior of the proportions  $\text{Leb}(A_{n-1} \cap H_n)/\text{Leb}(A_{n-1})$ . It is not hard to check that if the average

$$\frac{1}{n} \sum_{j=1}^n \frac{\text{Leb}(A_{j-1} \cap H_j)}{\text{Leb}(A_{j-1})}$$

is bounded away from 0 for large  $n$ , then  $\text{Leb}(\Delta_n)$  decays exponentially fast to 0. This happens, for instance, when  $f$  is uniformly expanding.

### 5. Transitional metric estimates

The goal of this section is to prove several estimates relating the Lebesgue measure of the sets  $A_n, A_n^\varepsilon, B_n$  and  $\{R = n\}$ . The first result shows that a fixed proportion of  $A_{n-1}$  having  $n$  as a hyperbolic time gives rise to new elements of the partition with return time not exceeding  $n$  too much. We discuss the relative proportion of the sets  $A_n$  and  $B_n$  in  $\Delta_n$ .

**Proposition 5.1.** *There exist  $c_0 > 0$  and  $N = N(\varepsilon)$  such that for every  $n \geq 1$*

$$\text{Leb}\left(\bigcup_{i=0}^N \{R = n + i\}\right) \geq c_0 \text{Leb}(A_{n-1} \cap H_n).$$

**Proof.** Take  $r = 5\delta_0 C_0^{N_0}$ , where  $N_0$  and  $C_0$  are given by Lemma 2.6 and Remark 2.7, respectively. Let  $\{z_j\}$  be a maximal set in  $f^n(A_{n-1} \cap H_n)$  with the property that  $B(z_j, r)$  are pairwise disjoint. By maximality we have

$$\bigcup_j B(z_j, 2r) \supset f^n(A_{n-1} \cap H_n).$$

Let  $Z_j$  be the set of points  $x \in H_n \cap A_{n-1}$  such that  $f^n(x) = z_j$ . Observe that  $f^n$  sends the hyperbolic pre-ball  $V_n(x)$  associated to  $x \in Z_j$  diffeomorphically onto a ball of radius  $\delta_1$  around  $z_j$  as in Lemma 2.1. Note that the hyperbolic pre-balls  $V_n(x)$  with  $x \in Z_j$  are pairwise disjoint.

Let us fix  $x \in Z_j$ . Given  $B \subset B(z_j, \delta_1)$ , we will simply denote  $(f^n|_{V_n(x)})^{-1}(B)$  by  $f_x^{-n}(B)$ . Our aim now is to prove that  $f_x^{-n}(B(z_j, r))$  contains some component of  $\{R = n + k_j\}$  with  $0 \leq k_j \leq N_\varepsilon + N_0$ . We start by showing that

$$t_{n+k_j}|f_x^{-n}(B(z_j, \varepsilon)) \text{ is not identically } 0 \text{ for some } 0 \leq k_j \leq N_\varepsilon + N_0. \tag{7}$$

Assume by contradiction that  $t_{n+k_j}|f_x^{-n}(B(z_j, \varepsilon)) = 0$  for all  $0 \leq k_j \leq N_\varepsilon + N_0$ . This implies that  $f_x^{-n}(B(z_j, \varepsilon)) \subset A_{n+k_j}^\varepsilon$  for all  $0 \leq k_j \leq N_\varepsilon + N_0$ . Using Lemma 2.4 we may find a hyperbolic pre-ball  $V_m \subset B(z_j, \varepsilon)$  with  $m \leq N_\varepsilon$ . Now, since  $f^m(V_m)$  is a ball  $B$  of radius  $\delta_1$  it follows from Lemma 2.6 that there is some  $V \subset B$  and  $m' \leq N_0$  with  $f^{m'}(V) = \Delta_0$ . Thus, taking  $k_j = m + m'$  we have that  $0 \leq k_j \leq N_\varepsilon + N_0$  and  $f_x^{-n}(V_m)$  is an element of  $\{R = n + k_j\}$  inside  $f_x^{-n}(B(z_j, \varepsilon))$ . This contradicts the fact that  $t_{n+k_j}|f_x^{-n}(B(z_j, \varepsilon)) = 0$  for all  $0 \leq k_j \leq N_\varepsilon + N_0$ , and so (7) holds.

Let  $k_j$  be the smallest integer  $0 \leq k_j \leq N_\varepsilon + N_0$  for which  $t_{n+k_j}|f^{-n}(B(z_j, \varepsilon))$  is not identically 0.



Since

$$f_x^{-n}(B(z_j, \varepsilon)) \subset A_{n-1}^\varepsilon \subset \{t_{n-1} \leq 1\},$$

there must be some element  $U_{n+k_j}^0(j, x)$  of  $\{R = n + k_j\}$  for which

$$f_x^{-n}(B(z_j, \varepsilon)) \cap U_{n+k_j}^1(j, x) \neq \emptyset.$$

Recall that by definition  $f^{n+k_j}$  sends  $U_{n+k_j}^1(j, x)$  diffeomorphically onto  $\Delta_0^1$ , the ball of radius  $2\delta_0$  around  $p$ . From time  $n$  to  $n + k_j$  we may have some final “bad” period of length at most  $N_0$  where the derivative of  $f$  may contract, however being bounded from below by  $1/C_0$  in each step. Thus, the diameter of  $f^n(U_{n+k_j}^1(j, x))$  is at most  $4\delta_0 C_0^{N_0}$ . Since  $B(z_j, \varepsilon)$  intersects  $f^n(U_{n+k_j}^1(j, x))$  and  $\varepsilon < \delta_0 < \delta_0 C_0^{N_0}$ , we have by definition of  $r$

$$f^{-n}(B(z_j, r)) \supset U_{n+k_j}^0(j, x).$$

Thus we have shown that  $f_x^{-n}(B(z_j, r))$  contains some component of  $\{R = n + k_j\}$  with  $0 \leq k_j \leq N_\varepsilon + N_0$ . Moreover, since  $n$  is a hyperbolic time for  $x_j$ , we have by the distortion control given by Lemma 2.1

$$\frac{\text{Leb}(f_x^{-n}(B(z_j, 2r)))}{\text{Leb}(f_x^{-n}(B(z_j, r)))} \leq D_1 \frac{\text{Leb}(B(z_j, 2r))}{\text{Leb}(B(z_j, r))} \tag{8}$$

and

$$\frac{\text{Leb}(f_x^{-n}(B(z_j, r)))}{\text{Leb}(U_{n+k_j}^0(j, x))} \leq D_1 \frac{\text{Leb}(B(z_j, r))}{\text{Leb}(f^n(U_{n+k_j}^0(j, x)))}. \tag{9}$$

Here we are implicitly assuming that  $2r < \delta_1$ . This can be done just by taking  $\delta_0$  small enough. Note that the estimates on  $N_0$  and  $C_0$  do not change when we diminish  $\delta_0$ .

From time  $n$  to time  $n + k_j$  we have at most  $k_j = m_1 + m_2$  iterates with  $m_1 \leq N_\varepsilon, m_2 \leq N_0$  and  $f^n(U_{n+k_j}^0(j, x))$  containing some point  $w_j \in H_{m_1}$ . By the definition of  $(\sigma, \delta)$ -hyperbolic time we have that  $\text{dist}_\delta(f^i(w_j), \mathcal{S}) \geq \sigma^{bN_\varepsilon}$  for every  $0 \leq i < m_1$ , which by the uniform distortion control implies that there is some constant  $D = D(\varepsilon) > 0$  such that  $|\det(Df^i(x))| \leq D$  for  $0 \leq i < m_1$  and  $x \in f^n(U_{n+k_j}^0(j, x))$ . On the other hand, since the first  $N_0$  preimages of  $\Delta_0$  are uniformly bounded away from  $\mathcal{S}$  we also have some  $D' > 0$  such that  $|\det(Df^i(x))| \leq D'$  for every  $0 \leq i \leq m_2$  and  $x$  belonging to an  $i$  preimage of  $\Delta_0$ . Hence,

$$\text{Leb}(f^n(U_{n+k_j}^0(j, x))) \geq \frac{1}{DD'} \text{Leb}(\Delta_0),$$

which combined with (9) gives

$$\text{Leb}(f_x^{-n}(B(z_j, r))) \leq C \text{Leb}(U_{n+k_j}^0(j, x)),$$

with  $C$  only depending on  $D_1, D, D', \delta_0$  and the dimension of  $M$ . We also deduce from (8) that

$$\text{Leb}(f_x^{-n}(B(z_j, 2r))) \leq C' \text{Leb}(f_x^{-n}(B(z_j, r)))$$

with  $C'$  only depending on  $D_1$  and the dimension of  $M$ . Finally let us compare the Lebesgue measure of the sets  $\bigcup_{i=0}^N \{R = n + i\}$  and  $A_{n-1} \cap H_n$ . We have

$$\text{Leb}(A_{n-1} \cap H_n) \leq \sum_j \sum_{x \in Z_j} \text{Leb}(f_x^{-n}(B(z_j, 2r))) \leq C' \sum_j \sum_{x \in Z_j} \text{Leb}(f_x^{-n}(B(z_j, r))).$$

On the other hand,

$$\sum_j \sum_{x \in Z_j} \text{Leb}(f_x^{-n}(B(z_j, r))) \leq C \sum_j \sum_{x \in Z_j} \text{Leb}(U_{n+k_j}^0(j, x)) \leq C \text{Leb}\left(\bigcup_{i=0}^N \{R = n + i\}\right).$$

We just have to take  $c_0 = (CC')^{-1}$ .  $\square$

Let us prove now a couple of useful lemmas. The first one gives a lower bound for the flow of mass from  $B_{n-1}$  to  $A_n$ , and second one gives a lower bound for the flow of mass from  $A_{n-1}$  to  $B_n$  and  $\{R = n\}$ .

**Lemma 5.2.** *There exists  $a_1 > 0$  such that for every  $n \geq 1$*

$$\text{Leb}(B_{n-1} \cap A_n) \geq a_1 \text{Leb}(B_{n-1}).$$

Moreover,  $a_1$  is bounded away from 0 independently from  $\delta_0$ .

**Proof.** It is enough to see this for each component of  $B_{n-1}$  at a time. Let  $C$  be a component of  $B_{n-1}$  and  $Q$  be its outer ring corresponding to  $t_{n-1} = 1$ . Observe that by Lemma 3.1 we have  $Q = C \cap A_n$ . Moreover, there must be some  $k < n$  and a component  $U_k^0$  of  $\{R = k\}$  such that  $f^k$  maps  $C$  diffeomorphically onto  $\bigcup_{i=k}^\infty I_i$  and  $Q$  onto  $I_k$ , both with uniform bounded distortion (not depending on  $\delta_0$  or  $n$ ). Thus, it is sufficient to compare the Lebesgue measures of  $\bigcup_{i=k}^\infty I_i$  and  $I_k$ . We have

$$\frac{\text{Leb}(I_k)}{\text{Leb}(\bigcup_{i=k}^\infty I_i)} \approx \frac{[\delta_0(1 + \sigma^{(k-1)/2})]^d - [\delta_0(1 + \sigma^{k/2})]^d}{[\delta_0(1 + \sigma^{(k-1)/2})]^d - \delta_0^d} \approx 1 - \sigma^{1/2}.$$

Clearly this proportion does not depend on  $\delta_0$ .  $\square$

The second item of the lemma below is apparently counterintuitive, since our main goal is to make the points in  $\Delta_0$  have small return times. However, this is needed for keeping  $\text{Leb}(A_n)$  uniformly much bigger than  $\text{Leb}(B_n)$ . This will help us in the statistical estimates of the last section.

**Lemma 5.3.** *There exist  $b_1 = b_1(\delta_0) > 0$  and  $c_1 = c_1(\delta_0) > 0$  with  $b_1 + c_1 < 1$  such that for every  $n \geq 1$*

- (1)  $\text{Leb}(A_{n-1} \cap B_n) \leq b_1 \text{Leb}(A_{n-1});$
- (2)  $\text{Leb}(A_{n-1} \cap \{R = n\}) \leq c_1 \text{Leb}(A_{n-1}).$

Moreover,  $b_1 \rightarrow 0$  and  $c_1 \rightarrow 0$  as  $\delta_0 \rightarrow 0$ .

**Proof.** It is enough to prove this for each neighborhood of a component  $U_n^0$  of  $\{R = n\}$ . Observe that by construction we have  $U_n^3 \subset A_{n-1}^\varepsilon$ , which means that  $U_n^2 \subset A_{n-1}$ , because  $\varepsilon < \delta_0 < \sqrt{\delta_0}$ . Using the uniform bounded distortion of  $f^n$  on  $U_n^3$  given by Lemma 2.4 and Lemma 2.6 (cf. Remark 2.7) we obtain

$$\frac{\text{Leb}(U_n^1 \setminus U_n^0)}{\text{Leb}(U_n^2 \setminus U_n^1)} \approx \frac{\text{Leb}(\Delta_0^1 \setminus \Delta_0^0)}{\text{Leb}(\Delta_0^2 \setminus \Delta_0^1)} \approx \frac{\delta_0^d}{\delta_0^{d/2}} \ll 1,$$

which gives the first estimate. Moreover,

$$\frac{\text{Leb}(U_n^0)}{\text{Leb}(U_n^2 \setminus U_n^1)} \approx \frac{\text{Leb}(\Delta_0^0)}{\text{Leb}(\Delta_0^2 \setminus \Delta_0^1)} \approx \frac{\delta_0^d}{\delta_0^{d/2}} \ll 1,$$

and this gives the second one.  $\square$

The next result is a consequence of the estimates we obtained in the last two lemmas. The proof is essentially the same of the uniformly hyperbolic case; see [29]. Here we need to be more careful on the estimates.

**Proposition 5.4.** *There exists  $a_0 = a_0(\delta_0) > 0$  such that for every  $n \geq 1$*

$$\text{Leb}(B_n) \leq a_0 \text{Leb}(A_n).$$

Moreover,  $a_0 \rightarrow 0$  as  $\delta_0 \rightarrow 0$ .

**Proof.** We have by Lemma 5.3

$$\text{Leb}(A_{n-1} \cap A_n) \geq \eta \text{Leb}(A_{n-1}), \tag{10}$$

where  $\eta = 1 - b_1 - c_1$ . We define

$$\hat{a} = \frac{b_1 + c_1}{a_1} \quad \text{and} \quad a_0 = \frac{(1 + a_1)b_1 + c_1}{a_1 \eta}.$$

The fact that  $a_0 \rightarrow 0$  when  $\delta_0 \rightarrow 0$  is a consequence of  $b_1 \rightarrow 0$  and  $c_1 \rightarrow 0$  when  $\delta_0 \rightarrow 0$  and  $a_1$  being bounded away from 0. Observe that  $0 < \eta < 1$  and  $\hat{a} < a_0$ . Now the proof of the proposition follows by induction. The result obviously holds for  $n$  up to  $R_0$ . Assuming that it holds for  $n - 1 \geq R_0$  we will show that it also holds for  $n$ , by considering separately the cases  $\text{Leb}(B_{n-1}) > \hat{a} \text{Leb}(A_{n-1})$  and  $\text{Leb}(B_{n-1}) \leq \hat{a} \text{Leb}(A_{n-1})$ .

Assume first that  $\text{Leb}(B_{n-1}) > \hat{a} \text{Leb}(A_{n-1})$ . We may write

$$\text{Leb}(B_{n-1}) = \text{Leb}(B_{n-1} \cap A_n) + \text{Leb}(B_{n-1} \cap B_n),$$

which by Lemma 5.2 gives

$$\text{Leb}(B_{n-1} \cap B_n) \leq (1 - a_1) \text{Leb}(B_{n-1}). \tag{11}$$

Since we also have

$$\text{Leb}(B_n) = \text{Leb}(B_n \cap B_{n-1}) + \text{Leb}(B_n \cap A_{n-1}),$$

it follows from (11) and Lemma 5.3 that

$$\text{Leb}(B_n) \leq (1 - a_1) \text{Leb}(B_{n-1}) + b_1 \text{Leb}(A_{n-1}),$$

which according to the case we are considering leads to

$$\text{Leb}(B_n) \leq (1 - a_1) \text{Leb}(B_{n-1}) + \frac{b_1 a_1}{b_1 + c_1} \text{Leb}(B_{n-1}) < \text{Leb}(B_{n-1}). \tag{12}$$

On the other hand, we have  $\text{Leb}(A_n) = \text{Leb}(A_n \cap A_{n-1}) + \text{Leb}(A_n \cap B_{n-1})$ , which together with Lemma 5.2 and (10) gives  $\text{Leb}(A_n) \geq \eta \text{Leb}(A_{n-1}) + a_1 \text{Leb}(B_{n-1})$ . Again by the case we are considering we have

$$\text{Leb}(A_n) \geq \eta \text{Leb}(A_{n-1}) + a_1 \hat{a} \text{Leb}(A_{n-1}) \geq \text{Leb}(A_{n-1}). \tag{13}$$

Inequalities (12) and (13), together with the inductive hypothesis, yield the result in this first case.

Assume now that  $\text{Leb}(B_{n-1}) \leq \hat{a} \text{Leb}(A_{n-1})$ . Since we have

$$\text{Leb}(B_n) = \text{Leb}(B_n \cap B_{n-1}) + \text{Leb}(B_n \cap A_{n-1}),$$

it follows from Lemma 5.3 that

$$\text{Leb}(B_n) \leq \text{Leb}(B_{n-1}) + b_1 \text{Leb}(A_{n-1}).$$

Hence

$$\frac{\text{Leb}(B_n)}{\text{Leb}(A_n)} < \frac{\text{Leb}(B_{n-1}) + b_1 \text{Leb}(A_{n-1})}{\eta \text{Leb}(A_{n-1})} \leq \frac{\hat{a} + b_1}{\eta} = a_0,$$

which gives the result also in this case.  $\square$

It will be useful to establish the following consequence of the last two results.

**Corollary 5.5.** *There exists  $c_2 > 0$  such that for every  $n \geq 1$*

$$\text{Leb}(\Delta_n) \leq c_2 \text{Leb}(\Delta_{n+1}).$$

**Proof.** By Lemma 5.3 we have

$$\text{Leb}(\Delta_{n+1}) \geq \text{Leb}(A_{n+1}) \geq (1 - b_1 - c_1) \text{Leb}(A_n).$$

On the other hand, by Proposition 5.4,

$$\text{Leb}(\Delta_n) = \text{Leb}(A_n) + \text{Leb}(B_n) \leq (1 + a_0^{-1}) \text{Leb}(A_n).$$

It is enough to take  $c_2 = (1 + a_0^{-1}) / (1 - b_1 - c_1)$ .  $\square$

### 6. Asymptotic metric estimates

We start this section by recalling that  $\theta > 0$  was obtained in Lemma 2.2 and gives a lower bound for the frequency of hyperbolic times; it only depends on the non-uniform expansion coefficient  $\lambda$  and the map  $f$ .

Before we go into the main proposition of this section which will enable us to conclude the proof of the Main Theorem, let us impose one more requirement on the choice of  $\delta_0$ : let  $\gamma > 0$  be some positive number (to be specified later) and take  $0 < \alpha < (\theta/12)^{\gamma+1}$ . Then we choose  $\delta_0 > 0$  small so that  $a_0 = a(\delta_0) < 2\alpha$ .

We define for each  $n \geq 1$

$$E_n = \left\{ j \leq n: \frac{\text{Leb}(A_{j-1} \cap H_j)}{\text{Leb}(A_{j-1})} < \alpha \right\},$$

and

$$F = \left\{ n \in \mathbb{N}: \frac{\#E_n}{n} > 1 - \frac{\theta}{12} \right\}.$$

**Proposition 6.1.** *Take any  $n \in F$  with  $n \geq R_0 > 12/\theta$ . If  $\text{Leb}(A_n) \geq 2\text{Leb}(\Gamma_n)$ , then there is some  $0 < k = k(n) < n$  for which*

$$\frac{\text{Leb}(A_n)}{\text{Leb}(A_k)} < \left(\frac{k}{n}\right)^\gamma.$$

**Proof.** We have for  $j \leq n$

$$\frac{\text{Leb}(A_n \cap H_j)}{\text{Leb}(A_n)} \geq \frac{\text{Leb}(A_n \setminus \Gamma_n)}{\text{Leb}(A_n)} \cdot \frac{\text{Leb}((A_n \setminus \Gamma_n) \cap H_j)}{\text{Leb}(A_n \setminus \Gamma_n)} \geq \frac{1}{2} \cdot \frac{\text{Leb}((A_n \setminus \Gamma_n) \cap H_j)}{\text{Leb}(A_n \setminus \Gamma_n)},$$

which together with the conclusion of Corollary 2.3 for the set  $A_n \setminus \Gamma_n$  gives

$$\frac{1}{n} \sum_{j=1}^n \frac{\text{Leb}(A_n \cap H_j)}{\text{Leb}(A_n)} \geq \frac{\theta}{2}. \tag{14}$$

Let

$$G_n = \left\{ j \in E_n: \frac{\text{Leb}(A_{j-1})}{\text{Leb}(A_n)} > \frac{\theta}{12\alpha} \right\}.$$

Since  $n \in F$ , we have

$$\frac{1}{n} \sum_{j=1}^n \frac{\text{Leb}(A_n \cap H_j)}{\text{Leb}(A_n)} \leq \frac{\theta}{12} + \frac{1}{n} \sum_{j \in E_n} \frac{\text{Leb}(A_n \cap H_j)}{\text{Leb}(A_n)} \leq \frac{\theta}{12} + \frac{1}{n} \sum_{j \in E_n \setminus G_n} \frac{\text{Leb}(A_n \cap H_j)}{\text{Leb}(A_n)} + \frac{\#G_n}{n}.$$

Now, for  $j \in E_n \setminus G_n$ ,

$$\begin{aligned} \frac{\text{Leb}(A_n \cap H_j)}{\text{Leb}(A_n)} &= \frac{\text{Leb}(A_n \cap H_j)}{\text{Leb}(A_{j-1})} \cdot \frac{\text{Leb}(A_{j-1})}{\text{Leb}(A_n)} \\ &\leq \left( \frac{\text{Leb}(A_n \cap A_{j-1} \cap H_j)}{\text{Leb}(A_{j-1})} + \frac{\text{Leb}((A_n \setminus A_{j-1}) \cap H_j)}{\text{Leb}(A_{j-1})} \right) \frac{\text{Leb}(A_{j-1})}{\text{Leb}(A_n)} \\ &\leq \left( \frac{\text{Leb}(A_{j-1} \cap H_j)}{\text{Leb}(A_{j-1})} + a_0 \right) \frac{\theta}{12\alpha}. \end{aligned}$$

For this last inequality we used the fact that  $(A_n \setminus A_{j-1}) \subset B_{j-1}$  and  $j \notin G_n$ . Hence

$$\begin{aligned} \frac{1}{n} \sum_{j=1}^n \frac{\text{Leb}(A_n \cap H_j)}{\text{Leb}(A_n)} &\leq \frac{\theta}{12} + \frac{1}{n} \sum_{j \in E_n \setminus G_n} \frac{\text{Leb}(A_{j-1} \cap H_j)}{\text{Leb}(A_{j-1})} \frac{\theta}{12\alpha} + a_0 \frac{\theta}{12\alpha} + \frac{\#G_n}{n} \\ &< \frac{\theta}{12} + \alpha \frac{\theta}{12\alpha} + a_0 \frac{\theta}{12\alpha} + \frac{\#G_n}{n}. \end{aligned}$$

By the choice of  $a_0$  we have that the third term in the last sum above is smaller than  $\theta/6$ . So, using (14) we obtain

$$\frac{\#G_n}{n} > \frac{\theta}{6}. \tag{15}$$

Now, defining

$$k = \max(G_n) - 1,$$

we have

$$\text{Leb}(A_n) < \frac{12\alpha}{\theta} \text{Leb}(A_k).$$

It follows from (15) that  $k + 1 > \theta n/6$ , and so  $k/n > \theta/12$ , because  $n \geq R_0 > 12/\theta$ . Since we have chosen  $\alpha < (\theta/12)^{\gamma+1}$ , it follows that

$$\left(\frac{k}{n}\right)^\gamma > \frac{12}{\theta} \left(\frac{\theta}{12}\right)^{\gamma+1} > \frac{12\alpha}{\theta}.$$

This completes the proof of the result.  $\square$

Now we are ready to conclude the proof of our Main Theorem, namely the decay estimate on the tail of return times. Observe that by Proposition 5.4 we have  $\text{Leb}(\Delta_n) \lesssim \text{Leb}(A_n)$ , and so it is enough to derive the tail estimate for  $\text{Leb}(A_n)$  in the place of  $\text{Leb}(\{R > n\}) = \text{Leb}(\Delta_n)$ .

Given any large integer  $n$ , we consider the following situations:

(1) If  $n \in \mathbb{N} \setminus F$ , then by (6) and Corollary 5.5 we have

$$\text{Leb}(\Delta_n) \leq c_2^N \exp\left(-\frac{b_0\theta\alpha}{12(N+1)}(n - R_0)\right) \text{Leb}(\Delta_0).$$

(2) If  $n \in F$ , then we distinguish the following two cases:

(a) If  $\text{Leb}(A_n) < 2\text{Leb}(F_n)$ , then nothing has to be done.

(b) If  $\text{Leb}(A_n) \geq 2\text{Leb}(F_n)$ , then we apply Proposition 6.1 and get some  $k_1 < n$  for which

$$\text{Leb}(A_n) < \left(\frac{k_1}{n}\right)^\gamma \text{Leb}(A_{k_1}).$$

The only situation we are left to consider is (2)(b). In such case, either  $k_1$  is in situation (1) or (2)(a), or by Proposition 6.1 we can find  $k_2 < k_1$  for which

$$\text{Leb}(A_{k_1}) < \left(\frac{k_2}{k_1}\right)^\gamma \text{Leb}(A_{k_2}).$$

Arguing inductively we show that there is a sequence of integers  $0 < k_s < \dots < k_1 < n$  for which one of the following cases eventually holds.

$$\text{Leb}(A_n) < \left(\frac{k_s}{n}\right)^\gamma c_2^N \exp\left(-\frac{b_0\theta\alpha}{12(N+1)}(k_s - R_0)\right) \text{Leb}(\Delta_0). \quad (\text{I})$$

$$\text{Leb}(A_n) < \left(\frac{k_s}{n}\right)^\gamma \text{Leb}(\Gamma_{k_s}). \quad (\text{II})$$

$$\text{Leb}(A_n) < \left(\frac{R_0}{n}\right)^\gamma \text{Leb}(\Delta_0). \quad (\text{III})$$

Case (III) corresponds to falling into situation (2)(b) repeatedly until  $k_s \leq R_0$ . Observe that until now  $\gamma > 0$  is arbitrary.

Now we check that our construction always gives rise to a partition  $\text{Leb } 0$  of  $\Delta_0$ , independently of the way the  $\text{Leb}(\Gamma_n)$  goes to 0. One easily sees that the only case we have to consider is (II). If  $k_s \geq \sqrt{n}$ , then  $\text{Leb}(A_n) < \text{Leb}(\Gamma_{\sqrt{n}})$ . Otherwise,  $\text{Leb}(A_n) \leq \text{Leb}(\Delta_0)/n^{\gamma/2}$ . In both cases  $\text{Leb}(A_n) \rightarrow 0$  when  $n \rightarrow \infty$ .

Let us finally prove item (4) of our Main Theorem. As before, we just have to consider case (II) above. Assume that  $\text{Leb}(\Gamma_n) \leq \mathcal{O}(n^{-\gamma})$  for some  $\gamma > 0$ . In this case there must be some  $C > 0$  such that  $k^\gamma \text{Leb}(\Gamma_k) \leq C$  for all  $k \in \mathbb{N}$ , which applied to  $k_s$  in case (II) leads to  $\text{Leb}(A_n) \leq \mathcal{O}(n^{-\gamma})$ .

## References

- [1] J.F. Alves, SRB measures for non-hyperbolic systems with multidimensional expansion, *Ann. Sci. École Norm. Sup.* 33 (4) (2000) 1–32.
- [2] J.F. Alves, V. Araújo, Random perturbations of nonuniformly expanding maps, *Astérisque* 286 (2003) 25–62.
- [3] J.F. Alves, C. Bonatti, M. Viana, SRB measures for partially hyperbolic systems whose central direction is mostly expanding, *Invent. Math.* 140 (2000) 351–398.
- [4] J.F. Alves, S. Luzzatto, V. Pinheiro, Lyapunov exponent and rates of mixing for one-dimensional maps, *Ergodic Theory Dynam. Systems* 24 (2004) 637–657.
- [5] J.F. Alves, M. Viana, Statistical stability for robust classes of maps with non-uniform expansion, *Ergodic Theory Dynam. Systems* 22 (2002) 1–32.
- [6] R. Adler, B. Weiss, Proc. Entropy, a complete metric invariant for automorphisms of the torus, *Nat. Acad. Sci. USA* 57 (1967) 1573–1576.
- [7] R. Adler, B. Weiss, Similarity of Automorphisms of the Torus, *Mem. Amer. Math. Soc.*, vol. 98, Amer. Math. Soc., Providence, RI, 1970.
- [8] V. Baladi, M. Benedicks, V. Maume-Deschamps, Almost sure rate of mixing for i.i.d unimodal maps, *Ann. Sci. Écol. Norm. Sup.* (4) (2002).
- [9] H. Bruin, S. Luzzatto, S. van Strien, Rates of decay of correlations in one-dimensional dynamics, *Ann. Sci. Écol. Norm. Sup.* 36 (2003) 621–646.
- [10] R. Bowen, D. Ruelle, The ergodic theory of Axiom A flows, *Invent. Math.* 29 (1975) 181–202.
- [11] R. Bowen, Markov partitions for Axiom A diffeomorphisms, *Amer. J. Math.* 92 (1970) 725–747.
- [12] R. Bowen, Equilibrium States and the Ergodic Theory of Axiom A Diffeomorphisms, *Lecture Notes in Math.*, vol. 92, Springer, 1975.
- [13] X. Bressaud, Subshifts on an infinite alphabet, *Ergodic Theory Dynam. Systems* 19 (5) (1999) 1175–1200.
- [14] L.A. Bunimovich, Ya.G. Sinai, N.I. Chernov, Statistical properties of two-dimensional hyperbolic billiards, *Russian Math. Surveys* 46 (4) (1991) 47–106.
- [15] J. Buzzi, Markov extensions for multi-dimensional dynamical systems, *Israel J. Math.* 112 (1999) 357–380.
- [16] J. Buzzi, V. Maume-Deschamps, Decay of correlations for piecewise invertible maps in higher dimensions, *Israel J. Math.* 131 (2002) 203–220.
- [17] J. Buzzi, V. Maume-Deschamps, Decay of correlations on towers with non-Hölder Jacobian and non-exponential return time, Preprint, 2002.

- [18] J. Buzzi, O. Sester, M. Tsujii, Weakly expanding skew-products of quadratic maps, *Ergodic Theory Dynam. Systems* 24 (2004) 385–405.
- [19] M. Holland, Slowly mixing systems and intermittency maps, *Ergodic Theory Dynam. Systems*, in press.
- [20] H. Hu, Decay of correlations for piecewise smooth maps with indifferent fixed points, *Ergodic Theory Dynam. Systems* 24 (2004) 495–524.
- [21] T. Krüger, S. Troubetzkoy, Markov partitions and shadowing for non-uniformly hyperbolic systems with singularities, *Ergodic Theory Dynam. Systems* 12 (3) (1992) 487–508.
- [22] C. Liverani, B. Saussol, S. Vaienti, A probabilistic approach to intermittency, *Ergodic Theory Dynam. Systems* 19 (1999) 671–685.
- [23] V. Maume-Deschamps, Correlation decay for Markov maps on a countable state space, *Ergodic Theory Dynam. Systems* 21 (2001) 165–196.
- [24] M. Pollicott, M. Yuri, Statistical properties of maps with indifferent periodic points, *Commun. Math. Phys.* 217 (2001) 503–520.
- [25] O. Sarig, Thermodynamical formalism for countable Markov shifts, *Ergodic Theory Dynam. Systems* 19 (1999) 1565–1593.
- [26] O. Sarig, Subexponential decay of correlations, *Invent. Math.* 150 (2002) 629–653.
- [27] Y. Sinai, Markov partitions and U-diffeomorphisms, *Funktsional Anal. i Prilozhen.* 32 (2) (1968) 70–80.
- [28] M. Viana, Multidimensional non-hyperbolic attractors, *Publ. Math. IHES* 85 (1997) 63–96.
- [29] L.-S. Young, Statistical properties of dynamical systems with some hyperbolicity, *Ann. Math.* 147 (1998) 585–650.
- [30] L.-S. Young, Recurrence times and rates of mixing, *Israel J. Math.* 110 (1999) 153–188.
- [31] M. Yuri, On the speed of convergence to equilibrium states for multi-dimensional maps with indifferent fixed points, *Nonlinearity* 15 (2002) 429–445.