

# Extreme value theory for non-uniformly expanding dynamical systems \*

Mark Holland, Matthew Nicol, Andrei Török

June 23rd, 2008. (Revised November 30th 2009.)

## Abstract

We establish extreme value statistics for functions with multiple maxima and some degree of regularity on certain non-uniformly expanding dynamical systems. We also establish extreme value statistics for time-series of observations on discrete and continuous suspensions of certain non-uniformly expanding dynamical systems via a general lifting theorem. The main result is that a broad class of observations on these systems exhibit the same extreme value statistics as i.i.d processes with the same distribution function.

## Contents

<b>1</b>	<b>Introduction and background</b>	<b>2</b>
<b>2</b>	<b>Statement of the main results</b>	<b>3</b>
2.1	Extreme value theory for observations with multiple maxima . . . . .	4
2.2	Extreme value theory for suspension flows . . . . .	7
2.3	Corollaries of the main results. . . . .	9
<b>3</b>	<b>Applications</b>	<b>11</b>
3.1	Non-uniformly hyperbolic flows . . . . .	12
3.2	Non-uniformly expanding maps with Young towers . . . . .	12
3.3	Gibbs-Markov maps . . . . .	13
3.4	Non-uniformly expanding maps and intermittent maps . . . . .	15

---

\*The research of MN and AT was supported in part by the National Science Foundation grants DMS-0607345 and DMS-0600927. We thank Henk Bruin for useful discussions, especially in connection with Lemma 3.10. We also wish to thank an anonymous referee for helpful suggestions and in particular the proof of Lemma 4.16.

<b>4</b>	<b>Proof of the main theorems</b>	<b>19</b>
4.1	Proof of Theorem 2.2 . . . . .	19
4.1.1	The maximal function technique of Collet . . . . .	20
4.1.2	Recurrent points . . . . .	20
4.1.3	Moving points . . . . .	22
4.1.4	Recurrence for multiple maxima . . . . .	23
4.1.5	Conclusion of proof of Theorem 2.2 . . . . .	24
4.2	Proof of Theorem 2.6 . . . . .	29
4.2.1	Mixing convergence . . . . .	29
4.2.2	Lifting to $X^h$ . . . . .	31
4.2.3	The normalization constants . . . . .	31
4.2.4	The lap number . . . . .	32
4.2.5	Conclusion of proof of Theorem 2.6. . . . .	33

# 1 Introduction and background

If  $\{X_i\}$  is a stochastic process we may define the process of successive maxima  $\{M_n\}$  by  $M_n := \max\{X_1, \dots, X_n\}$ . Analogously, if  $\{X_t\}$  is a continuous time stochastic process we define  $M_T := \sup_{0 \leq t \leq T} \{X_t\}$ . In order to simplify the discussion we will focus in this section on the discrete-time case.

Extreme value theory is concerned with the limiting distribution of  $\{M_n\}$  under linear scalings  $a_n(M_n - b_n)$  defined by constants  $a_n > 0, b_n \in \mathbb{R}$ . The theory is well-understood in the case that  $\{X_i\}$  are i.i.d [18, 11, 25] and certain progress has been made for stationary dependent random variables under mixing conditions [19, 18] and even for non-stationary stochastic processes [18]. In the i.i.d case it is known that there are only three possible non-degenerate limit distributions under linear scaling i.e. if  $\{X_i\}$  is i.i.d,  $a_n > 0, b_n \in \mathbb{R}$  are scaling constants and  $G(x)$  is a non-degenerate distribution defined by

$$\lim_{n \rightarrow \infty} P(a_n(M_n - b_n) \leq x) = G(x)$$

then  $G(x)$  has one of three possible forms (up to scale and location changes), which we call *extreme type distributions*:

**Type I**

$$G(x) = e^{-e^{-x}}, \quad -\infty < x < \infty.$$

**Type II**

$$G(x) = \begin{cases} 0 & \text{if } x \leq 0; \\ e^{-x^{-\alpha}} & \text{for some } \alpha > 0 \text{ if } x > 0. \end{cases}$$

### Type III

$$G(x) = \begin{cases} e^{-(-x)^\alpha} & \text{for some } \alpha > 0 \text{ if } x \leq 0; \\ 1 & \text{if } x > 0. \end{cases}$$

The extremal type distributions are defined only up to scale and location: if  $G(x)$  is Type I then so is  $G(ax + b)$  for any constants  $a > 0, b$ , and similarly for Type II and Type III.

We make the elementary remark:

**Lemma 1.1** *Assume a function  $x \mapsto g(x)$  has a minimum value of zero (we have in mind the function  $g(x) = d(x, x_0)$ ).*

*The following are equivalent, where  $\alpha > 0$ :*

1. *A Type I law for  $x \mapsto -\log g(x)$  with  $a_n = 1$  and  $b_n = \log n$ ;*
2. *A Type II law for  $x \mapsto g(x)^{-\alpha}$  with  $a_n = n^{-\alpha}$  and  $b_n = 0$ ;*
3. *A Type III law for  $x \mapsto C - g(x)^\alpha$  with  $a_n = n^\alpha$  and  $b_n = C$ ;*

*(and similarly for other choices of  $b_n$  in the first case).*

Recent work on extreme value theory in the setting of deterministic dynamics includes [2, 3, 4, 6, 8, 9, 10, 12, 13]. These works establish the extreme value laws for observables with a unique maximum in a variety of non-uniformly hyperbolic [2, 3, 8, 9, 10], partially hyperbolic [6, 12] and uniformly hyperbolic with singularities [13] settings. There is also a closely related literature on return time statistics [3, 16, 14, 15, 17, 22].

## 2 Statement of the main results

In this section we state two main theorems. The first is a result on the existence of extreme value distributions for a wide class of non-uniformly expanding dynamical systems where the observations on the system have multiple maxima. The second result is a theorem which gives conditions under which extreme value distributions lift from a base transformation  $f : X \rightarrow X$  to a discrete- or continuous-time suspension. Their proofs are given in Section 4. We also list in this section applications of these theorems, with details provided in Section 3.

## 2.1 Extreme value theory for observations with multiple maxima

We prove that under certain conditions on a dynamical system, for a family of observations with multiple maxima the extreme value distribution obtained is the same as that of a corresponding i.i.d. stochastic process. The conditions we impose on the  $x_k$ 's (the location of the maxima) are that they are density points and do not recur too fast, conditions which are satisfied for Lebesgue a.e.  $r$ -tuple  $(x_1, \dots, x_r)$  for certain classes of non-uniformly expanding maps. As far as we know these are the first extreme value statistic results for observations with multiple maxima. Our result is based on the maximal function argument of Collet [4] and the estimates developed there.

Suppose  $f : X \rightarrow X$  has an ergodic SRB measure  $\mu$  with density  $\rho$  with respect to Lebesgue measure (or volume)  $m$ . We refer to  $x_0$  as a *density point* of  $\mu$  if

$$\lim_{r \rightarrow 0^+} \frac{\mu(B_r(x_0))}{m(B_r(x_0))} = \rho(x_0) < \infty.$$

Our standing assumption is that if a function is maximized at a point then that point is a density point.

We use the notation  $\varphi(x) \approx \psi(x)$  near  $x_0$  to mean that

$$\varphi(x) = \psi(x) + e(x)$$

and  $\lim_{x \rightarrow x_0} \frac{e(x)}{\psi(x)} = 0$ . We have in mind that the function  $\psi$  has a unique local maximum at  $x_0$ , and  $e(x)$  is a negligible perturbation. For ease of exposition we introduce the following definitions.

**Definition 2.1** Consider a function  $\psi : X \rightarrow \mathbb{R}$  and points  $x_k \in X$ ,  $1 \leq k \leq r$ . Let  $d$  be a distance function on  $X$ . We now describe the different types of maxima at the points  $x_k$  for  $\psi$  that we consider. We assume that  $\psi$  is a.e. uniformly bounded on the complement of any open neighborhoods of the points  $x_k$ ,  $1 \leq k \leq r$ . We say that:

1.  $\psi$  has a *logarithmic singularity* at the point  $x_k$  if  $x_k$  has a neighbourhood where  $\psi(x) = -C_k \log d(x, x_k) + g_k(x)$  with  $C_k > 0$ , where  $g_k$  is bounded and has a finite limit as  $x \rightarrow x_k$ ;
2.  $\psi$  has a *power singularity* at the point  $x_k$  if  $\psi(x) \approx C_k d(x, x_k)^{s_k}$  near  $x_k$  with  $s_k < 0$ ,  $C_k > 0$ ;
3.  $\psi$  has a *power function maximum* at the points  $x_k$ ,  $1 \leq k \leq r$ , if  $\psi(x) \approx C - C_k d(x, x_k)^{s_k}$  near  $x_k$  with  $s_k > 0$ ,  $C_k > 0$ , and supremum of  $\psi$  on the complement of any open neighborhoods of the points  $x_k$  is less than  $C$ .

For bounded maxima or *power singularities* of the form  $d(x, x_k)^{s_k}$ ,  $s_k < 0$  the relation  $\approx$  suffices. But for technical reasons related to the form of Type I distributions we require that  $\psi(x) + C_k \log d(x, x_k)$  be bounded for logarithmic singularities. In Theorem 2.2 we consider the local behavior of a function  $\psi(x)$  near its maximal points  $x_k$  and suppose that at each  $x_k$  it is one of the types of maxima listed in Definition 2.1 above.

**Theorem 2.2** *Let  $X \subset \mathbb{R}^D$  be a compact set and let  $d$  denote Euclidean distance. Assume  $f : X \rightarrow X$  is an ergodic map with an absolutely continuous invariant probability measure  $\mu$  having density  $\rho \in L^{1+\delta}(m)$ ,  $\delta > 0$ . Suppose that for all Lipschitz  $\varphi_1, \varphi_2 \in L^\infty$ :*

$$\left| \int \varphi_1 \cdot \varphi_2 \circ f^n d\mu - \int \varphi_1 d\mu \int \varphi_2 d\mu \right| \leq \Theta_1(n) \|\varphi_2\|_{L^\infty} \|\varphi_1\|_{\text{Lip}},$$

Assume that:

- (a) *There exist functions  $\tilde{g}(n)$  and  $g(n)$  (increasing to  $\infty$  as  $n \rightarrow \infty$ ) such that for all  $v > 0$  there exists  $N(v) > 0$  with*

$$g(n) \leq \tilde{g}\left(\frac{n}{3v}\right) \text{ for } n > N(v) \quad (2.1)$$

and for some  $0 < \beta < 1$  and  $D\alpha_1 > \beta > 0$

$$\lim_{n \rightarrow \infty} \frac{g(n)}{n^\beta} = 0 \quad (2.2)$$

$$\tilde{g}(n) \leq Cn^{D(1-\alpha_1)}. \quad (2.3)$$

- (b) *For some  $\varepsilon > 0$ ,*

$$\Theta_1(g(n^{1/D})) \leq Cn^{-\zeta-\varepsilon} \text{ with } \zeta = (1+\beta) \left(1 + \frac{1+\delta}{\delta}\right) + \frac{1}{D} - 1. \quad (2.4)$$

- (c) *Defining*

$$E_n := \{x \in X : d(x, f^j x) \leq \frac{2}{n} \text{ for some } j \leq \tilde{g}(n)\},$$

there is  $\alpha > \beta$  such that

$$\mu(E_n) \leq \frac{C}{n^\alpha} \text{ for all } n. \quad (2.5)$$

Then, for each  $r \geq 1$  there is a set  $\mathcal{X}_r \subset X^r$  of full Lebesgue measure with the following property: if  $(x_1, \dots, x_r)$  in  $\mathcal{X}_r$  then, with the notations of Definition 2.1:

(I) if  $\psi : X \rightarrow \mathbb{R}$  has only logarithmic singularities, at the points  $x_1, \dots, x_r$ , then  $\psi$  satisfies a Type I extreme value law. More precisely, with  $m(B_1(0))$  denoting the volume of the unit ball,

$$\mu(\max\{\psi, \psi \circ f, \dots, \psi \circ f^n\} \leq v + \frac{C}{D} \log n) \rightarrow \exp \left( -m(B_1(0)) e^{-Dv/C} \sum_{\{k|C_k=C\}} \rho(x_k) e^{D\ell_k} \right)$$

where  $C = \max_k C_k$  and  $\ell_k = \lim_{x \rightarrow x_k} [\psi(x) + C_k \log d(x, x_k)]$ .

(II) if  $\psi$  is unbounded with power singularities (and perhaps logarithmic singularities also) then  $\psi$  satisfies a Type II extreme value law. Define  $s := \max_{s_k < 0} \{|s_k|\}$ . Then, for  $v > 0$ ,

$$\mu(\max\{\psi, \psi \circ f, \dots, \psi \circ f^n\} \leq n^{\frac{s}{D}} v) \rightarrow \exp \left( -m(B_1(0)) \sum_{\{k|-s_k=s\}} \left( \frac{C_k}{v} \right)^{D/s} \rho(x_k) \right)$$

(III) if  $\psi$  is bounded (there are only power function maxima i.e. all  $s_k \geq 0$ ) then  $\psi$  satisfies a Type III extreme value law. Define  $s := \max\{s_k\}$ . Then, for  $v \leq 0$ ,

$$\mu(\max\{\psi, \psi \circ f, \dots, \psi \circ f^n\} \leq \frac{v}{n^{\frac{s}{D}}} + C) \rightarrow \exp \left( -m(B_1(0)) \sum_{\{k|s_k=s\}} \left( \frac{-v}{C_k} \right)^{D/s} \rho(x_k) \right)$$

In Section 3 we show that Theorem 2.2 applies to various classes of dynamical systems. We summarize this in the following corollary. The precise description of the maps is given before each of the theorems listed below. We note that for observations with a unique maximum our results are not new for some of the dynamical systems classes listed below, see [4, 10].

**Corollary 2.3** *The hypotheses of Theorem 2.2 are satisfied by:*

- (a) *D-dimensional non-uniformly expanding maps which are modeled by Young Towers with exponential tails (see Theorem 3.1);*
- (b) *Gibbs-Markov maps in dimension one (see Theorem 3.3);*
- (c) *certain one-dimensional non-uniformly expanding maps with singularities (see Theorem 3.8).*

(d) *intermittent type interval maps with absolutely continuous invariant measures (such as the Pomeau-Manneville type and the Liverani-Saussol Vaienti map) of form  $x \mapsto x + ax^{1+\omega}$  near the indifferent fixed point, for small values of  $\omega > 0$  (see Theorem 3.8 and Remarks 3.9, 3.11).*

Therefore, extreme value laws hold for observations with multiple maxima over these systems, as described in Theorem 2.2.

**Remark 2.4** Because the measure  $\mu$  in Theorem 2.2 is absolutely continuous, it follows that  $\mu(B_{1/n}(y_0)) = \mathcal{O}(1/n^D)$  for Lebesgue-a.e.  $y_0$ . Hence (2.3) implies:

$$\text{for } m\text{-a.e. } y_0 \text{ we have } \tilde{g}(n)\mu(B_{1/n}(y_0)) \leq \frac{C_{y_0}}{n^{D\alpha_1}}. \quad (2.6)$$

This observation is used in the proof of Theorem 2.2 (see Lemma 4.3).

**Remark 2.5** The absolute continuity of  $\mu$  with respect to Lebesgue measure has been used in an essential way to establish our results on extreme value laws for dynamical systems, such as the results given in Corollary 2.3. Our techniques do not easily extend to general non-uniformly hyperbolic systems that have singular SRB measures (singular with respect to volume but absolutely continuous when restricted to unstable manifolds). In higher dimensional systems with singular SRB measures the regularity of the SRB measure  $\mu$  also becomes an issue, especially when estimating the measure of balls on asymptotically small scales. Such regularity is not apriori guaranteed when studying such systems, even if one assumes they are modeled by a Young tower.

## 2.2 Extreme value theory for suspension flows

The results we present here apply to both discrete suspensions (maps) or continuous suspensions (flows). We will use the same language for both. As far as we know these are the first results on extreme value statistics for non-uniformly hyperbolic flows. Statistical limit laws on suspension flows are derived in [21] for normalized ergodic sums. Here as in [21], a result of Eagleson [7] is used to conclude that the convergence in distribution is mixing [23]. The main part of the argument is to relate a random index (the *lap number*) to a deterministic one. The methods needed here differ from those of [21]. Combined with distributional mixing this allows us to lift the convergence in distribution from an observation on the base of the suspension to its pull-pack to the suspension. Our results are also an extension of [18, Theorem 13.3.2] and related results in [18, Chapter 13] where a discrete sampling at constant times is considered.

Assume that  $f : X \rightarrow X$  preserves the probability measure  $\mu$ . We suppose that  $h \in L^1(\mu)$  is a positive roof function. Consider the suspension space

$$X^h = \{(x, u) \in X \times \mathbb{R} \mid 0 \leq u \leq h(x)\} / \sim, \quad (x, h(x)) \sim (f(x), 0)$$

( $u \in \mathbb{Z}$  and  $h : X \rightarrow \mathbb{Z}_+$  for the discrete case). We denote the suspension (semi) flow by  $f_s : X^h \rightarrow X^h$ ,  $f_s(x, u) = (x, u + s) / \sim$ .

We introduce the projection map  $\pi^h : X^h \rightarrow X$  by  $\pi^h(x, u) = x$  for  $(x, u) \in X^h$ ,  $0 \leq u < h(x)$ . If  $\psi : X \rightarrow \mathbb{R}$ , we will often use the notation  $\widehat{\psi}$  for  $\psi \circ \pi^h : X^h \rightarrow \mathbb{R}$ .

On  $X^h$  introduce the flow-invariant probability measure  $\mu^h$  given by  $d\mu \times dm / \bar{h}$  (where for the discrete case  $dm$  will denote counting measure) and  $\bar{h} = \int_X h d\mu$ . We will use  $du$  to mean either the Lebesgue or the counting measure.

Consider a (measurable) observation  $\varphi : X^h \rightarrow \mathbb{R}$  and define  $\Phi : X \rightarrow \mathbb{R}$  by

$$\Phi(x) := \max\{\varphi(f_s(x)) \mid 0 \leq s < h(x)\}. \quad (2.7)$$

Denote

$$\begin{aligned} \varphi_t(p) &:= \max\{\varphi(f_s(p)) \mid 0 \leq s < t\} \\ \Phi_N(x) &:= \max\{\Phi(f^k(x)) \mid 0 \leq k < N\}. \end{aligned} \quad (2.8)$$

Our main theorem in the flow case is:

**Theorem 2.6 (discrete or continuous suspensions)** *Assume that  $f : (X, \mu) \rightarrow (X, \mu)$  is ergodic and  $h \in L^1(\mu)$ . Suppose also that the normalizing constants  $a_n > 0$  and  $b_n$  satisfy:*

$$\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} a_n |b_{[n+\varepsilon n]} - b_n| = 0, \quad (2.9)$$

$$\lim_{\varepsilon \rightarrow 0} \limsup_{n \rightarrow \infty} \left| 1 - \frac{a_{[n+\varepsilon n]}}{a_n} \right| = 0. \quad (2.10)$$

Then, with the notations described above in (2.7) and (2.8):

$$a_N(\Phi_N - b_N) \rightarrow_d G \implies a_{[T/\bar{h}]}(\varphi_T - b_{[T/\bar{h}]}) \rightarrow_d G. \quad (2.11)$$

If in addition  $h^{-1} \in L^1(\mu)$  then

$$a_N(\Phi_N - b_N) \rightarrow_d G \iff a_{[T/\bar{h}]}(\varphi_T - b_{[T/\bar{h}]}) \rightarrow_d G. \quad (2.12)$$

**Remark 2.7** Whether (2.9) and (2.10) are satisfied depends on the precise choice of observation. For example, one can apply Theorem 2.6 in the Type I case where  $b_n = \log n$  and  $a_n = 1$ , in the Type II and Type III scenarios when  $a_n$  is regularly varying and  $b_n$  is constant. Thus the assumptions on  $a_n, b_n$  hold for logarithmic singularities, power singularities and power function maxima.

**Remark 2.8** In the flow case (unlike the discrete setting) the constants  $a_n, b_n$  are not determined by the requirement  $\mu^h(\varphi > \frac{v}{a_n} + b_n) = O(\frac{1}{n})$ .

**Remark 2.9** We are mainly interested in lifting extreme value laws given by Theorem 2.2. In the setting  $a_n = 1$  and  $b_n = \log n$  the proof of Theorem 2.6 becomes much simpler.

### 2.3 Corollaries of the main results.

We show here that by combining Theorems 2.2 and 2.6 the extreme value theory for observations with multiple maxima on a broad class of non-uniformly hyperbolic flows is obtained. For simplicity of exposition we look at observations that depend on distance only.

We consider two situations. In the first we define a (local) metric on the suspension  $X^h$  starting from a metric  $d_X$  on  $X$ . In the second we start with a manifold  $M$  endowed with a metric  $d_M$  and view it as a suspension.

**Suspension flows.** Consider a suspension flow on  $X^h$ . Let  $d_X$  be a metric on  $X$  and define a (local) metric  $d_{X^h}$  on  $X^h$  by

$$d_{X^h}((x, u), (y, v)) = \sqrt{d_X(x, y)^2 + |u - v|^2}. \quad (2.13)$$

**Flows on manifolds.** Let  $M$  be a compact Riemannian manifold, and  $f_t : M \rightarrow M$  a  $C^1$ -flow. Assume that  $X \subset M$  is a transverse cross-section of the flow which is a  $C^1$ -submanifold with boundary (not necessarily connected). Let  $h : X \rightarrow \mathbb{R}$  be the first return time of the flow to  $X$  and assume that  $h$  is essentially bounded. We model the flow  $f_t : M \rightarrow M$  in the standard way by the suspension flow  $f_t : X^h \rightarrow X^h$  (with abuse of notation). Recall that the Riemannian metric on  $M$  is denoted by  $d_M$  and denote the metric induced on  $X$  by  $d_X$ . There is a projection  $\pi_M : X^h \rightarrow M$ ,  $(x, t) \mapsto f_t(x)$ , which is a local  $C^1$  diffeomorphism. Let  $\mu$  be an invariant probability measure for the first return map  $f : X \rightarrow X$ . This induces (in the standard way) an invariant measure  $\mu^h$  on the suspension  $X^h$ , which then determines a  $f_t$ -invariant measure  $\mu_M$  on  $M$  by  $\mu_M(A) = \mu^h(\pi_M^{-1}A)$  for measurable sets  $A$ .

For these situations we have the following consequence of Theorems 2.2 and 2.6:

**Theorem 2.10** *Assume that the measure  $\mu$  on  $X$  is non-atomic and ergodic for  $f$ , and the roof function  $h : X \rightarrow \mathbb{R}$  is in  $L^1(\mu)$  and lower semi-continuous  $\mu$ -a.e.*

*If statement (I), (II) or (III) of Theorem 2.2 holds for  $(f, \mu)$  and observations  $\Phi : X \rightarrow \mathbb{R}$  having  $r$  maxima of logarithmic singularity, power singularity or power function type with respect to  $d_X$  at  $\mu$ -a.e.  $r$ -tuple in  $X^r$ , then an extreme value law of the same type as in Theorem 2.2 with the same scaling constants holds:*

- (a) over the flow  $(\{f_t\}, \mu^h)$  for observations  $\varphi : X^h \rightarrow \mathbb{R}$  that have maxima of the same type with respect to  $d_{X^h}$  at  $\mu^h$ -a.e.  $r$ -tuple in  $(X^h)^r$ .
- (b) over the flow  $(\{f_t\}, \mu_M)$  for observations  $\varphi : M \rightarrow \mathbb{R}$  that have maxima of the same type with respect to  $d_M$  at  $\mu_M$ -a.e.  $r$ -tuple in  $(M)^r$ .

In particular, these results hold for suspensions with lower semi-continuous roof function over transformations for which Theorem 2.2 applies, for example those listed in Corollary 2.3.

Note that if a point  $p_k = (x_k, u_k)$  is in the  $r$ -tuple so is  $(x_k, u)$  for all  $u < h(x_k)$  so whole fibers are included.

**Proof:** Apply Lemmas 2.11 and 2.12 below, which relate the singularities of  $\Phi$  to those of  $\varphi$ . The conclusion then follows from Theorem 2.6. ■

For the local metric introduced on a suspension note that:

**Lemma 2.11** Consider on a suspension  $X^h$  the local metric defined by (2.13). Let  $\varphi : X^h \rightarrow \mathbb{R}$  be an observation that has logarithmic singularities, power singularities or power function maxima with respect to  $d_{X^h}$  at the points  $p_k = (x_k, u_k) \in X^h$ , where  $0 < u_k < h(x_k)$  and the  $x_k$ 's are distinct,  $1 \leq k \leq r$ .

Then if  $h$  is lower semi-continuous at the points  $x_k$ ,  $\Phi$  defined by (2.7) has the same type of maxima (with the same exponents  $s_k$  for power singularities and exponents  $s_k$  and constant  $C$  for power function maxima) with respect to the metric  $d_X$  at the corresponding points  $x_k$ ,  $1 \leq k \leq r$ .

**Proof:** We prove the case of a logarithmic singularity in detail. The proofs for power singularities and power function maxima are similar and straightforward.

By hypothesis, there is a neighborhood  $U_k \subset \{(x, u) \mid 0 < u < h(x)\} \subset X^h$  of  $p_k = (x_k, u_k)$  on which  $\varphi(p) = -\log d_{X^h}(p, p_k) + g_k(p)$  with  $g_k$  bounded and with  $\lim_{p \rightarrow p_k} g(p) = \ell_k$ ; we take  $C_k = 1$  to simplify the notation. We restrict these neighborhoods for the different  $p_k$ 's so that their projections  $V_k$  to  $X$  are distinct. By the semi-continuity of  $h$ , we can assume that  $(x, u_k) \in U_k$  if  $x \in V_k$ .

Since the observation  $\varphi$  is uniformly bounded outside the union of the  $U_k$ 's, its supremum on each fiber,  $\Phi(x) := \sup\{\varphi(x, u) : 0 \leq u < h(x)\}$ , is uniformly bounded outside the union of the  $V_k$ 's. If  $x$  is close to  $x_k$ , then the supremum is attained for  $(x, u) \in U_k$ .

Therefore, for  $x \neq x_k$ ,  $x \in V_k$  close enough to  $x_k$ ,

$$\Phi(x) = -\log d_X(x, x_k) + \sup_{(x, u) \in U_k} \left\{ -\log \frac{\sqrt{d_X(x, x_k)^2 + |u - u_k|^2}}{d_X(x, x_k)} + g_k(x, u) \right\}$$

Denote, for  $x \neq x_k$ ,

$$\gamma_k(x) := \sup_{(x,u) \in U_k} \left\{ -\log \frac{\sqrt{d_X(x, x_k)^2 + |u - u_k|^2}}{d_X(x, x_k)} + g_k(x, u) \right\}. \quad (2.14)$$

Then  $\gamma_k$  is bounded near  $x_k$  because

$$\gamma_k(x) \leq \sup_{(x,u) \in U_k} \left( -\log \frac{\sqrt{d_X(x, x_k)^2 + |u - u_k|^2}}{d_X(x, x_k)} \right) + \sup_{(x,u) \in U_k} g_k(x, u) \leq \sup_{(x,u) \in U_k} g_k(x, u)$$

where the last inequality follows by taking  $u = u_k$ . To check that  $\lim_{x \rightarrow x_k} \gamma_k(x) = \ell_k$  note that for  $\varepsilon$  small, if  $d_X(x, x_k) < \varepsilon$  and  $|u - u_k| > \varepsilon^{1/2}$  then the logarithmic term in (2.14) is at most  $(\log \varepsilon)/2$ ; hence the supremum is attained for  $|u - u_k| \leq \varepsilon^{1/2}$ , and the conclusion follows.

In the case of power singularities we note if  $s_k < 0$  then  $\sup_{(x,u) \in U_k} (d_X(x, x_k)^2 + |u - u_k|^2)^{s_k/2} = d_X(x, x_k)^{s_k}$  and similarly the exponent  $s_k > 0$  is preserved for power function maxima. The constant  $C$  is also preserved in the case of power function maxima as is the global maximum. ■

Next we prove the corresponding result for a  $C^1$ -flow on a manifold.

**Lemma 2.12** *Let the  $C^1$ -flow  $f_t : M \rightarrow M$  be modeled by the suspension  $X^h$  with  $X$  a  $C^1$  cross-section and  $h$  essentially bounded.*

*Assume that  $\varphi : M \rightarrow \mathbb{R}$  has logarithmic singularities, power singularities or power function maxima with respect to  $d_M$  at the points  $p_k \in M \setminus X$ ,  $1 \leq k \leq r$  with  $p_k \cong (x_k, u_k) \in X^h$ ,  $x_k$  distinct and not on the boundary  $\partial X$  of  $X$ . Then  $\Phi$  defined by (2.7) has the same type of maxima (with the same exponents  $s_k$  for power singularities and same exponents  $s_k$  and constant  $C$  for power function maxima) with respect to  $d_X$  at the corresponding points  $x_k$ ,  $1 \leq k \leq r$ .*

**Proof:** The only difference to be noted from the proof of Lemma 2.11 is that the  $C^1$  diffeomorphism  $\pi_M^{-1} : M \rightarrow X^h$  is not an isometry but in a neighborhood of diameter  $1/n$  about  $p_0$  is given by an invertible linear transformation as  $n \rightarrow \infty$ . This implies that logarithmic singularities are preserved and the exponents  $s_k$  are preserved in the case of power singularities and power function maxima. The constant  $C$  is preserved in the case of power function maxima as is the global maximum. ■

### 3 Applications

In this section we apply our results to specific dynamical systems. We consider applications of our result on multiple maxima, namely Theorem 2.2. We also explain how these yield, by Theorem 2.6, similar results for suspension flows over these maps.

In a given application we adopt a strategy of checking conditions (a), (b) and (c) of Theorem 2.2. For the applications we consider one can take  $g(n) = C(\log n)^\kappa$  or  $g(n) = Cn^\kappa$ ,  $\kappa > 0$  and then define  $\tilde{g}(n) := g(n)^{1+\varepsilon}$  for some small  $\varepsilon > 0$ . Thus, in condition (a) of Theorem 2.2, we can take  $D(1-\alpha_1) = \beta$  and only have to check (2.2).

Therefore, from the hypotheses of Theorem 2.2 we only have to check that the system has an absolutely continuous probability measure with density in  $L^{1+\delta}(m)$ , condition (2.2) from (a), the mixing condition (b), and the estimate (c) of  $\mu(E_n)$ . The interplay of the constants is non-trivial, but if the system has density in  $L^p(m)$ ,  $p > 1$ , fast enough decay of correlations (which includes fast polynomial decay), and power law recurrence statistics in the sense of (2.5) then the conclusions of Theorem 2.2 hold. In applications, much work has to be done to check (2.5) since this condition depends on short range dependence statistics and topology of the dynamics in the vicinity of the observable maxima. Analysis of long range dependence statistics can be controlled by the asymptotics of the correlation function.

### 3.1 Non-uniformly hyperbolic flows

Recall that Theorem 2.6 together with Theorem 2.10 gives conditions under which extreme value laws can be lifted to a suspension. In particular, extreme value laws for observations that are functions of distance can be lifted to suspensions over maps that satisfy the assumptions of Theorem 2.2. Thus Theorem 2.10 gives the extreme value theory for observations with multiple maxima for suspension flows with base maps the classes of maps discussed in this section. Hence we obtain extreme value laws for observations with multiple maxima on such non-uniformly hyperbolic flows.

### 3.2 Non-uniformly expanding maps with Young towers

We refer to Young's papers [27, 28] and Baladi's book [1] for details of the construction and properties of Young Towers. In this section we describe the structure that we use. We use the notation of [27]. The existence of a Young tower for a map  $f : X \rightarrow X$  gives an SBR measure  $\mu$  and allows the study of statistical properties of  $f$  with respect to  $\mu$ . Below, we give hypotheses for  $(f, X)$  which allow the system to be modeled by such a tower.

Let  $X \subset \mathbb{R}^D$  and suppose  $f : X \rightarrow X$  is locally  $C^{1+\gamma}$  and  $\|f'(x)\| \leq K$  for all  $x \in X$ . Suppose  $f : X \rightarrow X$  is modeled by a Young tower. Let  $m$  denote Lebesgue measure (volume). Suppose  $\Lambda_0$  is the base of the tower and  $R : \Lambda_0 \rightarrow \mathbb{N}$  is the return time function. Then we have the following result:

**Theorem 3.1 (Based on Theorem 1.1 [4] and Theorem 6, [10])** *Suppose that  $m\{R > n\} = \mathcal{O}(\theta^n)$  for some  $\theta < 1$ . Then the hypotheses of Theorem 2.2 are*

satisfied.

**Proof:** We recall key facts from [4, 10] that allow us to verify (a), (b) and (c) in Theorem 2.2. It is shown using exponential tails of  $m\{R > n\}$  that the density  $\rho$  must lie in  $L^{1+\delta}(m)$  for some  $\delta > 0$  and  $m(E_n) \leq n^{-\alpha}$  for some  $\alpha > 0$ . The mixing is also exponential,  $\Theta_1(n) \leq C\tilde{\theta}^n$  for some  $\tilde{\theta} < 1$ .

Hence condition (c) follows. Set  $g(n) = (\log n)^\kappa$  and  $\tilde{g}(n) = (\log n)^{\kappa(1+\varepsilon)}$  for  $\varepsilon$  small. Then condition (a) follows automatically while condition (b) requires

$$\tilde{\theta}^{(\log n)^\kappa} \leq Cn^{-\zeta-\varepsilon}.$$

Provided  $\kappa > 1$  this relation holds (asymptotically) for any  $\beta > 0$ . ■

**Remark 3.2** The hypotheses stated on  $f$  above are not strong enough to prove the corresponding extreme value laws when  $m\{R > n\}$  decays sub-exponentially. The crucial estimate that needs to be checked is that on  $m(E_n)$ . We check this estimate for certain kinds of intermittency maps in Section 3.4.

### 3.3 Gibbs-Markov maps

Let  $(\Lambda, m)$  be a Lebesgue space with a countable measurable partition  $\alpha$ . Without loss, we suppose that all partition elements  $a \in \alpha$  have  $m(a) > 0$ . Recall that a measure-preserving transformation  $f : \Lambda \rightarrow \Lambda$  is a *Markov map* if  $f(a)$  is a union of elements of  $\alpha$  and  $f|_a$  is injective for all  $a \in \alpha$ . Define  $\alpha'$  to be the coarsest partition of  $\Lambda$  such that  $fa$  is a union of atoms in  $\alpha'$  for all  $a \in \alpha$ . If  $a_0, \dots, a_{n-1} \in \alpha$ , we define the  $n$ -cylinder  $[a_0, \dots, a_{n-1}] = \cap_{i=0}^{n-1} f^{-i}a_i$ . It is assumed that  $f$  and  $\alpha$  separate points in  $\Lambda$  (if  $x, y \in \Lambda$  and  $x \neq y$ , then for  $n$  large enough there exist distinct  $n$ -cylinders that contain  $x$  and  $y$ ).

Let  $0 < \beta < 1$ . We define a metric  $d_\beta$  on  $\Lambda$  by  $d_\beta(x, y) = \beta^{s(x, y)}$  where  $s(x, y)$  is the greatest integer  $n \geq 0$  such that  $x, y$  lie in the same  $n$ -cylinder.

Define  $g = Jf^{-1} = \frac{dm}{d(m \circ f)}$  and set  $g_k = g \circ f \dots \circ f \circ f^{k-1}$ .

The map  $f : \Lambda \rightarrow \Lambda$  is a *Gibbs-Markov map* if it satisfies the additional properties:

- (i) *Big images property:* There exists  $c > 0$  such that  $m(fa) \geq c$  for all  $a \in \alpha$ .
- (ii) *Distortion:*  $\log g|_a$  is Lipschitz with respect to  $d_\beta$  for all  $a \in \alpha'$ .

It follows from assumptions (i) and (ii) that there exists a constant  $\tilde{D} \geq 1$  such that for all  $x, y$  lying in a common  $k$ -cylinder  $[a_0, \dots, a_{k-1}]$ ,

$$\left| \frac{g_k(x)}{g_k(y)} - 1 \right| \leq \tilde{D}d_\beta(f^k x, f^k y) \quad \text{and} \quad \tilde{D}^{-1} \leq \frac{m[a_0, \dots, a_{k-1}]}{g_k(x)} \leq \tilde{D}. \quad (3.15)$$

For the Gibbs-Markov map there is an invariant measure  $\mu$  with density  $\rho$  and  $\log \rho|_a$  is Lipschitz with respect to  $d_\beta$  for all  $a \in \alpha'$ .

**Theorem 3.3** *Suppose that  $f : \Lambda \rightarrow \Lambda$  is a Gibbs-Markov map with  $\Lambda \subset \mathbb{R}$  compact. Then the hypotheses of Theorem 2.2 are satisfied.*

**Proof:** We check conditions (a), (b) and (c) in Theorem 2.2.

By [27] Gibbs-Markov maps have exponential decay of correlations and the invariant density lies in  $L^{1+\delta}(m)$  for some  $\delta > 0$ . As in the proof of Theorem 3.1 we let  $g(n) = (\log n)^\kappa$  and  $\tilde{g}(n) = (\log n)^{\kappa(1+\varepsilon)}$  for  $\varepsilon$  small. Conditions (a) and (b) easily follow.

For condition (c) we use Lemma 3.4 below. It implies that  $m(E_n) = \mathcal{O}((\log n)^{\kappa(1+\varepsilon)}/n)$  and so condition (c) of Theorem 2.2 holds.  $\blacksquare$

**Lemma 3.4** *Denote by  $\mathcal{E}_n(\varepsilon) := \{x \in \Lambda : |f^n(x) - x| \leq \varepsilon\}$ . Then  $m(\mathcal{E}_n(\varepsilon)) \leq C\varepsilon$  for some fixed constant  $C$ .*

**Proof:** Since  $|f'|$  is not uniformly bounded above, the conclusion does not immediately follow from the results of [4, 10] we used to prove Theorem 3.1. The proof relies on one-dimensional arguments. Let  $\mathcal{P}_n = \bigvee_{k=0}^{n-1} f^{-k}\alpha$  and consider one partition element  $\alpha_0 \in \mathcal{P}_n$ . We estimate  $m(\mathcal{E}_n(\varepsilon) \cap \alpha_0)$  using bounded distortion and uniform expansion of the map (or at least some iterate).

Since  $|(f^n|_{\alpha_0})'| > 1$ , there is a unique solution to  $f^n(x^\pm) = x^\pm \pm \varepsilon$  with  $x^\pm \in \alpha_0$  (provided such a solution exists). Moreover for all  $x \notin [x^-, x^+] \cap \alpha_0$  we have  $x \notin \mathcal{E}_n(\varepsilon)$  since  $|(f^n|_{\alpha_0})'| > 1$  and there is no sign change of  $(f^n)'$  on  $\alpha_0$ . We now estimate  $m([x^-, x^+]) = m(\mathcal{E}_n(\varepsilon) \cap \alpha_0)$ . We have

$$f^n(x^+) - f^n(x^-) = (x^+ - x^-) + 2\varepsilon,$$

By the mean value theorem

$$|f^n(x^+) - f^n(x^-)| = |(f^n)'(c)| |x^+ - x^-|, \quad \text{some } c \in \alpha_0.$$

By equation (3.15) it follows that

$$m(\alpha_0)|(f^n)'(c)| \in [\tilde{D}^{-1}, \tilde{D}]$$

and hence assuming  $\tilde{D} \text{diam}(\mathcal{P}_n) < 1$  (which is true for  $n$  sufficiently large) we have

$$|x^+ - x^-| \leq \frac{2\varepsilon m(\alpha_0)}{1 - \tilde{D}m(\alpha_0)}. \quad (3.16)$$

We note that a similar estimate holds for small values of  $n$  (since we have uniform estimates over a finite number of iterates). The only concern that we have bypassed is that there may be no solution to  $f^n(x^\pm) = x^\pm \pm \varepsilon$ . However the estimate in (3.16) is actually improved if  $\alpha_0$  only partially crosses  $\mathcal{E}_n(\varepsilon)$  (or does not intersect at all). Putting these estimate together and summing over all partition elements we have

$$m(\mathcal{E}_n(\varepsilon)) \leq \sum_{\alpha_0} \frac{2\varepsilon m(\alpha_0)}{1 - \tilde{D}m(\alpha_0)} \leq C\varepsilon$$

and so the result follows since  $\mu$  is equivalent to  $m$ . ■

### Examples:

For illustration we consider two explicit examples which are Gibbs-Markov.

**Example 3.5 (Piecewise affine Markov maps)** Suppose that  $\{a_k\}$  is a monotonic infinite sequence with  $a_0 = 1$  and  $\lim a_k = 0$ . Consider a piecewise affine expanding map  $f : [0, 1] \rightarrow [0, 1]$  defined as follows: let  $J_k = [a_{k+1}, a_k]$  and suppose that  $f$  on  $(a_{k+1}, a_k)$  is given by

$$f(x) = \frac{1}{a_k - a_{k+1}}(x - a_{k+1}).$$

It is easy to see that  $f$  is full branch on each  $J_k$  and  $\text{diam}(J_k) \rightarrow 0$  as  $k \rightarrow \infty$ . Hence  $f$  is Gibbs-Markov with respect to the partition  $\{J_k\}$ .

**Example 3.6 (The Gauss Map)** Consider the Gauss map  $S : (0, 1] \rightarrow [0, 1]$  given by  $S(x) = \frac{1}{x} - \lfloor \frac{1}{x} \rfloor$ . This is Gibbs-Markov with respect to the partition  $\{\Lambda_j\}$  with  $\Lambda_j = [(j+1)^{-1}, j^{-1}]$  and it has an ergodic invariant density  $\rho(x) = \{\log 2(1+x)\}^{-1}$ .

## 3.4 Non-uniformly expanding maps and intermittent maps

We discuss here applications to maps with neutral fixed points, such as the Pomeau-Manneville maps, and other families of Markov maps which are not necessarily Gibbs-Markov. We stress that the novelty of these results is that they apply to observations with multiple maxima. We consider first a class of non-uniformly expanding convex maps.

More precisely, we consider an interval map  $f : I \rightarrow I$  satisfying the following conditions

- (M1) There is a finite Markov decomposition  $I = \cup_j I_j$  such that each  $f(I_j)$  is a union of  $I_j$ 's. Let  $\mathcal{P} := \{I_j\}$  and  $\mathcal{P}_k := \bigvee_{i=0}^{k-1} f^{-i}\mathcal{P}$ . Assume that  $\lim_k \text{diam}(\mathcal{P}_k) = 0$  and  $f$  is topologically mixing.

- (M2) For each  $x \in \text{int}(I_j)$  we have  $f'(x) \geq 1$  with equality at a finite number of points. In addition  $f$  is convex on each  $I_j$ .
- (M3) For each  $I_j \in \mathcal{P}$ ,  $f$  is differentiable on  $\text{int}(I_j)$  and satisfies the regularity conditions: let  $I_j = [c_j, d_j]$  then  $f(x)$  is either uniformly expanding on  $I_j$  or for some  $\gamma > 1$  we have  $f(x) = f(c_j) + (x - c_j) + \mathcal{O}((x - c_j)^\gamma)$  as  $x \rightarrow c_j^+$ . In addition as  $x \rightarrow d_j^-$ ,  $f(x) = f(d_j) + \mathcal{O}(|x - d_j|^{\tilde{\gamma}})$  for some  $\tilde{\gamma} \in [0, 1]$ .

**Remark 3.7** If  $f'(x) \geq 1$  with equality holding at a finite number of points then it follows by convexity that equality can only hold at the lower boundary of each  $I_j \in \mathcal{P}$ . In (M3) we allow for singularities in the derivative of  $f$  at the upper boundaries of the  $I_j$ 's. The map  $f$  need not be Gibbs-Markov since we do not require bounded distortion on the partition elements of  $\mathcal{P}_k$ .

**Theorem 3.8** *Suppose that  $f : I \rightarrow I$  is a non-uniformly expanding interval map satisfying (M1), (M2) and (M3). Assume further that  $f$  has an ergodic measure  $\mu$  with density  $\rho \in L^{1+\delta}(m)$ . Then there is a value  $\tilde{\alpha}_0 = \tilde{\alpha}_0(\delta, \gamma) > 0$  that does not depend on the map  $f$  such that the hypotheses of Theorem 2.2 are satisfied if  $f$  has rate of decay of correlations  $\mathcal{O}(n^{-\tilde{\alpha}})$  with  $\tilde{\alpha} \geq \tilde{\alpha}_0$ .*

**Remark 3.9** Examples of Markov maps that satisfy (M1), (M2) and (M3) are intermittent maps such as the Liverani-Saussol-Vaianti family, see equation (3.22) ahead, and certain convex Lorenz-like maps, i.e. non-uniformly expanding maps with singularities in the derivative. These latter families are described in [5] where the correlations decay exponentially fast, and in [2] where intermittent Lorenz maps exhibit subexponential decay of correlations.

The proof of Theorem 3.8 has two steps. First we estimate the measure of the set  $\mathcal{E}_n(\varepsilon) := \{x \in [0, 1] : |f^n(x) - x| \leq \varepsilon\}$  and obtain an estimate in the form (2.5). Second, we rely on fast enough decay of correlations and use the fact that  $\rho \in L^{1+\delta}(m)$ . We begin with the following estimate, which is based on an argument that we learned from Henk Bruin.

**Lemma 3.10** *Suppose that  $f : I \rightarrow I$  satisfies (M1), (M2) and (M3). Denote by  $\mathcal{E}_n(\varepsilon) := \{x \in [0, 1] : |f^n(x) - x| \leq \varepsilon\}$ . Then there is some fixed constant  $C$  such that  $m(\mathcal{E}_n(\varepsilon)) \leq C \min\{\varepsilon^{1/2}, \varepsilon^{1/\gamma}\}$  for all  $n \geq 1$ .*

**Proof:** For each branch  $\alpha_n \in \mathcal{P}_n$  it follows that  $f^n(\alpha_n)$  contains a union of intervals from  $\mathcal{P}$  and  $f^n|_{\alpha_n}$  is a convex map.

Consider the  $k$ -th branch,  $f^n : J_k \rightarrow \cup_j I_j$ , where  $J_k := [a_k, b_k]$  and each  $f^n(a_k), f^n(b_k)$  is an end point of some  $I_j \in \mathcal{P}$ . Suppose that there are points

$x_k^\pm \in (a_k, b_k)$  such that  $f^n(x_k^\pm) = x_k^\pm \pm \varepsilon$ . If these points exist then they are unique, and in particular  $\mathcal{E}_n(\varepsilon) \cap J_k = [x_k^-, x_k^+]$ .

Let  $x \geq x_k^-$  in  $J_k$ . Since  $(f^n)'$  is increasing on  $J_k$ ,

$$(f^n)'(x) \geq (f^n)'(x_k^-) \geq \frac{\int_{[a_k, x_k^-]} (f^n)'(t) dt}{x_k^- - a_k} = \frac{x_k^- - \varepsilon - f^n(a_k)}{x_k^- - a_k},$$

hence

$$(f^n)'(x) - 1 \geq \frac{x_k^- - \varepsilon - f^n(a_k)}{x_k^- - a_k} - 1 = \frac{a_k - \varepsilon - f^n(a_k)}{x_k^- - a_k} \geq \frac{a_k - f^n(a_k) - \varepsilon}{m(J_k)}.$$

This implies that

$$2\varepsilon = \int_{[x_k^-, x_k^+]} [(f^k)'(t) - 1] dt \geq (x_k^+ - x_k^-) \frac{(a_k - \varepsilon - f^n(a_k))}{m(J_k)}$$

and thus

$$m(\mathcal{E}_n(\varepsilon) \cap J_k) \leq \frac{2\varepsilon}{(a_k - f^n(a_k) - \varepsilon)} m(J_k). \quad (3.17)$$

This estimate is useful provided  $(a_k - f^n(a_k) - \varepsilon)$  is not small. Let  $W^\eta = \cup\{a_k : |a_k - f^n(a_k)| < \eta + \varepsilon\}$  and let  $Z^\eta = \cup_{a_k \in W^\eta} J_k$ . Then

$$m(\mathcal{E}_n(\varepsilon)) = m(\mathcal{E}_n(\varepsilon) \cap Z^\eta) + m(\mathcal{E}_n(\varepsilon) \cap (Z^\eta)^c) \leq m(Z^\eta) + \frac{2\varepsilon}{\eta} m((Z^\eta)^c).$$

We estimate these sets in two different ways depending on whether  $n$  is large or small. Assume  $\varepsilon < \eta$  and now suppose that  $n$  is large enough with  $\text{diam}(\mathcal{P}_n) < \varepsilon$ . Then for all  $J_k \subset Z^\eta$  we have  $\text{diam}(J_k) < \varepsilon$ . Since  $\cup_k \{f^n(a_k)\}$  is a finite set it follows that

$$m(\mathcal{E}_n(\varepsilon)) \leq C\eta + \frac{2\varepsilon|I|}{\eta}$$

for some  $C > 0$ . The constant  $C$  depends on the cardinality of  $\cup_k \{f^n(a_k)\}$  and moreover is independent of  $n$  if  $\text{diam}(\mathcal{P}_n) < \varepsilon$ . Optimizing over  $\eta \in (0, 1)$  we find that  $\eta = \mathcal{O}(\sqrt{\varepsilon})$  is the best choice and gives  $m(\mathcal{E}_n(\varepsilon)) \leq \sqrt{\varepsilon}$ . Note that the analysis above assumes that we can always solve for such a  $x^\pm$ . If we cannot, then the estimates are actually improved in the sense that either  $\mathcal{E}_n(\varepsilon) \cap J_k = \emptyset$  or  $J_k$  partially crosses  $\mathcal{E}_n(\varepsilon)$  and the required measure would be smaller than that computed in equation (3.17).

In the case that  $n$  is small we cannot ensure  $m(Z^\eta)$  is small in  $\eta$  since for each  $a_k \in W^\eta$  it does not follow that the sets  $[a_k, b_k]$  are small in diameter. We still assume

$\varepsilon < \eta$ . By convexity it suffices to consider the  $J_k$  corresponding to the largest (or smallest)  $a_k$  in  $W^\eta$ . Let  $J_{k'}$  denote the set corresponding to the largest such  $a_k$ . We again solve  $f^n(x_k^\pm) = x_k^\pm \pm \varepsilon$  and consider a worst case scenario: namely  $a_{k'} = f^n(a_{k'})$  and  $(f^n)'(a_{k'}) = 1$  so that  $a_{k'}$  is a neutral periodic point. In this case there is no  $x^-$  solution and we just solve for  $x^+$ . Let  $\tilde{x} = x^+ - a_{k'}$ . The length of the interval  $[a_{k'}, x^+]$  is  $\tilde{x}$ . By (M2) and (M3):

$$\begin{aligned} \varepsilon &= f^n(x^+) - x^+ = f^n(a_{k'} + \tilde{x}) - (a_{k'} + \tilde{x}) \\ &\geq f(a_{k'} + \tilde{x}) - (a_{k'} + \tilde{x}) = f(a_{k'}) - a_{k'} + \mathcal{O}(\tilde{x}^\gamma). \end{aligned}$$

Since  $f(a_{k'}) > a_{k'}$  and assuming a solution exists for  $x^+$  we must have  $\tilde{x} = \mathcal{O}(\varepsilon^{1/\gamma})$ . Hence the conclusion of the Lemma follows.  $\blacksquare$

**Proof of Theorem 3.8.** We check conditions (a), (b) and (c) of Theorem 2.2. It suffices to take  $\gamma < 1/2$  so that  $m(\mathcal{E}_n(\varepsilon)) \leq \varepsilon^{1/\gamma}$ . We begin by setting  $g(n) = n^\kappa$ ,  $\tilde{g}(n) = n^{\kappa(1+\xi)}$  and compute inequality relations between  $\delta$ ,  $\gamma$  and  $\tilde{\alpha}$ . We show that conditions (a), (b) and (c) are satisfied provided  $\tilde{\alpha}$  is sufficiently large.

We first of all have  $m(E_n) = \mathcal{O}(\tilde{g}(n)/n^{1/\gamma})$ , and so  $\mu(E_n) = \mathcal{O}\left(\left(n^{\kappa(1+\xi)-1/\gamma}\right)^{\delta/(1+\delta)}\right)$ . So for condition (c) we must have

$$(\gamma^{-1} - \kappa)(\delta/(1+\delta)) > \beta \quad \text{and so} \quad \kappa < \frac{1}{\gamma} - \frac{\beta(1+\delta)}{\delta}. \quad (3.18)$$

Condition (a) of Theorem 2.2 requires  $g(n) = o(n^\beta)$  and so we must take  $\kappa < \beta < 1$ . Decay of correlations (condition (b)) requires:

$$(n^\kappa)^{-\tilde{\alpha}} \leq C n^{-(1+\beta)(1+\frac{1+\delta}{\delta})-\varepsilon} \quad \text{and so} \quad \kappa > \left(\frac{1+\beta}{\tilde{\alpha}}\right) \left(2 + \frac{1}{\delta}\right).$$

Putting these equations together we have the following relations between  $\delta$ ,  $\gamma$  and  $\tilde{\alpha}$  with  $\beta$  freely varying in  $(0, 1)$ :

$$\tilde{\alpha} > \frac{1+\beta}{\beta} \left(2 + \frac{1}{\delta}\right) \quad (3.19)$$

$$\tilde{\alpha} > \left(\frac{\gamma^{-1}\delta}{1+\delta} - \beta\right)^{-1} \left(\frac{(1+\beta)(1+2\delta)}{1+\delta}\right). \quad (3.20)$$

We can solve these inequalities simultaneously for  $\tilde{\alpha}$  to obtain a lower bound in terms of  $\delta$  and  $\gamma$ . Let  $\beta_0 := \delta(2\gamma(1+\delta))^{-1}$  so that  $\beta_0 < 1$ . Substituting  $\beta = \beta_0$  into equations (3.19) and (3.20) we can choose any  $\tilde{\alpha} > \tilde{\alpha}_0$  where

$$\tilde{\alpha}_0 = \delta^{-2} \{(2\gamma\delta + 2\gamma + \delta)(2\delta + 1)\}, \quad (3.21)$$

and this gives a (not necessarily optimal) lower bound on  $\tilde{\alpha}$ .  $\blacksquare$

**Remark 3.11** Consider the Liverani-Saussol-Vaienti map  $f = f_\omega : [0, 1] \rightarrow [0, 1]$  given by

$$f_\omega(x) = \begin{cases} x(1 + 2^\omega x^\omega), & x \in [0, \frac{1}{2}) \\ 2x - 1, & x \in [\frac{1}{2}, 1] \end{cases} \quad (3.22)$$

where  $\omega \in (0, 1)$ .

For maps of the form  $x \mapsto x + x^{1+\omega}$  it is known that the density has form  $h(x) \sim x^{-\omega}$  near indifferent fixed points and so lies in  $L^{\frac{1}{\omega}-\varepsilon}$  for any  $\varepsilon > 0$ , see [26]. The mixing rate for Lipschitz functions is  $\Theta_1(n) \approx n^{(1-\frac{1}{\omega})}$ , see [28]. It is then a straightforward calculation to show there exists  $\omega_0 \in (0, 1)$  such that the hypotheses of Theorem 2.2 are satisfied by  $f_\omega$  for  $\omega \in (0, \omega_0)$  where  $\omega_0 \approx 1/13$ .

We conjecture that the conclusions of Theorem 2.2 are satisfied for all  $\omega$  which give an SRB probability measure, as is the case for observations with a single maximum.

## 4 Proof of the main theorems

### 4.1 Proof of Theorem 2.2

The proof is based on that of Collet [4, Theorem 1.1] and uses a blocking argument from extreme value statistics. Here is an outline in the simple case of a single maximum. We write  $n \sim p(n)q(n)$  and choose a gap length  $t(n)$ . In our setting  $p \sim n^\beta$ ,  $q \sim n^{1-\beta}$  and  $t = g(n)$ . We divide successive observations of length  $n$  into  $q$  blocks of length  $p + t$ , where  $t$  is large enough that successive  $p$  blocks are approximately independent yet small enough that  $\mu(M_n \leq u_n) \sim \mu(M_{q(p+t)} \leq u_n)$ . Using approximate independence of the  $p$  blocks we establish  $|\mu(M_{q(p+t)} \leq u_n) - (1 - \mu(M_p \geq u_n))^q| \leq E_n$  and show  $(1 - \mu(M_p \geq u_n))^q \sim (1 - p\mu(\varphi > u_n))^q$ . The error  $E_n$  has several terms but the most significant is of form  $n \sum_{j=1}^p \mu(\varphi > u_n, \varphi \circ f^j > u_n)$  which must be controlled by an estimate of the measure of points which recur quickly.

The proof of Theorem 2.2 has two stages. In the first stage we collect a series of estimates that allow us to measure the set  $\{\psi \circ f^j \geq u_n, \psi \geq u_n\}$  for  $j$  small. This uses precisely the asymptotics of  $m(E_n)$  together with the maximal function technique of [4, Section 2], see Lemmas 4.1 and 4.2. However, for observations with multiple maxima we must then derive a series of estimates which measure the set of points that move from one maximum to another. These estimates are summarized in Lemmas 4.3 and 4.4. For  $j$  large we can measure the set  $\{\psi \circ f^j \geq u_n, \psi \geq u_n\}$  using decay of correlations, and the estimate to use is Proposition 4.6. Similar estimates are derived in [4, Section 3] but we optimize as far as possible in order to study applications with subexponential decay of correlations. See Proposition 4.8.

### 4.1.1 The maximal function technique of Collet

We summarize the method used by Collet [4] and then adapt it to handle multiple maxima. One main tool is the following:

**Lemma 4.1** *Assume a sequence of measurable sets  $E_n \subset \mathbb{R}^D$  has the property*

$$\mu(E_n) \leq \frac{C}{n^\alpha} \text{ for some } \alpha > 0.$$

*Then for  $0 < \beta < \alpha$  and  $\gamma > 1/(\alpha - \beta)$  one has: for  $m$ -a.e.  $x_0$  there is  $N(x_0)$  such that*

$$n \geq N(x_0) \implies \mu(\{d(x, x_0) \leq \frac{1}{n^\gamma}\} \cap E_{n^\gamma}) = m(B_1(0)) \frac{1}{n^{\gamma(D+\beta)}}. \quad (4.23)$$

**Proof:** Let  $\chi_n$  be the indicator function of  $E_n$  and consider the Hardy-Littlewood maximal function for  $\rho\chi_n$

$$L_n(x) := \sup_{\ell > 0} \frac{1}{m(B_\ell(x))} \int_{B_\ell(x)} \rho\chi_n(z) dz.$$

Since, by Hardy-Littlewood [20, Theorem 2.19],  $m(L_n > \lambda) \leq \frac{C}{\lambda} \|\rho\chi_n\|_{L^1} \leq \frac{C}{n^\alpha \lambda}$ , taking  $\lambda = n^{-\beta}$  we have

$$m(L_n > n^{-\beta}) \leq \frac{C}{n^{\alpha-\beta}}$$

and therefore

$$\sum_n m(L_{n^\gamma} > n^{-\beta\gamma}) \leq \sum_n \frac{C}{n^{\gamma(\alpha-\beta)}}$$

is finite for  $0 < \beta < \alpha$  and  $\gamma(\alpha - \beta) > 1$ .

Then  $m$ -a.e.  $x_0$  belongs to only finitely many sets  $\{L_{n^\gamma} > n^{-\beta\gamma}\}$  hence there is an  $N(x_0)$  such that  $x_0 \notin \{L_{n^\gamma} > n^{-\beta\gamma}\}$  for  $n \geq N(x_0)$  and therefore, taking  $\ell = 1/n^\gamma$

$$\frac{1}{m(B_\ell(x_0))} \int_{B_\ell(x_0)} \rho\chi_{n^\gamma}(z) dz \leq n^{-\beta\gamma}$$

which is exactly (4.23). ■

### 4.1.2 Recurrent points

We want to get a bound for returns up to  $g(n)$ . Recall that  $g(n)$  and  $\tilde{g}(n)$  are functions such that conditions (2.1), (2.5) and (2.2) are satisfied, where  $0 < \beta < 1$ ,  $\beta < \alpha$ ,  $\beta < D\alpha_1$ , and

$$E_n := \{x \mid d(x, f^j x) \leq \frac{2}{n} \text{ for some } j \leq \tilde{g}(n)\}.$$

Pick  $\gamma > 1/(\alpha - \beta)$  and let  $x_0$  satisfy the conclusion of Lemma 4.1 for the sets  $E_n$ .

Assume given  $v > 0$ . We want to show that

$$k \sum_{j=1}^{g(k^{1/D})} \mu(\{x \mid d(x, x_0) \leq \frac{v}{k^{1/D}}, d(f^j x, x_0) \leq \frac{v}{k^{1/D}}\}) \rightarrow 0 \text{ as } k \rightarrow \infty \quad (4.24)$$

which follows from (4.25) proved below, in view of (2.2), in particular that  $k^{-\beta/D} g(k^{1/D}) \rightarrow 0$ .

**Lemma 4.2** *For m-a.e.  $x_0$  and each  $v > 0$ , there is  $K = K(x_0, v)$  such that for  $k \geq K$  one has*

$$\mu(\{x \mid d(x, x_0) \leq \frac{v}{k^{1/D}}, d(f^j x, x_0) \leq \frac{v}{k^{1/D}} \text{ for some } j \leq g(k^{1/D})\}) \leq \frac{C_v}{k^{1+\frac{\beta}{D}}}. \quad (4.25)$$

The constant  $C_v$  depends only on  $v$ .

**Proof:** For  $k \geq 1$  denote

$$n = \left( \frac{1}{3v} k^{1/D} \right)^{1/\gamma}.$$

Note that  $n^\gamma = \frac{k^{1/D}}{3v}$  hence  $\frac{v}{k^{1/D}} < \frac{1}{n^\gamma}$ . Then

$$\begin{aligned} A &:= \{x \mid d(x, x_0) \leq \frac{v}{k^{1/D}}, d(f^j x, x_0) \leq \frac{v}{k^{1/D}} \text{ for some } j \leq g(k^{1/D})\} \\ &\subset \{x \mid d(x, x_0) \leq \frac{1}{n^\gamma}, d(f^j x, x) \leq \frac{2}{n^\gamma} \text{ for some } j \leq g(k^{1/D})\} \\ &\subset \{x \mid d(x, x_0) \leq \frac{1}{n^\gamma}, d(f^j x, x) \leq \frac{2}{n^\gamma} \text{ for some } j \leq \tilde{g}(n^\gamma)\} \end{aligned}$$

provided  $g(k^{1/D}) \leq \tilde{g}(n^\gamma) = \tilde{g}(k^{1/D}/3v)$ , which is true for  $k$  large enough by (2.1). The last set above is exactly

$$\{d(x, x_0) \leq \frac{1}{n^\gamma}\} \cap E_{n^\gamma}$$

and by Lemma 4.1 its  $\mu$ -measure is at most

$$m(B_1(0)) n^{-D\gamma - \gamma\beta} = \frac{C_v}{k^{1+\frac{\beta}{D}}}$$

for  $k$  (and thus  $n$ ) large enough. ■

### 4.1.3 Moving points

We consider here points that move from one maximum to another under  $f$ .

We assume that  $g(n)$  and  $\tilde{g}(n)$  satisfy conditions (2.1), (2.2) and (2.6). Denote by  $\mathcal{G}_1 \subset X$  the set of full Lebesgue measure determined by (2.6).

For a fixed  $y_0 \in \mathcal{G}_1$ , denote

$$G_n(y_0) := \{x \mid d(f^j x, y_0) \leq \frac{1}{n} \text{ for some } j \leq \tilde{g}(n)\}.$$

Then, by  $f$ -invariance of the measure, (2.6) gives

$$\mu(G_n(y_0)) \leq \tilde{g}(n)\mu(B_{1/n}(y_0)) \leq \frac{C_{y_0}}{n^{D\alpha_1}}$$

and therefore one can apply Lemma 4.1 to the sets  $G_n(y_0)$  for values  $0 < \beta < D\alpha_1$  and  $\gamma_1 > 1/(D\alpha_1 - \beta)$ . We prove:

**Lemma 4.3** *Given  $y_0 \in \mathcal{G}_1$ , for  $m$ -a.e.  $x_0$  and each  $v > 0$  there is  $K = K(x_0, v, y_0)$  such that for  $k \geq K$  one has*

$$\mu(\{x \mid d(x, x_0) \leq \frac{v}{k^{1/D}}, d(f^j x, y_0) \leq \frac{v}{k^{1/D}} \text{ for some } j \leq g(k^{1/D})\}) \leq \frac{C'_v}{k^{1+\frac{\beta}{D}}}.$$

The constant  $C'_v$  depends only on  $v$ .

This implies, as in the previous section, that for any  $v > 0$

$$k \sum_{j=1}^{g(k^{1/D})} \mu(\{x \mid d(x, x_0) \leq \frac{v}{k^{1/D}}, d(f^j x, y_0) \leq \frac{v}{k^{1/D}}\}) \rightarrow 0 \text{ as } k \rightarrow \infty. \quad (4.26)$$

In this case we require  $k^{-\beta/D}g(k^{1/D}) \rightarrow 0$ , but this is true from condition (a) in the statement of Theorem 2.2.

**Proof of Lemma 4.3.** It is essentially the same as the proof of Lemma 4.2.

Let  $x_0$  be given by Lemma 4.1 applied for the sets  $G_n(y_0)$ .

For  $k \geq 1$  denote

$$n = \left( \frac{1}{3v} k^{1/D} \right)^{1/\gamma_1}.$$

Note that  $n^{\gamma_1} = \frac{k^{1/D}}{3v}$  hence  $\frac{v}{k^{1/D}} < \frac{1}{n^{\gamma_1}}$ . Then, provided  $g(k^{1/D}) \leq \tilde{g}(n^{\gamma_1}) = \tilde{g}(k^{1/D}/3v)$  for the last inclusion below (true for  $k$  large enough by (2.1)),

$$\begin{aligned} & \{x \mid d(x, x_0) \leq \frac{v}{k^{1/D}}, d(f^j x, y_0) \leq \frac{v}{k^{1/D}} \text{ for some } j \leq g(k^{1/D})\} \\ & \subset \{x \mid d(x, x_0) \leq \frac{1}{n^{\gamma_1}}, d(f^j x, y_0) \leq \frac{1}{n^{\gamma_1}} \text{ for some } j \leq \tilde{g}(n^{\gamma_1})\} \\ & = \{x \mid d(x, x_0) \leq \frac{1}{n^{\gamma_1}}\} \cap G_{n^{\gamma_1}}(y_0) \end{aligned}$$

By Lemma 4.1, for  $k$  (and thus  $n$ ) large enough, its  $\mu$ -measure is at most

$$m(B_1(0)) \frac{1}{n^{D\gamma_1}} \frac{1}{n^{\gamma_1\beta}} = \frac{C'_v}{k^{1+\frac{\beta}{D}}}.$$

■

#### 4.1.4 Recurrence for multiple maxima

We prove now one of the key estimates in the argument of Collet.

**Lemma 4.4** *Let  $f : X \rightarrow X$  be a map for which the invariant measure  $\mu$  is finite and absolutely continuous, and Lemmas 4.2 and 4.3 hold. Denote by  $\mathcal{D}$  the set of density points of  $\mu$ .*

*For each  $r \geq 1$  there is a set  $\mathcal{X}_r \subset \mathcal{D}^r$  of full Lebesgue measure in  $X^r$  such that if  $(x_1, \dots, x_r) \in \mathcal{X}_r$  and  $\varphi(x) := -\sum_{k=1}^r \log d(x, x_k)$  then*

$$n \sum_{j=1}^{g(n^{1/D})} \mu(\varphi > u_n, \varphi \circ f^j > u_n) \rightarrow 0 \text{ as } n \rightarrow \infty \quad (4.27)$$

for  $u_n = v + \frac{1}{D} \log n$ .

**Proof:** For simplicity, we discuss the case  $r = 2$ ,  $\varphi(x) = -\log d(x, x_0) - \log d(x, y_0)$ . The proof generalizes in a straightforward way to  $r > 2$ .

From Lemmas 4.2 and 4.3 we have: there is  $\delta > 0$  such that

(a) for  $m$ -a.e.  $x_0 \in X$ , for any  $v > 0$ , there exists an  $\ell(x_0, v) > 0$  such that

$$\frac{1}{n} \leq \ell(x_0, v) \implies \mu\{x \in B_{\frac{v}{n^{1/D}}}(x_0) : d(f^j x, x_0) \leq \frac{v}{n^{1/D}} \text{ for some } j \leq g(n^{1/D})\} \leq \frac{C_v}{n^{1+\delta}}$$

Denote this set by  $\mathcal{G}$ .

(b) for each  $y_0 \in \mathcal{G}_1$ , for  $m$ -a.e.  $x_0 \in X$ , for any  $v > 0$ , there exists a  $k(x_0, v, y_0) > 0$  such that

$$\frac{1}{n} \leq k(x_0, v, y_0) \implies \mu\{x \in B_{\frac{v}{n^{1/D}}}(x_0) : d(f^j x, y_0) \leq \frac{v}{n^{1/D}} \text{ for some } j \leq g(n^{1/D})\} \leq \frac{C_v}{n^{1+\delta}}$$

Denote this set by  $\mathcal{G}(y_0)$ .

Therefore, by intersecting the full Lebesgue-measure sets  $\mathcal{D} \times \mathcal{D}$ ,  $\mathcal{G} \times \mathcal{G}$ ,  $\cup_{y \in \mathcal{G}_1} \{y\} \times \mathcal{G}(y)$  and  $\cup_{y \in \mathcal{G}_1} \mathcal{G}(y) \times \{y\}$  in  $X \times X$ , we conclude that for  $m$ -a.e. pair  $(x_0, y_0) \in X \times X$ , for each  $v > 0$ , there is  $s(x_0, y_0, v) > 0$  such that if  $\frac{1}{n} \leq s(x_0, y_0, v)$  then

$$\begin{aligned} \mu\{x \in B_{v/\tilde{n}}(x_0) : f^j(x) \in B_{v/\tilde{n}}(y_0) \text{ for some } 1 \leq j \leq g(\tilde{n})\} &\leq \frac{C_v}{n^{1+\delta}}, \\ \mu\{y \in B_{v/\tilde{n}}(y_0) : f^j(y) \in B_{v/\tilde{n}}(x_0) \text{ for some } 1 \leq j \leq g(\tilde{n})\} &\leq \frac{C_v}{n^{1+\delta}}, \\ \mu\{x \in B_{v/\tilde{n}}(x_0) : f^j(y) \in B_{v/\tilde{n}}(x_0) \text{ for some } 1 \leq j \leq g(\tilde{n})\} &\leq \frac{C_v}{n^{1+\delta}}, \\ \mu\{x \in B_{v/\tilde{n}}(y_0) : f^j(y) \in B_{v/\tilde{n}}(y_0) \text{ for some } 1 \leq j \leq g(\tilde{n})\} &\leq \frac{C_v}{n^{1+\delta}} \end{aligned}$$

where  $\tilde{n} = n^{1/D}$ . Noting that  $\mu\{-\log d(x, x_0) > u_n\} = \mu\{d(x, x_0) < e^{-v} n^{-1/D}\}$  this implies

$$n \sum_{j=1}^{g(n^{1/D})} \mu(\varphi > u_n, \varphi \circ f^j > u_n) \rightarrow 0 \text{ as } n \rightarrow \infty$$

for the observation  $\varphi(x) = -\log d(x, x_0) - \log d(x, y_0)$ . ■

#### 4.1.5 Conclusion of proof of Theorem 2.2

Let  $\psi(x)$  have logarithmic singularities at the points  $x_i$ , with the  $r$ -tuple  $(x_1, \dots, x_r)$  in the full-measure set  $\mathcal{X}_r$  specified in Lemma 4.4. Assume that  $C_k = 1$  for each  $1 \leq k \leq r$  (the general case can be reduced to this, see Lemma 4.12 and the discussion preceding it). In particular, each  $x_k$  is a density point of  $\mu$  and  $\mu(\psi \geq v + \frac{1}{D} \log n) = \mathcal{O}(e^{-Dv}/n)$  as  $n \rightarrow \infty$ .

For the details of the argument we refer to [4, Page 415]. We collect the main results below.

**Remark 4.5** If the density of the invariant measure  $\mu$  is in  $L^{1+\delta}(m)$  then, using Hölder's inequality,  $\mu(E) \leq Cm(E)^{\delta/(1+\delta)}$ . This gives the conclusion of [4, Lemma 2.2] with  $\theta = \delta/(1 + \delta)$ .

Assume  $v \in \mathbb{R}$  is fixed and denote  $u_n := v + D^{-1} \log n$ . In the following let  $pq = n$  with  $t = g(n)$  and  $p, q$  freely chosen with  $q(p+t) \approx n$ .

**Proposition 4.6** *The following results holds for  $1 \leq \ell \leq q$ :*

$$|\mu(M_{\ell(p+t)} < u_n) - (1 - p\mu(\psi > u_n))\mu(M_{\ell-1(p+t)} < u_n)| \leq \Gamma_n$$

where

$$\begin{aligned}\Gamma_n &= t\mu(\psi \geq u_n) + 2p \sum_{j=1}^p \mu(\{\psi \geq u_n\} \cap \{\psi \circ f^j \geq u_n\}) \\ &\quad + p\mathcal{O}(1)r_n^{\theta(D+\eta)} + p\mathcal{O}(1)r_n^{-(1+\eta)}\Theta_1(t)\end{aligned}$$

and  $r_n$  is the diameter of a connected component of  $\{\psi \geq u_n\}$ .

**Remark 4.7** In this formula  $\eta > 0$  can be chosen arbitrarily (see [4, Lemma 3.3] and the proof below), affecting only the constants  $\mathcal{O}(1)$ .

**Proof:** We first of all have the following estimates, see [4, Proposition 3.2]: for any integers  $t, r, m, k, p \geq 0$

$$0 \leq \mu(M_r < u) - \mu(M_{r+k} < u) \leq k\mu(\psi \geq u),$$

and

$$\begin{aligned}&\left| \mu(M_{m+p+t} < u) - \mu(M_m < u) + \sum_{j=1}^p \mathbb{E}(\chi_{\{\psi \geq u\}} \chi_{\{M_m \circ f^{p+t-j} < u\}}) \right| \\ &\leq 2p \sum_{j=1}^p \mathbb{E}(\chi_{\{\psi \geq u\}} \chi_{\{\psi \circ f^j \geq u\}}) + t\mu(\psi > u).\end{aligned}$$

The second estimate we need is that of decay of correlations for piecewise continuous observables (namely characteristic functions). Here we must approximate the indicator function  $\chi_{\{X > u_n\}}$  by a Lipschitz function.

The required bound is given in [12, Lemma 3.6]: for any  $\eta > 0$ , any integer  $t > 0$ , any ball  $B_\ell$ , and any measurable set  $A$ ,

$$|\mu(B_\ell \cap f^{-t}(A)) - \mu(B_\ell)\mu(A)| \leq C\|\rho\|_{1+\delta}\ell^{\theta(D+\eta)} + \mathcal{O}(1)\ell^{-(1+\eta)}\Theta_1(t). \quad (4.28)$$

where  $\Theta_1$  is the decay of Lipschitz functions against the  $L^\infty$ -norm,

$$\left| \int \varphi \cdot \psi \circ f^n d\mu - \int \varphi d\mu \int \psi d\mu \right| \leq \Theta_1(n)\|\psi\|_{L^\infty}\|\varphi\|_{\text{Lip}}.$$

This is obtained by approximating  $\chi_{B_\ell}$  with a radial function  $\Phi_B$  which is equal to 1 on the ball  $B_{\ell'}$  and decaying to zero at a linear rate between  $\ell'$  and  $\ell > \ell'$ . We take  $\ell' = \ell - \ell^{1+\eta}$  and  $B_{\ell'}$  is concentric to  $B_\ell$ . We first obtain

$$\|\rho(\Phi_B - \chi_{B_\ell})\|_1 \leq \mu(B_\ell \setminus B_{\ell'}) = \|\rho\chi_{B_\ell \setminus B_{\ell'}}\| \leq C\|\rho\|_{1+\delta}\ell^{\theta(D+\eta)} \quad (4.29)$$

where we have used Hölder's inequality in the right hand inequality. Using decay of correlations for Lipschitz functions (where the Lipschitz constant of  $\Phi_B$  is  $\mathcal{O}(\ell^{-(1+n)})$ ) and the approximation given by (4.29) the result follows.

In our case  $B = \{X \geq u_n\}$  is a finite union of balls since  $X$  has finitely many maxima. However, (4.28) remains valid as long as the balls that form  $\{X \geq u_n\}$  have comparable diameters (i.e., the ratios stay bounded).

Putting this together gives:

$$|\mu(M_n < u_n) - \mu(M_{q(p+t)} < u_n)| \leq qt\mu(\psi \geq u_n),$$

and for  $1 \leq \ell \leq q$  we have

$$\begin{aligned} & |\mu(M_{\ell(p+t)} < u_n) - (1 - p\mu(\psi > u_n))\mu(M_{(\ell-1)(p+t)} < u_n)| \\ & \leq \left| p\mu(\psi \geq u_n)\mu(M_{(\ell-1)(p+t)} < u_n) - \sum_{j=1}^p \mathbb{E} \left( \chi_{\{\psi \circ f^j \geq u_n\}} \chi_{M_{(\ell-1)(p+t)} \circ f^{p+t} < u_n} \right) \right| \\ & \quad + t\mu(\psi \geq u_n) + 2p \sum_{j=1}^p \mu(\{\psi \geq u_n\} \cap \{\psi \circ f^j \geq u_n\}). \end{aligned}$$

This latter expression is bounded above by  $\Gamma_n$  by the estimates above. ■

Continuing Collet's argument, we see that if  $p\mu(\psi \geq u_n) < 2$  then

$$|\mu(M_{q(p+t)} < u_n) - (1 - p\mu(\psi \geq u_n))^q| \leq q\Gamma_n$$

and so

$$|\mu(M_n < u_n) - (1 - p\mu(\psi \geq u_n))^q| \leq q\Gamma_n + qt\mu(\psi \geq u_n)$$

by the results above.

Hence if  $q\Gamma_n \rightarrow 0$  and  $qt\mu(\psi \geq u_n) \rightarrow 0$  then

$$\mu(\max\{\psi, \dots, \psi \circ f^n\} \leq u_n) \rightarrow \exp(-\lim_n n\mu(\psi \geq u_n)).$$

With our choice of  $\psi$ , since  $x_k$  are density points of  $\mu$ ,

$$\begin{aligned} \lim_n n\mu(\psi \geq u_n) &= \lim_n n \sum_{k=1}^r \mu(\{x \mid d(x, x_k) \leq \psi^{-1}(u_n)\}) \\ &= m(B_1(0))e^{-Dv} \sum_{k=1}^r \rho(x_k)e^{D\ell_k} \end{aligned} \tag{4.30}$$

where  $\ell_k = \lim_{x \rightarrow x_k} [\psi(x) + \log d(x, x_k)]$ .

**Proposition 4.8** *If  $t = g(n^{1/D})$ ,  $p = n^\beta$  and  $q = n^{1-\beta}$  then  $qt\mu(\psi \geq u_n) \rightarrow 0$  and  $q\Gamma_n \rightarrow 0$ .*

**Proof:** That  $qt\mu(\psi \geq u_n) \rightarrow 0$  follows from (2.2) since  $\mu(\psi \geq u_n) = \mathcal{O}(e^{-Dv}/n)$ .

Given the choice of constants, we verify in Sublemmas 4.10 and 4.11 below that each term in the formula for  $q\Gamma_n$  tends to zero.

The summation term that appears in  $q\Gamma_n$  is considered in Sublemma 4.11, which uses both the fast decay of correlations and the recurrence statistics. ■

**Remark 4.9** In our setting the choice of sequences for  $p$  and  $q$  is in fact optimal (this is important for specific applications): if we choose  $p = n^\gamma$  and  $q = n/p$ , then  $\gamma = \beta$  is optimal given our type of assumptions.

**Sublemma 4.10** *With the choice of  $t = g(n^{1/D})$ ,  $p = n^\beta$  and  $q = n^{1-\beta}$  the following estimates hold:*

$$qq(n^{1/D})\mu(\psi \geq u_n) \rightarrow 0, \quad pq r_n^{-(1+\eta)}\Theta_1(g(n^{1/D})) \rightarrow 0, \quad pq r_n^{\theta(D+\eta)} \rightarrow 0$$

where  $r_n$  is the diameter of a connected component of  $\{\psi \geq u_n\}$ .

**Proof:** By direct computation we have

$$qq(n^{1/D})\mu(\psi \geq u_n) = n^{1-\beta}g(n^{1/D})\mathcal{O}(e^{-Dv}/n) \rightarrow 0.$$

This is exactly condition (2.2), because  $\mu(\psi \geq u_n) = \mathcal{O}(e^{-Dv}/n)$  and the  $x_k$ 's are density points. For fixed  $v$ , we note that  $r_n = \mathcal{O}(1/n^{1/D})$ . Hence the following series of estimates hold.

$$pq r_n^{-(1+\eta)}\Theta_1(g(n^{1/D})) = n^{1+(1+\eta)/D}\Theta_1(g(n^{1/D})) \rightarrow 0.$$

This follows from (4.32). Finally we have

$$pq r_n^{\theta(D+\eta)} = n^{-\theta(1+\eta/D)+1} \rightarrow 0$$

since  $\theta(1 + \eta/D) > 1$ , again by (4.32). ■

**Sublemma 4.11** *With the choice of  $t = g(n^{1/D})$ ,  $p = n^\beta$  and  $q = n^{1-\beta}$  we have*

$$pq \sum_{j=1}^p \mu(\{\psi \geq u_n\} \cap \{\psi \circ f^j \geq u_n\}) \rightarrow 0 \tag{4.31}$$

**Proof:** We split the summation as  $\sum_{j=1}^p = \sum_{j \leq g(\tilde{n})} + \sum_{j \geq g(\tilde{n})}$  with  $\tilde{n} = n^{1/D}$ . For the part  $\sum_{j \geq g(\tilde{n})}$  we use mixing. We have:

$$\begin{aligned} & \mu(\{\psi \geq u_n\} \cap \{\psi \circ f^j \geq u_n\}) \\ & \leq r_n^{-(1+\eta)} \Theta_1(j) + \mathcal{O}(1) r_n^{\theta(D+\eta)} + \mu(\{\psi \geq u_n\})^2 \\ & \leq n^{(1+\eta)/D} \Theta_1(g(\tilde{n})) + \frac{1}{n^{\theta(1+\eta/D)}} + \frac{1}{n^2}, \end{aligned}$$

and hence we can bound  $pq \sum_{j=g(\tilde{n})}^p$  by

$$n^{1+\beta+(1+\eta)/D} \Theta_1(g(\tilde{n})) + \frac{\mathcal{O}(1)}{n^{\theta(1+\eta/D)-1-\beta}} + \frac{\mathcal{O}(1)}{n^{1-\beta}}.$$

This requires

$$n^{1+\beta+(1+\eta)/D} \Theta_1(g(\tilde{n})) \rightarrow 0, \quad \theta(1 + \eta/D) > 1 + \beta. \quad (4.32)$$

By Remark 4.5,  $\theta = \delta/(1 + \delta)$ . Then, by condition (2.4), there is an  $\eta > 0$  for which (4.32) holds.

The bound of  $pq \sum_{j \leq g(\tilde{n})}$  follows from (4.27), which relies on the the maximal function argument.  $\blacksquare$

We now conclude the proof of Theorem 2.2 by proving case (I) without the assumption  $C_k \equiv 1$ , and cases (II) and (III). We do this in detail only for case (II) via the next simple lemma. The proof of the other cases is the same mutatis mutandis.

**Lemma 4.12** *Suppose that  $\psi$  is a function with maxima at points  $(x_1, \dots, x_r) \in \mathcal{X}_r$  and at each  $x_k$   $\psi$  has a logarithmic singularity or  $\psi(x) \approx C_k d(x, x_k)^{s_k}$  with  $s_k < 0$ . Let  $\{I_k\}$  be a collection of disjoint neighborhoods of the points  $x_k$ ,  $k = 1, \dots, r$ . Let  $s = \max |s_k|$  and  $G$  be the set of  $x_k$ 's such that  $|s_k| = s$ . Let  $\tilde{\psi}$  be a function which equals  $\psi$  on  $I_i$  if  $x_i \in G$ , and which outside  $\cup_{x_i \in G} I_i$  is bounded and satisfies  $\tilde{\psi} \leq \psi$ . Let  $u_n = n^{s/D} v$ ,  $M_n = \max_{1 \leq j \leq n} \psi \circ f^j$  and  $\tilde{M}_n = \max_{1 \leq j \leq n} \tilde{\psi} \circ f^j$ . Suppose  $\lim_n \mu(\tilde{M}_n \leq u_n)$  exists. Then*

$$\lim_n \mu(M_n \leq u_n) = \lim_n \mu(\tilde{M}_n \leq u_n)$$

**Proof:** For large  $n$ ,  $\mu(\tilde{M}_n \geq u_n) \leq \mu(M_n \geq u_n) \leq \mu(\tilde{M}_n \geq u_n) + \sum_{x_i \notin G} n \mu(x \in I_i : \psi(x) > u_n)$ . By our choice of scaling  $u_n$ ,  $\lim_n \sum_{x_i \notin G} n \mu(x \in I_i : \psi(x) > u_n) = 0$ . Thus if  $\lim_n \mu(\tilde{M}_n \leq u_n)$  exists then  $\lim_n \mu(M_n \leq u_n)$  exists and  $\lim_n \mu(M_n \leq u_n) = \lim_n \mu(\tilde{M}_n \leq u_n)$ .  $\blacksquare$

Thus it suffices to consider functions of form  $\tilde{\psi}$  i.e. just the local behavior of  $\varphi$  at each  $x_k \in G$ . Finally the corresponding estimate in equation (4.30) yields

$$\begin{aligned} \lim_n n\mu(\psi \geq u_n) &= \lim_n n \sum_{x_k \in G} \mu(\{x \mid d(x, x_k) \leq \frac{1}{n^{1/D}} \left(\frac{C_k}{v}\right)^{1/s}\}) \\ &= m(B_1(0)) \sum_{x_k \in G} \rho(x_k) \left(\frac{C_k}{v}\right)^{D/s} \end{aligned}$$

## 4.2 Proof of Theorem 2.6

We will use the notations (2.7) and (2.8) throughout this section. In the proof of Theorem 2.6 first we show that the convergence in distribution is mixing [7]. Next we derive distributional convergence of  $\varphi_T(p)$  from knowledge of distributional convergence of  $\Phi_N(x)$  with  $N = N(x, T)$  a random function of time (governed by measure  $\mu$ ). The key result we prove in this derivation is Lemma 4.16 and this estimate is used when approximating  $T$  by  $N(x, T)$ . Extra care is needed if  $h$  is unbounded, see Sublemma 4.18. Finally the estimates are collected in Section 4.2.5 where Theorem 2.6 is proved.

### 4.2.1 Mixing convergence

Recall [23] that a sequence of random variables  $S_n : X \rightarrow \mathbb{R}$  on a probability space  $(X, \mu)$  converges mixing in distribution to  $G$  if for each  $A \subset X$  with positive measure,  $S_n|_A \rightarrow_d G$  with respect to the conditional measure  $\mu_A(B) := \mu(B \cap A)/\mu(A)$  on  $A$ .

Let  $(X, \mu, f)$  be a m.p.t. on a probability space and  $\Phi : X \rightarrow \mathbb{R}$  an (a.e. finite) random variable.

Denote  $\Phi_{n,m}(x) := \max\{\Phi \circ f^k(x) \mid n \leq k < m\}$ , for  $n, m \geq 0$ . Hence  $\Phi_n = \Phi_{0,n}$ .

**Lemma 4.13** *In the above setting, if  $f$  is ergodic then:*

- (a)  $\Phi_n \rightarrow \text{ess-sup}(\Phi)$  a.e.
- (b)  $a_n(\Phi_{0,n} - \Phi_{k,n+\ell}) \rightarrow_p 0$  as  $n \rightarrow \infty$ , for each  $k \geq 0$  and  $\ell$ .
- (c) Therefore, by [7, Thm. 6], if  $a_n(\Phi_n - b_n) \rightarrow_d G$ , then the convergence is mixing. Moreover, for each  $k \geq 0$  and  $\ell$ ,  $a_n(\Phi_{k,n+\ell} - b_n) \rightarrow_d G$  mixing as well.
- (d) Let  $h \in L^1(\mu)$  be a roof function and denote  $\widehat{\Phi}_n := \Phi_n \circ \pi^h$ .

Similarly to (c), if  $a_n(\widehat{\Phi}_n - b_n) \rightarrow_d G$  on  $(X^h, \mu^h)$ , then the convergence is mixing.

**Remark 4.14** Note that if  $a_n(\Phi_n - b_n) \rightarrow_d G$ , then (b) implies that  $a_n(\Phi_{n+k} - b_n) \rightarrow_d G$  for any fixed  $k$ . Hence, by Khintchine's Theorem (see, e.g., [18, Theorem 1.2.3]),  $a_{n+k}/a_n \rightarrow 1$  and  $a_{n+k}(b_n - b_{n+k}) \rightarrow 0$  as  $n \rightarrow \infty$ . Conditions (2.9) and (2.10) are stronger versions of this.

**Proof:** Part (a) is a straightforward consequence of the Birkhoff ergodic theorem.

The result in (b) is a consequence of (a), with some extra care needed if  $\text{ess-sup}(\Phi) = \infty$  or if  $\Phi$  is maximized on a non-isolated set of points (this latter case does not arise in our applications). We consider the case when  $\ell \geq 0$ , the other being similar. Then

$$\Phi_{0,n} - \Phi_{0,n+\ell} \leq \Phi_{0,n} - \Phi_{k,n+\ell} \leq \Phi_{0,n+\ell} - \Phi_{k,n+\ell}$$

and it follows by ergodicity that we get a.e. convergence to zero if  $\text{ess-sup}(\Phi)$  is finite. For the case in which  $\text{ess-sup}(\Phi) = \infty$ , we show that the left side (which is non-positive) and right side (which is non-negative) in the above inequalities converge in probability to zero. Let  $\varepsilon > 0$  and pick  $a < \text{ess-sup}(\Phi)$  such that  $\mu(\Phi_{0,k} \geq a) \leq \varepsilon$  (this is possible because  $\Phi$  is a.e. finite and  $\text{ess-sup}(\Phi) = \infty$ ). Then

$$\begin{aligned} \mu(a_n(\Phi_{0,n+\ell} - \Phi_{k,n+\ell}) \geq \varepsilon) &= \mu(\Phi_{0,k} \geq \Phi_{k,n+\ell} + \varepsilon/a_n) \\ &\leq \mu(\Phi_{0,k} \geq a) + \mu(\Phi_{0,k} < a, \Phi_{0,k} \geq \Phi_{k,n+\ell} + \varepsilon/a_n) \\ &\leq \mu(\Phi_{0,k} \geq a) + \mu(\Phi_{k,n+\ell} < a - \varepsilon/a_n) \\ &\leq \mu(\Phi_{0,k} \geq a) + \mu(\Phi_{k,n+\ell} < a) \end{aligned}$$

which converges to  $\mu(\Phi_{0,k} \geq a)$  by (a). The other side is dealt with in a similar fashion, using the stationarity of the process:

$$\begin{aligned} \mu(a_n(\Phi_{0,n+\ell} - \Phi_{0,n}) \geq \varepsilon) &\leq \mu(\Phi_{n,n+\ell} \geq a) + \mu(\Phi_{0,n} < a - \varepsilon/a_n) \\ &\leq \mu(\Phi_{n,n+\ell} \geq a) + \mu(\Phi_{0,n} < a). \end{aligned}$$

The first claim of (c) follows from ergodicity, (b), and the following result of Eagleson [7, Thm. 6]: if  $X_1, X_2, \dots$  have a trivial invariant  $\sigma$ -field,  $T_n(X_1, \dots, X_n) \rightarrow_d T$  for a sequence of statistics  $T_n$ , and  $T_n(X_1, \dots, X_n) - T_n(X_{k+1}, \dots, X_{k+n}) \rightarrow_p 0$  for each  $k$ , then  $T_n(X_1, \dots, X_n) \rightarrow_d T$  mixing.

The ‘‘moreover’’ part of (c) follows immediately invoking (b) once more: let  $A \subset X$  be a set of positive measure; then, in the induced measure  $\mu_A$ ,  $a_n(\Phi_n - b_n)|_A \rightarrow_d G$  by the first claim of (c), and  $a_n(\Phi_{0,n} - \Phi_{k,n+\ell})|_A \rightarrow_p 0$  from (b); thus  $a_n(\Phi_{k,n+\ell} - b_n)|_A \rightarrow_d G$  as well.

The result in (d) follows as the first claim of (c) if we notice that  $\widehat{\Phi}_{1,n+1} - \widehat{\Phi}_{0,n} \rightarrow_p 0$  on  $(X^h, \mu^h)$  because  $\Phi_{1,n+1} - \Phi_{0,n} \rightarrow_p 0$  on  $(X, \mu)$  and  $h \in L^1(\mu)$ . Note that the invariant  $\sigma$ -field of  $\{(\Phi \circ f^k) \circ \pi^h\}_k$  is the pull-back through  $\pi^h$  of the invariant  $\sigma$ -field on  $X$ , hence it is still trivial.  $\blacksquare$

### 4.2.2 Lifting to $X^h$

Mixing convergence allows to relate the extreme value laws for observations on  $X$  to observations on  $X^h$ . We prove the following lemma:

**Lemma 4.15** *Let  $f : (X, \mu) \rightarrow (X, \mu)$  be ergodic and  $X^h$  the suspension space with a roof function  $h \in L^1(\mu)$ . Let  $\Phi : X \rightarrow \mathbb{R}$  be an observation.*

*Set  $\Phi_N(x) = \max_{k \leq N-1} \Phi(f^k(x))$  and define  $\widehat{\Phi}_N : X^h \rightarrow \mathbb{R}$  by  $\widehat{\Phi}_N(x, u) = \Phi_N \circ \pi^h(x, u) = \Phi_N(x)$ .*

(a) *If  $a_N(\Phi_N - b_N) \rightarrow_d G$  on  $X$ , then  $a_N(\widehat{\Phi}_N - b_N) \rightarrow_d G$  on  $X^h$ .*

(b) *If  $1/h \in L^1(\mu)$  and  $a_N(\widehat{\Phi}_N - n_N) \rightarrow_d G$  on  $X^h$ , then  $a_N(\Phi_N - b_N) \rightarrow_d G$  on  $X$ .*

**Proof:** By [24], mixing convergence in distribution on  $(X, \mu)$  implies convergence in distribution on  $(X, \nu)$  whenever  $\nu$  is absolutely continuous with respect to  $\mu$ .

For (a), take  $\nu$  to be the probability measure on  $X$  give by  $d\nu := h/\bar{h}d\mu$ . Then

$$\int_X \exp\{ita_N(\Phi_N - b_N)\} h/\bar{h} d\mu \rightarrow \mathbb{E}(e^{itG}), \quad \forall t \in \mathbb{R}$$

because the convergence  $a_N(\Phi_N - b_N) \rightarrow_d G$  is mixing by Lemma 4.13(c). However,

$$\begin{aligned} \int_{X^h} \exp\{ita_N(\widehat{\Phi}_N - b_N)\} d\mu^h &= \frac{1}{\bar{h}} \int_X \int_0^{h(x)} \exp\{ita_N(\widehat{\Phi}_N - b_N)\} du d\mu \\ &= \int_X \exp\{ita_N(\Phi_N - b_N)\} h/\bar{h} d\mu. \end{aligned}$$

Hence  $\int_{X^h} \exp\{ita_N(\widehat{\Phi}_N - b_N)\} d\mu^h \rightarrow \mathbb{E}(e^{itG})$ , and the result follows.

For (b), note that the probability measure  $d\nu^h := \bar{h}/h d\mu^h$  is an absolutely continuous probability with respect to  $d\mu^h$ . Repeat the argument from above and use Lemma 4.13(d). ■

### 4.2.3 The normalization constants

**Lemma 4.16** *Assume that  $g_n : X \rightarrow \mathbb{Z}$  are measurable and such that  $g_n(x)/n \rightarrow 0$  a.e. Let  $S_n$  be an increasing sequence of random variables on  $X$ . If conditions (2.9) and (2.10) hold then*

$$a_n(S_n - b_n) \rightarrow_d G \iff a_n(S_{n+g_n(x)}(x) - b_n) \rightarrow_d G.$$

**Proof:** If we define  $X_n = a_n(S_n - b_n)$  and  $Y_n = a_n(S_{n+g_n} - b_n)$  then the lemma is a straightforward consequence of the inequality

$$X_n \leq Y_n \leq \frac{a_n}{a_{n+\varepsilon n}} X_{a_{n+\varepsilon n}} + a_n(b_{n+\varepsilon n} - b_n) \quad (4.33)$$

valid on the set  $|\frac{g_n}{n}| < \varepsilon$ . ■

We will also need the following consequence of (4.33):

**Remark 4.17** (a) If  $\liminf_{n \rightarrow \infty} \mu(a_n(S_{n+g_n(x)}(x) - b_n) \leq v) \geq G(v)$  at each continuity point  $v$  of  $G$ , then  $\liminf_{m \rightarrow \infty} \mu(a_m(S_m - b_m) \leq v) \geq G(v)$  at each continuity point  $v$  of  $G$ .

(b) If  $\limsup_{n \rightarrow \infty} \mu(a_n(S_{n+g_n(x)}(x) - b_n) \leq v) \leq G(v)$  at each continuity point  $v$  of  $G$ , then  $\limsup_{m \rightarrow \infty} \mu(a_m(S_m - b_m) \leq v) \leq G(v)$  at each continuity point  $v$  of  $G$ .

#### 4.2.4 The lap number

Assume that  $h(x)$  satisfies the Strong Law of Large Numbers (SLLN), namely

$$h_N = N\bar{h} + o(N) \text{ a.e. as } N \rightarrow \infty \quad (4.34)$$

where  $h_N(x) = h(x) + h(f(x)) + \dots + h(f^{N-1}(x))$ .

Given a time  $T \geq 0$  define the *lap number*  $N(x, T)$  by

$$h_{N(x, T)}(x) \leq T < h_{N(x, T)+1}(x). \quad (4.35)$$

The SLLN implies that  $h$  is a.e. finite, hence

$$\lim_{T \rightarrow \infty} N(x, T) = \infty \text{ a.e.} \quad (4.36)$$

and thus

$$\lim_{T \rightarrow \infty} \frac{T}{N(x, T)} = \bar{h} \text{ a.e.} \quad (4.37)$$

We also need an estimate comparing  $N(x, T + h(x))$  to  $N(x, T)$ . These quantities are not uniformly comparable, especially if  $\inf h = 0$  and  $\sup h = \infty$ . However the next result confirms that they are comparable almost surely.

**Sublemma 4.18** *For  $\mu$  almost all  $x \in X$  we have*

$$\lim_{T \rightarrow \infty} \frac{N(x, T + h(x))}{N(x, T)} = 1$$

**Proof:** Let  $Z_a = \{x \in X : h(x) \leq a\}$ , and given  $\varepsilon > 0$  and  $T > 0$  let

$$X_{\varepsilon,T} = \{x \in X : \left| \frac{N(x, t+h(x))}{N(x, t)} - 1 \right| \geq \varepsilon \text{ for some } t \geq T\}.$$

Then we have

$$\mu(X_{\varepsilon,T}) \leq \mu \left( x \in X : \left| \frac{N(x, t+h(x))}{N(x, t)} - 1 \right| \geq \varepsilon \text{ for some } t \geq T, h(x) \in [0, a] \right) + \mu(X \setminus Z_a).$$

Now for given  $a > 0$  we have that  $\mu$ -a.e.

$$\frac{t}{N(x, t)} = \bar{h} + o(1), \quad \frac{t+a}{N(x, t+a)} = \bar{h} + o(1) \text{ as } t \rightarrow \infty,$$

and therefore  $N(x, t+a)/N(x, t) \rightarrow 1$  almost surely as  $t \rightarrow \infty$ . Hence by taking  $a$  arbitrarily large and then  $T \rightarrow \infty$  it follows that  $\mu(X_{\varepsilon,T}) \rightarrow 0$ . The result follows.  $\blacksquare$

#### 4.2.5 Conclusion of proof of Theorem 2.6.

Note that the hypotheses about  $f$  and  $h$  imply all the statements in Section 4.2.4.

The main observation is that for  $(x, u) \in X^h$  with  $0 \leq u < h(x)$

$$\Phi_{1, N(x, T)}(x) \leq \varphi_T(x, u) \leq \Phi_{N(x, T+h(x))+1}(x) \quad (4.38)$$

(recall from Section 4.2.1 that  $\Phi_{n,m}(x) := \max\{\Phi \circ f^k(x) \mid n \leq k < m\}$ ). Indeed,  $\Phi_N(x) = \varphi_{h_N(x)}(x, 0)$  for  $x \in X$  and thus, taking into account the identifications of  $X^h$ :

$$\begin{aligned} \Phi_{1, N(x, T)}(x) &= \max\{\varphi(x, t) \mid h(x) \leq t < h_{N(x, T)}(x)\} \\ \varphi_T(x, u) &= \max\{\varphi(x, t) \mid u \leq t < u + T\} \\ \Phi_{N(x, T+h(x))+1}(x) &= \max\{\varphi(x, t) \mid 0 \leq t < h_{N(x, T+h(x))+1}(x)\}. \end{aligned}$$

The definition (4.35) of the lap number gives

$$h_{N(x, T)}(x) \leq T, \quad u + T < T + h(x) < h_{N(x, T+h(x))+1}(x),$$

and (4.38) follows.

We will also use that

$$\frac{N(x, T)}{\lfloor T/\bar{h} \rfloor} \rightarrow 1, \quad \frac{N(x, T+h(x))+1}{\lfloor T/\bar{h} \rfloor} \rightarrow 1 \text{ a.e. on } X \text{ (and hence on } X^h), \quad (4.39)$$

which follow from (4.37) and Sublemma 4.18.

We first prove the implication

$$a_n(\Phi_n - b_n) \rightarrow_d G \text{ on } X \implies a_{T/\bar{h}}(\varphi_T - b_{T/\bar{h}}) \rightarrow_d G \text{ on } X^h. \quad (4.40)$$

By Lemma 4.15(a), the left hand side of (4.40) implies

$$a_{\lfloor T/\bar{h} \rfloor}(\widehat{\Phi}_{\lfloor T/\bar{h} \rfloor} - b_{\lfloor T/\bar{h} \rfloor}) \rightarrow_d G \text{ on } X^h, \quad (4.41)$$

where  $\lfloor r \rfloor$  denotes the largest integer not exceeding  $r$ . By Lemma 4.13(b)

$$a_{\lfloor T/\bar{h} \rfloor}(\Phi_{1, \lfloor T/\bar{h} \rfloor} - \Phi_{\lfloor T/\bar{h} \rfloor}) \rightarrow_p 0 \text{ on } X$$

and, because  $h \in L^1(\mu)$ , this convergence in probability also holds on  $X^h$  if we extend the function  $\Phi_{1,N}$  to  $\widehat{\Phi}_{1,N} := \Phi_{1,N} \circ \pi^h$ . Together with (4.41) this implies

$$a_{\lfloor T/\bar{h} \rfloor}(\widehat{\Phi}_{1, \lfloor T/\bar{h} \rfloor} - b_{\lfloor T/\bar{h} \rfloor}) \rightarrow_d G \text{ on } X^h. \quad (4.42)$$

By Lemma 4.16, (4.39), (4.42) and (4.41) imply

$$a_{\lfloor T/\bar{h} \rfloor}(\widehat{\Phi}_{1,N(x,T)}(x,u) - b_{\lfloor T/\bar{h} \rfloor}) \rightarrow_d G, \quad a_{\lfloor T/\bar{h} \rfloor}(\widehat{\Phi}_{N(x,T+h(x))+1}(x,u) - b_{\lfloor T/\bar{h} \rfloor}) \rightarrow_d G \quad \text{on } X^h.$$

Use (4.38) to obtain

$$a_{\lfloor T/\bar{h} \rfloor}(\varphi_T - b_{\lfloor T/\bar{h} \rfloor}) \rightarrow_d G \text{ on } X^h,$$

from where the desired conclusion follows.

We prove next the converse implication,

$$a_{\lfloor T/\bar{h} \rfloor}(\varphi_T - b_{\lfloor T/\bar{h} \rfloor}) \rightarrow_d G \text{ on } X^h \implies a_n(\Phi_n - b_n) \rightarrow_d G \text{ on } X. \quad (4.43)$$

Denote by  $\Omega \subset \mathbb{R}$  the continuity points of  $G$ .

Since (4.38) implies

$$\begin{aligned} \mu^h(a_{\lfloor T/\bar{h} \rfloor}(\widehat{\Phi}_{1,N(x,T)}(x,u) - b_{\lfloor T/\bar{h} \rfloor}) \leq v) &\geq \mu^h(a_{\lfloor T/\bar{h} \rfloor}(\varphi_T(x,u) - b_{\lfloor T/\bar{h} \rfloor}) \leq v) \\ \mu^h(a_{\lfloor T/\bar{h} \rfloor}(\widehat{\Phi}_{N(x,T+h(x))+1}(x,u) - b_{\lfloor T/\bar{h} \rfloor}) \leq v) &\leq \mu^h(a_{\lfloor T/\bar{h} \rfloor}(\varphi_T(x,u) - b_{\lfloor T/\bar{h} \rfloor}) \leq v) \end{aligned}$$

we obtain from the left hand side of (4.43) that

$$\begin{aligned} \liminf_{T \rightarrow \infty} \mu^h(a_{\lfloor T/\bar{h} \rfloor}(\widehat{\Phi}_{1,N(x,T)}(x,u) - b_{\lfloor T/\bar{h} \rfloor}) \leq v) &\geq G(v), \\ \limsup_{T \rightarrow \infty} \mu^h(a_{\lfloor T/\bar{h} \rfloor}(\widehat{\Phi}_{N(x,T+h(x))+1}(x,u) - b_{\lfloor T/\bar{h} \rfloor}) \leq v) &\leq G(v), \quad v \in \Omega. \end{aligned}$$

By Remark 4.17 and (4.39), we conclude that

$$\liminf_{T \rightarrow \infty} \mu^h(a_{\lfloor T/\bar{h} \rfloor}(\widehat{\Phi}_{1, \lfloor T/\bar{h} \rfloor} - b_{\lfloor T/\bar{h} \rfloor}) \leq v) \geq G(v), \quad (4.44)$$

$$\limsup_{T \rightarrow \infty} \mu^h(a_{\lfloor T/\bar{h} \rfloor}(\widehat{\Phi}_{\lfloor T/\bar{h} \rfloor} - b_{\lfloor T/\bar{h} \rfloor}) \leq v) \leq G(v), \quad v \in \Omega. \quad (4.45)$$

Use that, by Lemma 4.13(b),  $a_{[T/\bar{h}]}(\widehat{\Phi}_{[T/\bar{h}]} - \widehat{\Phi}_{1,[T/\bar{h}]}) \rightarrow_p 0$  on  $X^h$  to deduce from the first relation above that

$$\liminf_{T \rightarrow \infty} \mu^h(a_{[T/\bar{h}]}(\widehat{\Phi}_{[T/\bar{h}]} - b_{[T/\bar{h}]}) \leq v) \geq G(v), \quad v \in \Omega. \quad (4.46)$$

From (4.45) and (4.46) it follows that

$$a_{[T/\bar{h}]}(\widehat{\Phi}_{[T/\bar{h}]} - b_{[T/\bar{h}]}) \rightarrow_d G \text{ on } X^h.$$

and Lemma 4.15(b) completes the proof. ■

## References

- [1] V. Baladi. *Positive Transfer Operators and Decay of Correlations*. Advanced Series in Nonlinear Dynamics **16**, World Scientific, Singapore, 2000.
- [2] R. J. Bhansali and M. P. Holland. Frequency analysis of chaotic intermittency maps with slowly decaying correlations. *Statist. Sinica* **17** (2007), no. 1, 15–41.
- [3] H. Bruin, B. Saussol, S. Troubetsky and S. Vaienti. Return time statistics via inducing. , *Ergod. Th. Dyn. Sys.* **23** (2003) 991-1013.
- [4] P. Collet ‘Statistics of closest return for some non-uniformly hyperbolic systems’ *Erg. Th. Dyn. Syst.* **21** (2001), 401-420.
- [5] K. Díaz-Ordaz, M. P. Holland and S. Luzzatto. Statistical properties of one-dimensional maps with critical points and singularities. *Stoch. Dyn.* **6** (2006), no. 4, 423–458.
- [6] D. Dolgopyat. ‘Limit theorems for partially hyperbolic systems’ *Trans AMS* **356** (2004) 1637-1689.
- [7] G. K. Eagleson. *Some simple conditions for limit theorems to be mixing*. (Russian) *Teor. Veroyatnost. i Primenen.* **21** (1976), no. 3, 653–660. English translation: *Theor. Probability Appl.* **21** (1976), no. 3, 637–642 (1977)
- [8] J. Freitas and A. Freitas. ‘Extreme values for Benedicks Carleson maps’. To appear *Erg. Thy. Dyn. Sys.*.
- [9] J. Freitas and A. Freitas. “On the link between dependence and independence in Extreme Value Theory for Dynamical Systems” *Stat. Probab. Lett.* **78** (2008) 1088–1093.

- [10] J. Freitas, A. Freitas and M. Todd. “Hitting Times and Extreme Values”, *Preprint*.
- [11] J. Galambos. *The Asymptotic Theory of Extreme Order Statistics*, John Wiley and Sons, 1978.
- [12] C. Gupta. “Extreme Value Distributions for some classes of Non-Uniformly Partially Hyperbolic Dynamical Systems” *Preprint*
- [13] C. Gupta, M. P. Holland and M. Nicol. “E.V.T. for Lozi maps, Lorenz-like maps and dispersing billiards.” *Preprint* 2009.
- [14] N. Haydn, Y. Lacroix and S. Vaienti. “Hitting time and return time statistics in ergodic dynamical systems”, *Annals of Probability*, **33**, (2005) 2043–2050.
- [15] N. Haydn and S. Vaienti, “The compound Poisson distribution and returns times in dynamical systems”, To appear in *Probability Theory and Related Fields*.
- [16] M. Hirata. “Poisson law for Axioma A diffeomorphisms”. *Erg. Thy. Dyn. Sys.* **13** (1993) 533–556.
- [17] M. Hirata, B. Saussol and S. Vaienti. “Statistics of return times: a general framework and new applications”, *Comm. Math. Phys.* **206** (1999) 33–55.
- [18] M. R. Leadbetter, G. Lindgren and H. Rootzén. *Extremes and Related Properties of Random Sequences and Processes*, Springer-Verlag, 1980.
- [19] R. M. Loynes. ‘Extreme values in uniformly mixing stationary stochastic processes’ *Ann. Math. Statist.* **36**, (1965) 993-999.
- [20] P. Mattila, *Geometry of sets and measures in Euclidean spaces*, **44**, Cambridge studies in advanced mathematics, 1995.
- [21] I. Melbourne and A. Török. Statistical limit theorems for suspension flows. *Israel Journal of Math.* **144**, (2004) 191-209.
- [22] B. Pitskel “Poisson limit law for Markov chains”, *Erg. Thy. Dyn. Sys* **11** (1991) 501-513.
- [23] Rényi, A. ‘On mixing sequences of sets’. *Acta Math. Acad. Sci. Hungar.* **9** (1958), 215–228.
- [24] Rényi, A. ‘Contributions to the theory of independent random variables’. (Russian. English summary) *Acta Math. Acad. Sci. Hungar.* **1** (1950), 99–108.

- [25] S. I. Resnick. *Extreme values, regular variation, and point processes*, Applied Probability Trust, **4**, Springer-Verlag, 1987.
- [26] M. Thaler Transformations on  $[0, 1]$  with infinite invariant measures. *Israel J. Math.* **46**, (1983), no. 1-2, 67–96.
- [27] L.-S. Young. Statistical properties of dynamical systems with some hyperbolicity. *Ann. of Math.* **147** (1998) 585–650.
- [28] L.-S. Young. Recurrence times and rates of mixing. *Israel J. Math.* **110** (1999) 153–188.