SPEED OF CONVERGENCE FOR LAWS OF RARE EVENTS AND ESCAPE RATES

ANA CRISTINA MOREIRA FREITAS, JORGE MILHAZES FREITAS, AND MIKE TODD

Abstract. We obtain error terms on the rate of convergence to Extreme Value Laws for a general class of weakly dependent stochastic processes. The dependence of the error terms on the ‘time’ and ‘length’ scales is very explicit. Specialising to data derived from a class of dynamical systems we find even more detailed error terms, one application of which is to consider escape rates through small holes in these systems.

1. Introduction

The study of the statistics of extreme events is both of classical importance, and a crucial topic across contemporary science. Classically, the underlying stochastic processes are assumed to be independently distributed, but many more recent developments in this topic relate to the study of Extreme Value Laws (EVL) for dependent systems. A standard approach to prove the existence of EVL in this setting is to check some conditions on the underlying process, for example Leadbetter’s conditions $D(u_n)$ and $D'(u_n)$ in [Lea74]. Inspired by [Col01], in a series of works [FF08, FFT10, FFT12, FFT12] the authors developed these conditions so that they had wider application, the main motivation being to stochastic processes coming from dynamical systems. A natural question to now ask is: how fast is the convergence to the EVL? (To rephrase, at a given finite stage in the process, what is the difference, or error term, between the law observed up to this time and the asymptotic law?) For example, if the convergence were to be very slow, in a simulation the laws would be essentially invisible. In this paper we address this question, particularly in the context of the latter papers above.

Error terms in the i.i.d. context are rather well-known, see for example [HW79, Smi82] and the discussion in [Res08, Section 2.4]. However, the literature on dependent processes
is much less extensive (for one case, see [MS01]). On the other hand, if we think of our stochastic process as coming from a Markov chain or, more generally, a dynamical system, then there is an equivalence between EVL and Hitting Time Statistics (HTS) (see [FFT10]), which then yields a significant body of literature coming from that side on these error terms [GS97, Aba01, Aba04, AV09, AS11, Kel12]. In this paper, taking inspiration from all these areas, we obtain sophisticated estimates on rates of convergence, where the dependence on time and ‘length’ (i.e., the distance from the maximum) scales is made explicit. We will first give general error terms under very general mixing conditions (in a way that unifies both clustering and non-clustering cases), and then impose some stronger conditions on our underlying process to obtain better estimates.

1.1. A more technical introduction. Let \( X_0, X_1, \ldots \) be a stationary stochastic process, where each random variable (r.v.) \( X_i : \mathcal{Y} \rightarrow \mathbb{R} \) is defined on the measure space \( (\mathcal{Y}, \mathcal{B}, \mathbb{P}) \).

We assume, without loss of generality, that \( \mathcal{Y} \) is a sequence space with a natural product structure so that each possible realisation of the stochastic process corresponds to a unique element of \( \mathcal{Y} \) and there exists a measurable map \( T : \mathcal{Y} \rightarrow \mathcal{Y} \), the time evolution map, which can be seen as the passage of one unit of time, so that

\[
X_{i-1} \circ T = X_i, \quad \text{for all } i \in \mathbb{N}.
\]

**Note.** There is an obvious relation between \( T \) and the shift map but we avoid that comparison here because we are definitely not reduced to the usual shift dynamics, in the sense that normally the shift map acts on sequences from a finite or countable alphabet, while here \( T \), acts on spaces like \( \mathbb{R}^\mathbb{N} \), in the sense that the sequences can be thought as being obtained from an alphabet like \( \mathbb{R} \).

Stationarity means that \( \mathbb{P} \) is \( T \)-invariant. Note that \( X_i = X_0 \circ T^i \), for all \( i \in \mathbb{N}_0 \), where \( T^i \) denotes the \( i \)-fold composition of \( T \), with the convention that \( T^0 \) denotes the identity map on \( \mathcal{Y} \).

We denote by \( F \) the cumulative distribution function (d.f.) of \( X_0 \), i.e., \( F(x) = \mathbb{P}(X_0 \leq x) \). Given any d.f. \( F \), let \( \bar{F} = 1 - F \) and let \( u_F \) denote the right endpoint of the d.f. \( F \), i.e., \( u_F = \sup\{x : F(x) < 1\} \). We say we have an exceedance of the threshold \( u < u_F \) at time \( j \in \mathbb{N}_0 \) whenever \( \{X_j > u\} \) occurs.

We define a new sequence of random variables \( M_1, M_2, \ldots \) given by

\[
M_n = \max\{X_0, \ldots, X_{n-1}\}. \tag{1.1}
\]

We say that we have an Extreme Value Law (EVL) for \( M_n \) if there is a non-degenerate d.f. \( H : \mathbb{R} \rightarrow [0, 1] \) with \( H(0) = 0 \) and, for every \( \tau > 0 \), there exists a sequence of levels \( u_n = u_n(\tau), n = 1, 2, \ldots, \) such that

\[
n\mathbb{P}(X_0 > u_n) \rightarrow \tau, \quad \text{as } n \rightarrow \infty, \tag{1.2}
\]

and for which the following holds:

\[
\mathbb{P}(M_n \leq u_n) \rightarrow \bar{H}(\tau), \quad \text{as } n \rightarrow \infty. \tag{1.3}
\]
where the convergence is meant at the continuity points of $H(\tau)$.

Now let us assume that our underlying system is an ergodic measure-preserving dynamical system $f : \mathcal{X} \to \mathcal{X}$ where $(\mathcal{X}, \mathcal{B}, \mu)$ is a probability space. (Note that in this context we will often use $\mathbb{P}$ in place of $\mu$ to be more consistent with the rest of the paper.) Given an observable $\varphi : \mathcal{X} \to \mathbb{R} \cup \{\pm \infty\}$, we can define $X_n = \varphi \circ f^n$ for each $n \in \mathbb{N}$ (see Section 3.2 for more details) and ask what the EVL here is. If $\varphi$ is sufficiently regular and takes its maximum at a unique point $\zeta \in \mathcal{X}$ then the EVL here is related to the following concept.

Consider a set $A \in \mathcal{B}$. We define a function that we refer to as first hitting time function to $A$, denoted by $r_A : \mathcal{X} \to \mathbb{N} \cup \{+\infty\}$ where

$$r_A(x) = \min \{j \in \mathbb{N} \cup \{+\infty\}: f^j(x) \in A\}.$$  \hspace{1cm} (1.4)

The restriction of $r_A$ to $A$ is called the first return time function to $A$. We define the first return time to $A$, which we denote by $R(A)$, as the infimum of the return time function to $A$, i.e.,

$$R(A) = \inf_{x \in A} r_A(x).$$ \hspace{1cm} (1.5)

Given a point $\zeta \in \mathcal{X}$, by Kac Lemma the expected value of $r_{B_\varepsilon(\zeta)}$ when restricted to the $\varepsilon$-ball $B_\varepsilon(\zeta)$ is $1/\mu(B_\varepsilon(\zeta))$. We say that the system has HTS $H$ for balls around $\zeta$ if for each $t \in [0, \infty)$,

$$\lim_{\varepsilon \to 0} \mu(\{\mu(B_\varepsilon(\zeta))r_{B_\varepsilon(\zeta)} > t\}) = H(t)$$

for some d.f. $H : [0, \infty) \to [0, 1]$.

In [FFT10] the link between HTS and EVL was demonstrated and exploited. So the main question in this paper is what are the orders of

$$|\mathbb{P}(M_n \leq u_n) - \bar{H}(\tau)|$$

and

$$|\mu(\{\mu(B_\varepsilon(\zeta))r_{B_\varepsilon(\zeta)} > t\}) - H(t)|?$$

For the rest of this introduction we will refer to $\tau$ and $t$ as the time scale; and $n$ and $\mu(B_\varepsilon(\zeta))$ as the length scale (although, strictly speaking, the true time is rescaled by the inverse of the length).

In the context of HTS, the early results were mostly for sequences of shrinking dynamically defined cylinders, that is sets $A_n$ where $A_n$ was a maximal subset on which $f^n : A_n \to \mathcal{X}$ is a homeomorphism and $\zeta \in A_n$ for all $n \in \mathbb{N}$. Moreover, certain strong mixing conditions were imposed. For example [GS97] gave error terms in terms of a power of the length scale for $\psi$-mixing systems. A similar result was obtained, but in the context of $\varphi$-mixing systems, in [Aba01], where lower bounds were also found. Various similar results are described in [AG01], including a description of results of [Ald82] for Markov chains and the results in [HSV99].

A major breakthrough was made in [Aba04] where for $\varphi$-mixing systems with cylinders, the error terms also incorporated the time scale, so that increasing time $t$ meant that the
error terms decreased by a factor comparable to $H(t)$. Further refinements were also made in [AV09]. A similar approach was used for $\alpha$-mixing processes in [AS11], but the error terms were not as powerful, in particular, the time scale was not so nicely decoupled from length in the estimates. We remark that in all the results mentioned so far there is a parameter $\xi_{A_n}$ present, which converges in the limit, but allows a convenient perturbation of the asymptotic law $H$ to improve the apparent convergence. The results we present here are of a similar form to [AV09], but are not restricted to balls and do not include this extra factor.

In the dependent EVL context, [MS01] obtained error terms of the order of the length, but not independently of $\tau$. This is also the case for many of the results in the EVL literature in the i.i.d. case. We should also mention the rather complete results in [Kel12] (see also [KL09]) where, under an assumption of the nice behaviour of a transfer operator, similar results to ours were proved and applied in the EVL and HTS context (see further comments on this in Section 3.3, in particular Remark 3.10). One of the main results there, and one of our key motivations here, is to use the error terms to find limit laws for escape of mass through a hole (a small ball), see for example [DY06]. By decoupling the time parameter in a suitable way in our error terms we are able to estimate how the escape rate depends on the size of the hole.

1.2. Structure of the paper. In Section 2 we first go into a few more technicalities and history of this topic, in particular explaining quantities which deal with clustering and the extremal index, before going on to state and prove our first main result, Theorem 2.2. In Section 3 we suppose that our data comes from a dynamical system with a nice mixing property and state and comment on our other main results, Theorems 3.4 and 3.6 and Corollary 3.12. In Section 4 we prove some preparatory results we’ll use to prove these theorems, while in Section 5 we prove Theorem 3.4 and in Section 6 we prove Theorem 3.6. In Section 7 we give a natural application of our main results.

Acknowledgements. MT would like to thank A. Galves for an inspiring conversation. The authors would like to thank H. Bruin and N. Haydn for helpful comments and suggestions.

2. Error terms for general stationary stochastic processes under conditions of the type $D_2$ and $D'$

In what follows for every $A \in \mathcal{B}$, we denote the complement of $A$ as $A^c := \mathcal{X} \setminus A$.

For some $u \in \mathbb{R}$, $q \in \mathbb{N}$, we define the events:

$$U(u) := \{X_0 > u\} \quad \text{and} \quad A^{(q)}(u) := U(u) \cap \bigcap_{i=1}^{q} T^{-i}(U(u)^c) = \{X_0 > u, X_1 \leq u, \ldots, X_q \leq u\}. \quad \text{(2.1)}$$

We also set $A^{(0)}(u) := U(u), \; U_n := U(u_n) \; \text{and} \; A^{(q)}_n := A^{(q)}(u_n)$, for all $n \in \mathbb{N}$ and $q \in \mathbb{N}_0$. 
Let $B \in \mathcal{B}$ be an event. For some $s \geq 0$ and $\ell \geq 0$, we define:
\[
W_{s,\ell}(B) = \bigcap_{i=[s]}^{[s]+\max\{\ell,1\}-1} T^{-i}(B^c).
\] (2.2)
We will write $W^c_{s,\ell}(B) := (W_{s,\ell}(B))^c$. Whenever is clear or unimportant which event $B \in \mathcal{B}$ applies, we will drop the $B$ and write just $W_{s,\ell}$ or $W^c_{s,\ell}$. Observe that
\[
W_{0,\ell}(U(u)) = \{M_n \leq u\}.
\]

After the success of the classical Extremal Types Theorem of Fisher-Tippet and Gnedenko in the i.i.d. setting, there has been a great deal of interest in studying the existence of EVL for dependent stationary stochastic processes. Building upon the work of Loynes and Watson, in [Lea74], Leadbetter proposed two conditions on the dependence structure of the stochastic processes, which he called $D(u_n)$ and $D'(u_n)$, that guaranteed the existence of the same EVLs of the i.i.d. applied to the partial maxima of sequences of random variables satisfying those conditions.

Condition $D(u_n)$ is a sort of uniform mixing condition adapted to this setting of extreme values where the main events of interest are exceedances of the threshold $u_n$. Condition $D'(u_n)$ precludes the existence of clusters of exceedances of $u_n$. Under these two conditions the EVL obtained corresponds to a standard exponential distribution, where $\bar{H}(\tau) = e^{-\tau}$.

In certain circumstances, observed data clearly showed the existence of clusters of exceedances, which meant that $D'(u_n)$ did not hold. This motivated the study of EVL and the affect of clustering. In fact it was observed that clustering of exceedances essentially produced the same type of EVL but with an parameter $0 \leq \theta \leq 1$, the Extremal Index (EI), so that $\bar{H}(\tau) = e^{-\theta\tau}$: here $\theta$ quantifies the intensity of clustering. In order to show the existence of EVLs with a certain EI $\theta \leq 1$, new conditions (replacing $D'(u_n)$) were devised. We mention condition $D''(u_n)$ of [LN89] and particularly the more general condition $D^{(k)}(u_n)$ of [CHM91], which also includes the case of absence of clustering. In fact, $D^{(k)}(u_n)$ (in the formulation of [CHM91, Equation (1.2)]) is equal to $D'(u_n)$, when $k = 1$ and to $D''(u_n)$ when $k = 2$. Together with condition $D(u_n)$, the condition $D^{(k)}(u_n)$ gave an EVL, where $\bar{H}(\tau) = e^{-\theta\tau}$, with an EI $\theta$ given by O’Brien’s formula:
\[
\theta = \lim_{n \to \infty} \frac{\mathbb{P}\left(A^{(q)}_{n}\right)}{\mathbb{P}(U_n)}.
\] (2.3)

When the stochastic processes arise from dynamical systems as described in Section 3 below, condition $D(u_n)$ cannot be verified using the usual available information about mixing rates of the system except in some very special situations, and even then only for certain subsequences of $n$. This means that the theory developed by Leadbetter and others (such as Nandagopalan, Chernick, Hsing and McCormick) is not practical in this dynamical systems context. For that reason, motivated by the work of Collet ([Col01]), the first and second named authors proposed a new condition called $D_2(u_n)$ for general
stationary stochastic processes, much weaker than $D(u_n)$, which together with $D'(u_n)$ admitted a proof of the existence of EVL in the absence of clustering (with $\theta = 1$). The great advantage of $D_2(u_n)$ is that it is so much weaker than $D(u_n)$ and follows easily for stochastic processes arising from systems with sufficiently fast decay of correlations. In the argument of [FF08], this weakening was achieved by a fuller application of condition $D'(u_n)$.

In [FFT12], the authors proved a connection between periodicity and clustering. Motivated by the behaviour at periodic points, which created the appearance of clusters of exceedances, the authors proposed new conditions in order to prove the existence of EVLs with EI less than 1. The main idea was that, under a condition $SP$, the authors proposed new conditions in order to prove the existence of EVLs activated by the behaviour at periodic points, which created the appearance of clusters of exceedances.

In [FFT12], the authors proved a connection between periodicity and clustering. Motivated by the behaviour at periodic points, which created the appearance of clusters of exceedances, the authors proposed new conditions in order to prove the existence of EVLs with EI less than 1. The main idea was that, under a condition $SP$, the authors proposed new conditions in order to prove the existence of EVLs activated by the behaviour at periodic points, which created the appearance of clusters of exceedances. In [FFT12], the authors proved a connection between periodicity and clustering. Motivated by the behaviour at periodic points, which created the appearance of clusters of exceedances, the authors proposed new conditions in order to prove the existence of EVLs activated by the behaviour at periodic points, which created the appearance of clusters of exceedances.

In [FFT12], the authors proved a connection between periodicity and clustering. Motivated by the behaviour at periodic points, which created the appearance of clusters of exceedances, the authors proposed new conditions in order to prove the existence of EVLs with EI less than 1. The main idea was that, under a condition $SP$, the authors proposed new conditions in order to prove the existence of EVLs activated by the behaviour at periodic points, which created the appearance of clusters of exceedances. In [FFT12], the authors proved a connection between periodicity and clustering. Motivated by the behaviour at periodic points, which created the appearance of clusters of exceedances, the authors proposed new conditions in order to prove the existence of EVLs activated by the behaviour at periodic points, which created the appearance of clusters of exceedances.

In [FFT12], the authors proved a connection between periodicity and clustering. Motivated by the behaviour at periodic points, which created the appearance of clusters of exceedances, the authors proposed new conditions in order to prove the existence of EVLs with EI less than 1. The main idea was that, under a condition $SP$, the authors proposed new conditions in order to prove the existence of EVLs activated by the behaviour at periodic points, which created the appearance of clusters of exceedances. In [FFT12], the authors proved a connection between periodicity and clustering. Motivated by the behaviour at periodic points, which created the appearance of clusters of exceedances, the authors proposed new conditions in order to prove the existence of EVLs activated by the behaviour at periodic points, which created the appearance of clusters of exceedances.

In [FFT12], the authors proved a connection between periodicity and clustering. Motivated by the behaviour at periodic points, which created the appearance of clusters of exceedances, the authors proposed new conditions in order to prove the existence of EVLs with EI less than 1. The main idea was that, under a condition $SP$, the authors proposed new conditions in order to prove the existence of EVLs activated by the behaviour at periodic points, which created the appearance of clusters of exceedances. In [FFT12], the authors proved a connection between periodicity and clustering. Motivated by the behaviour at periodic points, which created the appearance of clusters of exceedances, the authors proposed new conditions in order to prove the existence of EVLs activated by the behaviour at periodic points, which created the appearance of clusters of exceedances.
For some fixed $q \in \mathbb{N}_0$, consider the sequence $(t_n)_{n \in \mathbb{N}}$, given by condition $\mathcal{D}(u_n)$ and let $(k_n)_{n \in \mathbb{N}}$ be another sequence of integers such that

$$k_n \to \infty \quad \text{and} \quad k_n t_n = o(n). \quad (2.5)$$

**Condition** $(\mathcal{D}_q'(u_n))$. We say that $\mathcal{D}_q'(u_n)$ holds for the sequence $X_0, X_1, X_2, \ldots$ if there exists a sequence $(k_n)_{n \in \mathbb{N}}$ satisfying $(2.5)$ and such that

$$\lim_{n \to \infty} n \sum_{j=1}^{[n/k_n]} \mathbb{P} \left( A^{(q)}_n \cap T^{-j}(A^{(q)}_n) \right) = 0. \quad (2.6)$$

**Remark 2.1.** Note that condition $\mathcal{D}_q'(u_n)$ is condition $D^{(q)}(u_n)$ from [CHM91]. Moreover, if $q = 0$ then we get condition $D'(u_n)$ from Leadbetter. Thus the following result, Theorem 2.2, gives in particular a generalisation of [CHM91, Corollary 1.3] since our condition $\mathcal{D}(u_n)$ is much weaker that the original $D(u_n)$ of Leadbetter. Moreover, as discussed above, condition $\mathcal{D}(u_n)$ follows from sufficiently fast decay of correlations of the underlying stochastic processes.

The following is the most general of the main theorems in this paper.

**Theorem 2.2.** Let $X_0, X_1, \ldots$ be a stationary stochastic process and $(u_n)_{n \in \mathbb{N}}$ a sequence satisfying $(1.2)$, for some $\tau > 0$. Assume that conditions $\mathcal{D}(u_n)$ and $\mathcal{D}_q'(u_n)$, for some $q \in \mathbb{N}_0$, are satisfied. Then, there exists $C > 0$ such that for all $n \in \mathbb{N}$ we have

$$\left| \mathbb{P}(M_n \leq u_n) - e^{-\theta \tau} \right| \leq C \left( k_n t_n \frac{\tau}{n} + n \gamma(q, n, t_n) + n \sum_{j=1}^{[n/k_n]} \mathbb{P} \left( A^{(q)}_n \cap T^{-j}(A^{(q)}_n) \right) \right.
$$

$$+ \left. e^{-\theta \tau} \left( \theta \tau - n \mathbb{P} \left( A^{(q)}_n \right) \right) + \frac{\tau^2}{k_n} + q \mathbb{P} \left( U_n \setminus A^{(q)}_n \right) \right),$$

where the EI $\theta$ is given by equation (2.3).

**Remark 2.3.** Behind the convergence result there is a ‘blocking argument’ and several quantities are related to it. Namely, the number of blocks taken $(k_n)$ and the size of the gaps between the blocks $(t_n)$ have to satisfy $(2.5)$. The first error term, depending on the choices for adequate $k_n$ and $t_n$, typically, decays like $n^{-\delta}$, for some $0 < \delta < 1$. The second term depends on the long range mixing rates of the process $(\mathcal{D}(u_n))$. The third term takes into account the short range recurrence properties $(\mathcal{D}'(u_n))$. The fourth has two components, the first depends on the asymptotics of relation $(2.3)$ and the second on the number of blocks, which must be traded off with the first term. Note that the term $e^{-\theta \tau} \frac{\tau^2}{k_n}$ also appears in the i.i.d. case since it results from expansion $(2.4)$ below. The fifth term results from replacing $U_n$ by $A^{(q)}_n$ (see Proposition 2.4) and should decay like $1/n$. The constant $C$ may depend on the rates just mentioned but not on $\tau$.

The rest of this section is devoted to the proof of Theorem 2.2.
The following result gives a simple estimate but a rather important one. It is crucial in removing condition $SP_{p,\theta}(u_n)$ from [FFT12], to present in a unified way the results under the presence and absence of clustering in Theorem 2.2 and, most of all, to obtain the sharper results in Theorems 3.4 and 3.6.

**Proposition 2.4.** Given an event $B \in \mathcal{B}$, let $q, n \in \mathbb{N}$ be such that $q < n$ and define $A = B \setminus \bigcup_{j=1}^{q} f^{-j}(B)$. Then

$$|\mathbb{P}(\mathcal{W}_{0,n}(B) - \mathbb{P}(\mathcal{W}_{0,n}(A))| \leq \sum_{j=1}^{q} \mathbb{P}(\mathcal{W}_{0,n}(A) \cap f^{-n+j}(B \setminus A)).$$

**Proof.** Since $A \subset B$, then clearly $\mathcal{W}_{0,n}(B) \subset \mathcal{W}_{0,n}(A)$. Hence, we have to estimate the probability of $\mathcal{W}_{0,n}(A) \setminus \mathcal{W}_{0,n}(B)$ which corresponds to the set of points that at some time before $n$ enter the $B$ but never enter its subset $A$.

Let $x \in \mathcal{W}_{0,n}(A) \setminus \mathcal{W}_{0,n}(B)$. Then the orbit of $x$ must enter $B$ during the time period from 1 to $n - 1$ but it must never enter $A$ during the same time. We will see that there exists $j \in \{1, \ldots, q\}$ such that $f^{n-j}(x) \in B$. In fact, suppose that no such $j$ exists. Then let $\ell = \max\{i \in \{1, \ldots, n - 1\} : f^i(x) \in B\}$ be the last moment the orbit of $x$ enters $B$ during the time period in question. Then, clearly, $\ell < n - q$. Hence, if $f^i(x) \notin B$, for all $i = \ell + 1, \ldots, n - 1$ then we must have that $f^\ell(x) \in A$ by definition of $A$. But this contradicts the fact that $x \in \mathcal{W}_{0,n}(A)$. Consequently, we have that there exists $j \in \{1, \ldots, q\}$ such that $f^{n-j}(x) \in B$ and since $x \in \mathcal{W}_{0,n}(A)$ then we can actually write $f^{n-j}(x) \in B \setminus A$.

This means that $\mathcal{W}_{0,n}(A) \setminus \mathcal{W}_{0,n}(B) \subset \bigcup_{j=1}^{q} f^{-n+j}(B \setminus A) \cap \mathcal{W}_{0,n}(A)$ and then

$$|\mathbb{P}(\mathcal{W}_{0,n}(B) - \mathbb{P}(\mathcal{W}_{0,n}(A))| \leq \mathbb{P}(\mathcal{W}_{0,n}(A) \setminus \mathcal{W}_{0,n}(B))$$

$$\leq \mathbb{P} \left( \bigcup_{j=1}^{q} f^{-n+j}(B \setminus A) \cap \mathcal{W}_{0,n}(A) \right) \leq \sum_{j=1}^{q} \mathbb{P}(\mathcal{W}_{0,n}(A) \cap f^{-n+j}(B \setminus A)), $$

as required. \qed

In what follows we will need the error term of the limit expression $\lim_{k \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x$, for every $x \in \mathbb{R}$, which is given by:

$$\left(1 + \frac{x}{n}\right)^n = e^x \left(1 - \frac{x^2}{2n} + \frac{x^3(8 + 3x)}{24n^2} + \ldots \right). \quad (2.7)$$

Also, by Taylor’s expansion, for every $\delta \in \mathbb{R}$ and $x \in \mathbb{R}$ we have

$$|e^{x+\delta} - e^x| \leq e^x \left(|\delta| + e^{|\delta|} \delta^2 / 2 \right). \quad (2.8)$$

The strategy is to use a blocking argument, that goes back to Markov, which consists of splitting the data into blocks with gaps of increasing length. There are three main steps. The first step is to estimate the error produced by neglecting the data corresponding to the gaps. The second is to use essentially the mixing condition $\mathcal{D}(u_n)$ (with some help from
Proof of Theorem 2.2. Letting $A = A_n^{(q)}$, $\ell = \lfloor n/k_n \rfloor$, $k = k_n$ and $t = t_n$ on Proposition 2.7, we obtain

$$
\left| \mathbb{P}(\mathcal{W}_{0,n}(A^{(q)}_n)) - \left( 1 - \frac{n}{k_n} \right) \mathbb{P}(A^{(q)}_n) \right|^k_n \leq 2k_n t_n \mathbb{P}(U_n) + n \sum_{j=1}^{[n/k_n]} \mathbb{P}(A^{(q)}_n \cap \mathcal{W}_{0|n/k_n}(A^{(q)}_n)) + k_n \left( \frac{n}{k_n} \right) \mathbb{P}(A^{(q)}_n \cap \mathcal{W}_{\ell+t-j,t}(A^{(q)}_n)) \right|.
$$

(2.9)

Using condition $\mathcal{D}(u_n)$, we have that for the third term:

$$
k_n \left( \frac{n}{k_n} \right) \mathbb{P}(A^{(q)}_n \cap \mathcal{W}_{0|n/k_n}(A^{(q)}_n)) - \sum_{j=0}^{[n/k_n]} \mathbb{P}(A^{(q)}_n \cap \mathcal{W}_{\ell+t-j,t}(A^{(q)}_n)) \right| \leq n \gamma(q, n, t_n).
$$

(2.10)
By (2.8), we have that there exists $C$ such that
\[ |e^{-\floor{n/k_n} P(A_n^{(q)} - e^{-\theta\tau}}| \leq C e^{-\theta\tau} \left\{ \left| \theta\tau - n P(A_n^{(q)}) \right| + o(\left| \theta\tau - n P(A_n^{(q)}) \right|) \right\} \]
\[ \leq C e^{-\theta\tau} |\theta\tau - n P(A_n^{(q)})|. \]
Using (2.7) and (2.8), there exists $C' > 0$ such that
\[ |1 - \left( \frac{n}{k_n} \right) P(A_n^{(q)}) | e^{-\theta\tau} \leq e^{-\theta\tau} \left( \frac{n P(A_n^{(q)})}{2k_n} + \frac{\tau^2}{k_n} \right) + o(1) \]
\[ \leq C' e^{-\theta\tau} \frac{\tau^2}{k_n}. \]
Hence, there exists $C'' > 0$, depending on $\varepsilon$ but not on $\tau$, such that
\[ \left| 1 - \left( \frac{n}{k_n} \right) P(A_n^{(q)}) \right|^k_n e^{-\theta\tau} \leq C'' e^{-\theta\tau} \left( \left| \theta\tau - n P(A_n^{(q)}) \right| + \frac{\tau^2}{k_n} \right). \] (2.11)
Finally, by Proposition 2.4 we have
\[ \left| P(M_n \leq u_n) - P(\mathcal{W}_{0,n} \cap T^{n+j}(U_n \setminus A_n^{(q)})) \right| \leq q P(U_n \setminus A_n^{(q)}) \] (2.12)
Note that when $q = 0$ both sides of inequality (2.12) equal 0.
The estimate in Theorem 2.2 follows from joining the estimates in (2.9), (2.10), (2.11) and (2.12). \qed

3. Sharper estimates under stronger assumptions on the system
In this section we make stronger assumptions on our system in order to obtain better error estimates. Firstly we make an important mixing assumption and secondly we ask for a bit more geometric structure around our points of interest.
Take a system $(\mathcal{X}, \mathcal{B}, P, f)$, where $\mathcal{X}$ is a Riemannian manifold, $\mathcal{B}$ is the Borel sigma-algebra, $f : \mathcal{X} \rightarrow \mathcal{X}$ is a measurable map and $P$ an $f$-invariant probability measure.

3.1. Decay of correlations against $L^1$. Our basic assumption will be decay of correlations against $L^1$ observables. The definition of this is as follows. Let $C_1, C_2$ denote Banach
spaces of real valued measurable functions defined on $X$. We denote the correlation of non-zero functions $\phi \in C_1$ and $\psi \in C_2$ w.r.t. a measure $\mathbb{P}$ as

$$\text{Cor}_\mathbb{P}(\phi, \psi, n) := \frac{1}{\|\phi\|_{C_1} \|\psi\|_{C_2}} \left| \int \phi (\psi \circ f^n) \, d\mathbb{P} - \int \phi \, d\mathbb{P} \int \psi \, d\mathbb{P} \right|. $$

We say that we have decay of correlations, w.r.t. the measure $\mathbb{P}$, for observables in $C_1$ against observables in $C_2$ if, for every $\phi \in C_1$ and every $\psi \in C_2$ we have

$$\text{Cor}_\mathbb{P}(\phi, \psi, n) \rightarrow 0, \quad \text{as} \quad n \rightarrow \infty. $$

We say that we have decay of correlations against $L^1$ observables whenever this holds for $C_2 = L^1(\mathbb{P})$ and $\|\psi\|_{C_2} = \|\psi\|_1 = \int |\psi| \, d\mathbb{P}$.

**Remark 3.1.** Examples of systems for which we have decay of correlations against $L^1$ observables include: non-uniformly expanding maps of the interval, like the ones considered by Rychlik in [Ryc83], for which $C_1$ is the space of functions of bounded variation (see Section sec:Rych for more details); and non-uniformly expanding maps in higher dimensions, like the ones studied by Saussol in [Sau00], for which $C_1$ is the space of functions with bounded quasi-Hölder norm.

**Remark 3.2.** We remark that in most situations, decay of correlations against $L^1$ observables is a consequence of the existence of a gap in the spectrum of the map’s corresponding Perron-Frobenius operator. However, in [Dol98] Dolgopyat proves exponential decay of correlations for certain Axiom A flows but along the way he proves it for semiflows against $L^1$ observables. This is done via estimates on families of twisted transfer operators for the Poincaré map, but without considering the Perron-Frobenius operator for the flow itself. This means that the discretisation of this flow by using a time-1 map provides an example of a system with decay of correlations against $L^1$ for which it is not known if there exists a spectral gap of the corresponding Perron-Frobenius operator. As we have said, existence of a spectral gap for the map’s Perron-Frobenius operator, defined in some nice function space, appears to be a stronger property than decay of correlations against $L^1$ observables. However, the latter is still a very strong property. In fact, from decay of correlations against $L^1$ observables, regardless of the rate, as long as it is summable, one can actually show that the system has exponential decay of correlations of Hölder observables against $L^\infty$. (See [AFLV11, Theorem B]).

**Remark 3.3.** The results below assume that the system has decay of correlations against $L^1$ observables. However, we do not need this assumption in full strength since we do not need it to hold for all $L^1$ observables. Namely, we only need that for all $\phi \in C_1$, where $C$ is some Banach space of real valued functions, and all $\psi$ of the form $\psi = 1_A$, for some $A \in \mathcal{B}$, we have that $\gamma(n) := \text{Cor}(\phi, \psi, n)$ is such that $n^2 \gamma(n) \xrightarrow{n \to \infty} 0$, where (and this is a crucial point) the $\| \cdot \|_{C_2}$ appearing in Cor$(\phi, \psi, n)$ is the $L^1(\mathbb{P})$-norm, i.e., $\|\psi\|_{C_2} = \|1_A\|_1 = \mathbb{P}(A)$. 

3.2. Some geometric structure. Suppose that the time series \( X_0, X_1, \ldots \) arises from such a system simply by evaluating a given observable \( \varphi : \mathcal{X} \to \mathbb{R} \cup \{\pm \infty\} \) along the orbits of the system, or in other words, the time evolution given by successive iterations by \( f \):

\[
X_n = \varphi \circ f^n, \quad \text{for each } n \in \mathbb{N}.
\]  

(3.1)

Clearly, \( X_0, X_1, \ldots \) defined in this way is not an independent sequence. However, \( f \)-invariance of \( \mathbb{P} \) guarantees that this stochastic process is stationary.

We suppose that the r.v. \( \varphi : \mathcal{X} \to \mathbb{R} \cup \{\pm \infty\} \) achieves a global maximum at \( \zeta \in \mathcal{X} \) (we allow \( \varphi(\zeta) = +\infty \)). We assume that \( \varphi \) and \( \mathbb{P} \) are sufficiently regular so that:

1. For \( u \) sufficiently close to \( u_F := \varphi(\zeta) \), the event
   
   \[
   U(u) := \{x \in \mathcal{X} : \varphi(x) > u\} = \{X_0 > u\}
   \]
   
   corresponds to a topological ball centred at \( \zeta \). Moreover, the quantity \( \mathbb{P}(U(u)) \), as a function of \( u \), varies continuously on a neighbourhood of \( u_F \).

2. If \( \zeta \in \mathcal{X} \) is a repelling periodic point, of prime period \( p \in \mathbb{N} \), then we have that the periodicity of \( \zeta \) implies that for all large \( u \), \( \{X_0 > u\} \cap f^{-p}(\{X_0 > u\}) \neq \emptyset \) and the fact that the prime period is \( p \) implies that \( \{X_0 > u\} \cap f^{-j}(\{X_0 > u\}) = \emptyset \) for all \( j = 1, \ldots, p - 1 \). Moreover, the fact that \( \zeta \) is repelling means that we have backward contraction implying that there exists \( 0 < \theta < 1 \) such that
   
   \[
   \mathbb{P}\left(\{X_0 > u\} \cap f^{-p}(\{X_0 > u\})\right) \sim (1 - \theta)\mathbb{P}(X_0 > u),
   \]
   
   for all \( u \) sufficiently large.

We are interested in studying the extremal behaviour of the stochastic process \( X_0, X_1, \ldots \) which is tied with the occurrence of exceedances of high levels \( u \). The occurrence of an exceedance at time \( j \in \mathbb{N}_0 \) means that the event \( \{X_j > u\} \) occurs, where \( u \) is close to \( u_F \). Observe that a realisation of the stochastic process \( X_0, X_1, \ldots \) is achieved if we pick, at random and according to the measure \( \mathbb{P} \), a point \( x \in \mathcal{X} \), compute its orbit and evaluate \( \varphi \) along it. Then saying that an exceedance occurs at time \( j \) means that the orbit of the point \( x \) hits the ball \( U(u) \) at time \( j \), i.e., \( f^j(x) \in U(u) \).

3.3. Main results. Next we give sharper error terms for the distributional limit of the partial maximum and of the first hitting time, which have both to do with rare events corresponding to entrances in shrinking balls around a point \( \zeta \). The basic assumption is decay of correlations against \( L^1 \) observables.

These results relate with the main result from [Aba04], where sharp error terms of this type were obtained for the first hitting time of \( \psi \) and \( \phi \) mixing processes, arising from shift dynamics over a finite alphabet, and where the hitting targets were cylinders rather than balls.

In [Kel12], a similar result to that of [Aba04] was obtained using a very powerful technique developed in [KL09], gave essentially the same estimates, in the more general context

\[ i.e., \text{the smallest } n \in \mathbb{N} \text{ such that } f^n(\zeta) = \zeta. \]

Clearly \( f^k(\zeta) = \zeta \) for any \( i \in \mathbb{N} \).
of balls around some $\zeta$ (rather than cylinders), for systems with a spectral gap for the corresponding Perron-Frobenius operator. We comment more on the comparisons between these works and ours in Remark 3.11.

As we have seen, from [FFT10, FFT11], the existence of EVLs and HTS are two sides of the same coin. In here, we present the error terms in these two different contexts. On one hand, our tools are developed on EVL context which makes it natural to write Theorem 3.4 below. However, the normalising sequences on the EVL context are traditionally designed in such a way that they implicitly define a relation between the radii of the target balls around $\zeta$ and the time scale $\tau$. On the other hand, in the HTS context this intrinsic relation does not exist, which leads to complications in the proof, but then allows us to apply our results to obtain estimates for the escape rate in Corollary 3.12; hence, we also state Theorem 3.6.

3.3.1. Improved error terms for EVLs. Next we present our main result of this section in the context of EVLs, which provides sharper error terms for the convergence in distribution of the partial maximum of stochastic processes as defined in (3.1).

**Theorem 3.4.** Assume that the system has decay of correlations of observables in a Banach space $C$ against observables in $L^1$ with rate function $\gamma : \mathbb{N} \to \mathbb{R}$, where $\gamma$ is independent of the given observables and there exists $\delta > 0$ such that $n^{2+\delta} \gamma(n) \to 0$, as $n \to \infty$. Let $u_n$ be as in (1.2). Assume that there exists $q \in \mathbb{N}_0$ such that $q := \min \{ j \in \mathbb{N}_0 : \lim_{n \to \infty} R(A_n^{(j)}) = \infty \}.$ For each $n \in \mathbb{N}$, let $A_n := A_n^{(q)}$, $R_n := R(A_n^{(q)})$, where $R$ is defined as in (1.5), and let $k_n, t_n$ be such that

$$k_n t_n \mathbb{P}(A_n) + \frac{n^2}{k_n} \gamma(t_n) + \frac{(n \mathbb{P}(A_n))^2}{k_n} = \inf_{k, t \in \mathbb{N}, kt < n} \left\{ kl \mathbb{P}(A_n) + \frac{n^2}{k} \gamma(t) + \frac{(n \mathbb{P}(A_n))^2}{k} \right\}.$$ 

Assume that there exists $M > 0$ such that $\| 1_{A_n} \|_C \leq M$ for all $n \in \mathbb{N}$.

Then there exists $C > 0$ such that for all $n \in \mathbb{N}$

$$|\mathbb{P}(M_n \leq u_n) - e^{-\theta \tau}| \leq Ce^{-\theta \tau} \left(|\theta \tau - n \mathbb{P}(A_n)| + k_n t_n \frac{\theta \tau}{n} + \frac{n^2}{k_n} \gamma(t_n) + \frac{(\theta \tau)^2}{k_n} + \theta \tau \sum_{j=R_n}^{\ell_n-1} \gamma(j) \right),$$

where $\ell_n = \lfloor n/k_n \rfloor - t_n$ and the EI $\theta$ is given by equation (2.3).

**Remark 3.5.** Observe that by assumption we must have that

$$k_n t_n \mathbb{P}(A_n) + \frac{n^2}{k_n} \gamma(t_n) + \frac{(n \mathbb{P}(A_n))^2}{k_n} \to 0 \quad \text{as } n \to \infty.$$ 

In particular,

- $k_n, t_n \to \infty$ as $n \to \infty$;
- $k_n t_n = o(n)$. 

Indeed, if we define a priori rates for the above divergences/convergences along with the rate of growth of \( R_n \), then the constant \( C \) in the theorem depends only on those rates, on \( M \) and on a bound on the deviation from the limiting constants in [L2] and [R2]. That is to say that to apply this theorem, we don’t need to use the optimal sequences \((k_n)_{n \in \mathbb{N}}, (t_n)_{n \in \mathbb{N}}\), so long as the conditions above hold. Note that \( C \) can be taken independently of \( \tau \).

For example, if \( n^{2+\delta} \gamma(n) \leq \kappa \) where \( \delta > 0 \), then setting \( k_n = n^{\delta/2} \) and \( t_n = n^{1-\delta} \),

\[
\left| \mathbb{P}(M_n \leq u_n) - e^{-\theta t} \right| \leq C e^{-\theta t} \left( k_n t_n \frac{\tau}{n} + \frac{n^2}{k_n} \gamma(t_n) + \frac{\tau^2}{k_n} + \tau \sum_{j=R_n}^{\infty} \gamma(j) \right)
\]

\[
\leq C e^{-\theta t} \left( n^{-\frac{\delta}{2}} \tau (1 + \kappa + \tau) + \frac{R_n^{-1+\delta} \kappa \tau}{1 + \delta} \right).
\]

For comments on the existence of \( q \) as in the theorem, see Remark 3.7 below.

### 3.3.2. Improved error terms for HTS

Here we prove a result analogous to Theorem 3.4. The key differences are that it applies to HTS rather than EVLs, it is more directly applicable to balls of general diameter, and that it explicitly decouples the time scale \( \tau \) from the radius of the ball \( \epsilon \).

In what follows \( B_{\epsilon}(\zeta) \) denotes the open ball of radius \( \epsilon \), around the point \( \zeta \in X \), w.r.t. a given metric on \( X \). Also set \( A_{\epsilon}^{(0)}(\zeta) := B_{\epsilon}(\zeta) \) and, for each \( q \in \mathbb{N} \), let

\[
A_{\epsilon}^{(q)}(\zeta) := B_{\epsilon}(\zeta) \cap \bigcap_{i=1}^{q} f^{-i}((B_{\epsilon}(\zeta))^c).
\]

**Theorem 3.6.** Assume that the system has decay of correlations of observables in a Banach space \( C \) against observables in \( L^1 \) with rate function \( \gamma : \mathbb{N} \to \mathbb{R} \), where \( \gamma \) is independent of the given observables and there exists \( \delta > 0 \) such that \( n^{2+\delta} \gamma(n) \to 0 \), as \( n \to \infty \). Fix some point \( \zeta \in X \) and assume that there exists \( q \in \mathbb{N}_0 \) such that

\[
q := \min \left\{ j \in \mathbb{N}_0 : \lim_{\epsilon \to 0} R(A_{\epsilon}^{(j)}(\zeta)) = \infty \right\}.
\]

For each \( \epsilon > 0 \), let \( B_{\epsilon} := B_{\epsilon}(\zeta) \), \( A_{\epsilon} := A_{\epsilon}^{(q)}(\zeta) \), \( R_{\epsilon} := R(A_{\epsilon}^{(q)}(\zeta)) \), where \( R \) is defined as in [L5], and \( k_\epsilon, t_\epsilon \in \mathbb{N} \) be such that

\[
k_\epsilon t_\epsilon \mathbb{P}(B_{\epsilon}) + \frac{\mathbb{P}(B_{\epsilon})^2}{k_\epsilon^2} \gamma(t_\epsilon) + \frac{1}{k_\epsilon} = \inf_{k, t \in \mathbb{N} : k t \mathbb{P}(B_{\epsilon}) \leq 1} \left\{ k t \mathbb{P}(B_{\epsilon}) + \frac{\mathbb{P}(B_{\epsilon})^2}{k} \gamma(t) + \frac{1}{k} \right\}.
\]

Let \( \ell_\epsilon = \frac{\mathbb{P}(B_{\epsilon})^{-1} / k_\epsilon}{t_\epsilon} \) and assume that there exists \( M > 0 \) such that \( \|1_{A_{\epsilon}}\|_C \leq M \) for all \( \epsilon > 0 \).

Then there exists \( C > 0 \), depending on \( \epsilon \) but not on \( \tau \), such that for all \( \epsilon > 0 \) and \( \tau > 0 \)

\[
\left| \mathbb{P} \left( \frac{r_{B_{\epsilon}(\zeta)}}{\mathbb{P}(B_{\epsilon})} > \frac{\tau}{\mathbb{P}(B_{\epsilon})} \right) - e^{-\theta \tau} \right| \leq C \left( \tau^2 \alpha_{\epsilon} \Gamma_{\epsilon} + \frac{\tau^2}{k_\epsilon} \Gamma_{\epsilon} + \frac{\tau^3}{k_\epsilon \alpha_{\epsilon} \Gamma_{\epsilon}} \right) e^{-(\theta - k_\epsilon \gamma_{A_{\epsilon}}) \tau},
\]
where \( \Gamma_\varepsilon = \left( k_\varepsilon t_\varepsilon \mathbb{P}(A_\varepsilon) + \frac{\mathbb{P}(B_\varepsilon) - \varepsilon \gamma(t_\varepsilon)}{k_\varepsilon} + \frac{1}{k_\varepsilon} + \sum_{j=R_\varepsilon}^{\gamma(j)} \right), \ \alpha_\varepsilon = |\theta - \frac{\mathbb{P}(A_\varepsilon)}{\mathbb{P}(B_\varepsilon)}| + t_\varepsilon k_\varepsilon \mathbb{P}(A_\varepsilon), \ \Upsilon_{A_\varepsilon} \) is given as in Lemma 4.3 and is such that \( k_\varepsilon \Upsilon_{A_\varepsilon} \leq C \Gamma_\varepsilon, \) for some \( C > 0 \) (see (6.3)) and

\[
\theta = \lim_{\varepsilon \to 0} \frac{\mathbb{P}(A_\varepsilon)}{\mathbb{P}(B_\varepsilon)}.
\]

**Remark 3.7.** If \( f \) is continuous, for example, and \( \mathbb{P} \) is sufficiently regular so that the content of condition \([R1]\) holds for balls around all \( \zeta \in \mathcal{X} \), then if \( \zeta \) is not periodic then Theorem 3.6 holds with \( q = 0 \) and \( \theta = 1 \). Moreover, if \( \zeta \) is a repelling periodic point of prime period \( p \), in the sense described in condition \([R2]\), then Theorem 3.6 also applies with \( q = p \) and \( \theta \) is given as in \([R2]\) by the backward contraction rate at \( \zeta \).

**Remark 3.8.** Theorem 3.6 can also be applied to discontinuity points \( \zeta \) of \( f \) as considered in [AFV12, Section 3.3] so that, if for example, there exist \( p^- \) and \( p^+ \) such that \( f^{p^+}(\zeta^+) = \zeta = f^{p^-}(\zeta^-) \) (see [AFV12, Section 3.3] for details) then \( q = \max\{p^-, p^+\} \) and \( \theta \) is given by the formulas in [AFV12, Proposition 3.4].

**Remark 3.9.** Observe that by assumption,

\[
k_\varepsilon t_\varepsilon \mathbb{P}(A_\varepsilon) + ((\mathbb{P}(A_\varepsilon))^2/k_\varepsilon) \gamma(t_\varepsilon) + 1/k_\varepsilon \to 0, \quad \text{as } \varepsilon \to 0.
\]

In particular,

- \( k_\varepsilon, t_\varepsilon \to \infty \) as \( \varepsilon \to 0 \);
- \( k_\varepsilon t_\varepsilon = o(\mathbb{P}(B_\varepsilon)^{-1}) \).

So as in Remark 3.5, we can fix growth rates on the parameters in this theorem to show that the constant \( C \) depends only on those rates.

**Remark 3.10.** Observe that the error terms are dominated by \( \Gamma_\varepsilon \) and to some extent by \( \alpha_\varepsilon \). There are four terms in \( \Gamma_\varepsilon \). The first and the third terms of \( \Gamma_\varepsilon \) result from the fact that we use a blocking argument and essentially take into account the balance between the size of \( \mathbb{P}(A_\varepsilon) \), the number of blocks and the size of the gaps between the blocks. As seen in Remark 3.5, this ultimately leads to an error that typically goes down as \( \mathbb{P}(A_\varepsilon)^{\delta} \) for some \( 0 < \delta < 1 \). For the second term we need to balance the size of gaps between the blocks and the loss of memory of the system. Since, as observed above, decay of correlations against \( L^1 \) is a strong mixing property, then typically \( \gamma \) decays exponentially fast which means that this term is usually negligible when compared to the others. The fourth term in \( \Gamma_\varepsilon \) accounts for the effect of fast recurrence from \( A_\varepsilon \) to \( A_\varepsilon \) itself. So if \( \gamma \) decays exponentially fast, then the fourth term should decay as \( \gamma(R_\varepsilon) \). The quantity \( R_\varepsilon \) has been studied in [ACS00, AL13, AV08, HV10, STV02], for example, and we generally expect it to behave like \( -\log(\mathbb{P}(A_\varepsilon)) \), which means that the fourth term should also decay like a power of \( \mathbb{P}(A_\varepsilon) \). As a consequence the behaviour of \( \Gamma_\varepsilon \) should ruled by a trade-off between the first, the third and the fourth terms of \( \Gamma_\varepsilon \).

Regarding \( \alpha_\varepsilon \), note that for a sufficiently regular dynamical system \( f \) and invariant measure \( \mathbb{P} \), the dominant term should be again the first term of \( \Gamma_\varepsilon \).
Remark 3.11. We note that, when compared with the sharp estimates in [Aba04, Kel12] in terms of $\tau$ we have a loss here from $\tau e^{-\xi A_1 \tau}$ in [Aba04] (or $\tau e^{-\xi \tau}$ in [Kel12]), to $\tau^3 e^{-(\theta - k_2 Y_1) \tau}$, which is explained by the fact that we compute the error terms with respect to the asymptotic limit $e^{-\theta \tau}$ with no correcting factors such as $\xi A_1$, used in [Aba04], and $\xi$, in [Kel12]. Observe that even in the ideal i.i.d. case, as noted in Remark 2.3 when the error terms are computed with respect to the asymptotic limit ($e^{-\tau}$, in this case) then an error term $\tau^2 e^{-\tau}$ already appears (see [LLR83, Theorem 2.4.2]). Moreover, the deeper analysis we perform here explains how $\theta - k_2 Y_1$ goes to zero, as $\varepsilon \to \infty$, which, as we have seen, depends on the fast recurrence of the point $\zeta$ to itself.

3.4. Escape rates in the zero-hole limit. One way of studying the recurrence properties of a system is to fix a hole $B_{\varepsilon}(\zeta)$ around some chosen point $\zeta$ and compute the rate of escape of mass through the hole, i.e., find the limit, if it exists,

$$- \lim_{\tau \to \infty} \frac{1}{\tau} \log \mathbb{P} \left( r_{B_{\varepsilon}(\zeta)} > \tau \right).$$

Moreover, one can consider, as in [KL09, FP12], what happens when the size of the ball goes to zero too. In those papers, it is shown that we should expect this to scale with $\mathbb{P}(B_{\varepsilon}(\zeta))$. Theorem 3.6 gives error estimates independently of the time scale $\tau$, which yields the following corollary.

Corollary 3.12. Suppose that the system is as in Theorem 3.6. Then

$$- \lim_{\varepsilon \to 0} \frac{1}{\mathbb{P}(B_{\varepsilon}(\zeta))} \limsup_{\tau} \frac{1}{\tau} \log \mathbb{P} \left( r_{B_{\varepsilon}(\zeta)} > \frac{\tau}{\mathbb{P}(B_{\varepsilon})} \right) \geq \theta.$$

Note that for many dynamical systems it is known that the upper limit is also $\theta$, see for example the systems in the papers referenced above, but we were unable to prove this in the generality of the setting above.

Proof. In Theorem 3.6 the error estimate is

$$\left| \mathbb{P} \left( r_{B_{\varepsilon}(\zeta)} > \frac{\tau}{\mathbb{P}(B_{\varepsilon})} \right) - e^{-\theta \tau} \right| \leq C \left( \tau^2 \alpha_\varepsilon \Gamma_\varepsilon + \frac{\tau^2}{k_\varepsilon} \Gamma_\varepsilon + \frac{\tau^3}{k_\varepsilon^2} \alpha_\varepsilon \Gamma_\varepsilon \right) e^{-(\theta - k_2 Y_1) \tau},$$

where $\Gamma_\varepsilon = \left( k_\varepsilon t_\varepsilon \mathbb{P}(A_\varepsilon) + \frac{\mathbb{P}(B_{\varepsilon})^{-2} \gamma(t_\varepsilon)}{k_\varepsilon} + \frac{1}{k_\varepsilon} + \sum_{j=1}^{\ell_\varepsilon-1} \gamma(j) \right)$, $\alpha_\varepsilon = \left| \theta - \frac{\mathbb{P}(A_\varepsilon)}{\mathbb{P}(B_{\varepsilon})} + t_\varepsilon k_\varepsilon \mathbb{P}(A_\varepsilon) \right|$ and

$$\theta = \lim_{\varepsilon \to 0} \frac{\mathbb{P}(A_\varepsilon)}{\mathbb{P}(B_{\varepsilon}(\zeta))}.$$

Since $n^{2+\delta} \gamma(n) \to 0$ as $n \to \infty$, for $\kappa > 0$, for all large enough $n$, we have $n^{2+\delta} \gamma(n) \leq \kappa$. Then setting $k_\varepsilon = \frac{1}{\mathbb{P}(B_{\varepsilon})^{\delta}}$ and $t_\varepsilon = \frac{1}{\mathbb{P}(B_{\varepsilon})^{-\gamma}}$, we deduce that

$$\Gamma_\varepsilon \leq \frac{1}{\mathbb{P}(B_{\varepsilon})^{\delta}} \left( \frac{\mathbb{P}(A_\varepsilon)}{\mathbb{P}(B_{\varepsilon})} + \kappa + 1 \right).$$
and
\[ \alpha_\varepsilon \leq \left| \theta - \frac{\mathbb{P}(A_\varepsilon)}{\mathbb{P}(B_\varepsilon)} \right| + \frac{1}{\mathbb{P}(B_\varepsilon)^2} \mathbb{P}(A_\varepsilon). \]

In particular,
\[ \limsup_{\tau} \frac{1}{\tau} \log \mathbb{P} \left( r_{B_\varepsilon(\zeta)} > \frac{\tau}{\mathbb{P}(B_\varepsilon)} \right) \leq -\left( \theta - k_\varepsilon \Upsilon A_\varepsilon \right) \mathbb{P}(A_\varepsilon). \]

Therefore
\[ -\lim_{\varepsilon \to 0} \frac{1}{\mathbb{P}(B_\varepsilon(\zeta))} \limsup_{\tau} \frac{1}{\tau} \log \mathbb{P} \left( r_{B_\varepsilon(\zeta)} > \frac{\tau}{\mathbb{P}(B_\varepsilon)} \right) \geq \theta, \]
as required.

4. Preparatory Lemmas

In this section, we give the estimates necessary to prove Theorems 3.4 and 3.6. As in previous such arguments, the core of the strategy is a blocking argument as described just before Lemma 2.5. The estimates here are enhanced versions of Lemmas 2.5, 2.6 and Proposition 2.7, obtained using the stronger assumption given by the decay of correlations against \( L^1 \) observables.

Next we give an estimate on the error resulting from neglecting the random variables in the gaps between the blocks.

**Lemma 4.1.** Assume that the systems has decay of correlations against \( L^1 \) with a rate function \( \gamma : \mathbb{N} \to \mathbb{R} \), independent of the observables \( \phi, \psi \) such that \( \gamma(n) \to 0 \), as \( n \to \infty \). For any fixed \( A \in \mathcal{B} \) and integers \( s, t, m \), with \( t < m \), we have:
\[ \left| \mathbb{P}(\mathcal{W}_{0,s+t+m}(A)) - \mathbb{P}(\mathcal{W}_{0,s}(A) \cap \mathcal{W}_{s+t,m}(A)) \right| \leq t [ \mathbb{P}(A) + \|1_A\|_{\mathcal{C}} \gamma(t)] \mathbb{P}(\mathcal{W}_{0,m-t}(A)). \]

**Proof.** Using stationarity and the decay of correlations against \( L^1 \), with \( \phi = 1_A \) and \( \psi = 1_{\mathcal{W}_{s+2t,m-t}} \), we have
\[ \mathbb{P}(\mathcal{W}_{0,s} \cap \mathcal{W}_{s+t,m}) - \mathbb{P}(\mathcal{W}_{0,s+t+m}) = \mathbb{P}(\mathcal{W}_{0,s} \cap \mathcal{W}_{s+t,m}) \leq \mathbb{P}(\mathcal{W}_{0,t} \cap \mathcal{W}_{t,m}) \leq \sum_{j=0}^{t-1} \mathbb{P}(f^{-j}(A) \cap \mathcal{W}_{2t,m-t}) \leq \sum_{j=0}^{t-1} [ \mathbb{P}(A) + \|1_A\|_{\mathcal{C}} \gamma(t)] \mathbb{P}(\mathcal{W}_{0,m-t}), \]
and the result follows immediately.

In the result that follows we give an estimate for the probability of not visiting the set \( A \) within a block, which will be used later in the recursive argument mounted in Lemma 4.3.
Lemma 4.2. Assume that the systems has decay of correlations of observables $\phi$, in a Banach space $C$, against $\psi \in L^1$, with a rate function $\gamma : \mathbb{N} \to \mathbb{R}$, independent of the observables $\phi, \psi$ such that $\gamma(n) \to 0$, as $n \to \infty$. Then for any fixed $A \in B$ and integers $s, t, m$, we have:

$$
\left| \mathbb{P}(\mathcal{W}_{s,t}(A) \cap \mathcal{W}_{s+m}(A)) - \mathbb{P}(\mathcal{W}_{s,t}(A))(1 - s\mathbb{P}(A)) \right| \leq \mathbb{P}(\mathcal{W}_{s,t}(A)) \left\{ 1_A \|c s \gamma(t) \right\} + (s - R(A))^2 \left[ (\mathbb{P}(A))^2 + 1_A \|c \gamma(t) \right] + 1_A \|c (s - R(A)) \mathbb{P}(A) + 1_A \|c \gamma(t) \sum_{q=R(A)}^{s-1} \gamma(q) \right\}.
$$

Proof. Observe that

$$
\left| \mathbb{P}(\mathcal{W}_{s,t}(A) \cap \mathcal{W}_{s+t}(A)) - \mathbb{P}(\mathcal{W}_{s,t}(A))(1 - s\mathbb{P}(A)) \right| \leq \mathbb{P}(A)\mathbb{P}(\mathcal{W}_{s,t}(A)) - \sum_{j=0}^{s-1} \mathbb{P}(A \cap \mathcal{W}_{s+t-j}(A)) \right| + \left| \mathbb{P}(\mathcal{W}_{s,t}(A) \cap \mathcal{W}_{s+t}(A)) - \mathbb{P}(\mathcal{W}_{s,t}(A)) - \sum_{j=0}^{s-1} \mathbb{P}(A \cap \mathcal{W}_{s+t-j}(A)) \right| . \tag{4.1}
$$

Using stationarity and decay of correlations against $L^1$, it follows immediately that for the first term on right:

$$
\left| s\mathbb{P}(A)\mathbb{P}(\mathcal{W}_{s,t}(A)) - \sum_{j=0}^{s-1} \mathbb{P}(A \cap \mathcal{W}_{s+t-j}(A)) \right| \leq s\gamma(t) 1_A \|c \|1_{\mathcal{W}_{s,t}} \|_1 = s\gamma(t) 1_A \|c \mathbb{P}(\mathcal{W}_{s,t}). \tag{4.2}
$$

For the second term on the right we have

$$
\mathbb{P}(\mathcal{W}_{s,t}(A) \cap \mathcal{W}_{s+t}(A)) = \mathbb{P}(\mathcal{W}_{s+t}(A)) - \mathbb{P}(\mathcal{W}_{s,t}(A) \cap \mathcal{W}_{s+t}(A)).
$$

Moreover the fact that $\mathcal{W}_{s,t} \cap \mathcal{W}_{s+t} = \bigcup_{i=0}^{s-1} f^{-i}(A) \cap \mathcal{W}_{s+t}$ implies

$$
0 \leq \left( \sum_{j=0}^{s-1} \mathbb{P}(A \cap \mathcal{W}_{s+t-j}(A)) \right) - \mathbb{P}(\mathcal{W}_{s,t} \cap \mathcal{W}_{s+t}) \leq \sum_{j=0}^{s-1} \sum_{i=j}^{s-1} \mathbb{P}((f^{-i-j}(A) \cap f^{-i}(A) \cap \mathcal{W}_{s+t}).
$$

Hence, by definition of $R(A)$, stationarity and decay of correlations against $L^1$ (with $\phi = 1_A$ and $\psi = 1_{A \cap \mathcal{W}_{s+t-i}}$) we have

$$
\left| \mathbb{P}(\mathcal{W}_{s,t}(A) \cap \mathcal{W}_{s+t}(A)) - \mathbb{P}(\mathcal{W}_{s,t}(A)) + \sum_{j=0}^{s-1} \mathbb{P}(A \cap \mathcal{W}_{s+t-j}(A)) \right| \leq \sum_{j=0}^{s-1} \sum_{i=j}^{s-1} \mathbb{P}(A \cap f^{-i-j}(A) \cap \mathcal{W}_{s+t-j}(A)) \leq \sum_{j=0}^{s-1} \sum_{i=j}^{s-1} \mathbb{P}(A \cap \mathcal{W}_{s+t-i}(A)) + \sum_{j=0}^{s-1} \sum_{i=j}^{s-1} \|1_A \|c \gamma(i-j) \mathbb{P}(A \cap \mathcal{W}_{s+t-i}(A)) =: I + II.
$$
Using stationarity, decay of correlations against $L^1$ (with $\phi = 1_A$ and $\psi = 1_{W_{0,m}}$), the fact that $s + t - i > t$ and $\gamma$ is non-increasing, we have

$$I \leq \sum_{j=0}^{s-1} \sum_{i=j+R(A)}^{s-1} \mathbb{P}(A) \left[ \mathbb{P}(A) + \|1_A\|c\gamma(t) \right] \mathbb{P}(W_{0,m}) \leq \mathbb{P}(W_{0,m})(s - R(A))^2 \left[ (\mathbb{P}(A))^2 + \|1_A\|c\gamma(t) \right].$$

and

$$II \leq \sum_{j=0}^{s-1} \sum_{i=j+R(A)}^{s-1} \|1_A\|c \left[ \mathbb{P}(A) + \|1_A\|c\gamma(t) \right] \mathbb{P}(W_{0,m}) \leq \mathbb{P}(W_{0,m}) \|1_A\|c \mathbb{P}(A) + \|1_A\|c\gamma(t) \sum_{q=R(A)}^{s-1} (s - q)\gamma(q) \leq \mathbb{P}(W_{0,m}) \|1_A\|c(s - R(A)) \left[ \mathbb{P}(A) + \|1_A\|c\gamma(t) \right] \sum_{q=R(A)}^{s-1} \gamma(q).$$

The result now follows from plugging (4.2) and the estimates on $I$ and $II$ into (4.1). \hfill \Box

The following result is contains the main estimate that will be used throughout the rest of the paper.

Lemma 4.3. Assume that the system has decay of correlations of observables in a Banach space $C$ against observables in $L^1$ with rate function $\gamma : \mathbb{N} \to \mathbb{R}$, where $\gamma$ is independent of the given observables and $\gamma(n) \to 0$, as $n \to \infty$. Fix $A \in \mathcal{B}$ and $n \in \mathbb{N}$. Let $\ell, k, b \in \mathbb{N}$ be such that $n = k\lfloor n/k \rfloor + b$ and $\lfloor n/k \rfloor \mathbb{P}(A) < 2$. Also, consider an integer $t$ such that $t < \lfloor n/k \rfloor$, and set $\ell = \lfloor n/k \rfloor - t$, $L := 1 - \ell\mathbb{P}(A)$. Also, let

$$t_{A,t} := \sup_{i \in \mathbb{N}} \frac{\mathbb{P}(W_{0,i(t+\ell+1)-2\ell})}{\mathbb{P}(W_{0,i(t+\ell+1)})}$$

and

$$\Upsilon_A := t(\mathbb{P}(A) + \|1_A\|c\gamma(t))[1 + \ell_{A,t}t(\mathbb{P}(A) + \|1_A\|c\gamma(t))] + \|1_A\|c\ell\gamma(t) + (\ell - R(A))^2 \left[ (\mathbb{P}(A))^2 + \|1_A\|c\gamma(t) \right] + \|1_A\|c(\ell - R(A)) \left[ \mathbb{P}(A) + \|1_A\|c\gamma(t) \right] \sum_{q=R(A)}^{\ell-1} \gamma(q).$$

We have

$$|\mathbb{P}(W_{0,n}(A)) - L^k| \leq k\Upsilon_A(L + \Upsilon_A)^{k-1}(1 + L + \Upsilon_A).$$

Proof. The basic idea is to split the time interval $[0, n]$ into $k$ blocks of size $\lfloor n/k \rfloor$. Then, using Lemma 4.2 we simply disregard the last $t$ observations of each block, which will create a time gap between the blocks that will allow us to apply Lemma 4.2. This argument is used recursively until we exhaust all the blocks.
Using Lemma 4.1 with \( s = 0 \), we begin by making the following approximation:

\[
\left| P(W_{0,n}) - P(W_{0,k(\ell + t)}) \right| \leq b[P(A) + \|1_A\|c\gamma(t)]P(W_{0,k(\ell + t)-t}). \tag{4.3}
\]

Next, we turn to the main recursive relation which is (recall that \( n/k = \ell + t \)):

\[
\left| P(W_{0,i(\ell + t)}) - LP(W_{0,(i-1)(\ell + t)}) \right| \leq \gamma_A P(W_{0,(i-1)(\ell + t)}), \quad \text{for every} \ i = 2, \ldots, k, \tag{4.4}
\]

In order to prove (4.4) we note first that, by Lemma 4.1 with \( s = 0 \), we have, for all \( i \in \mathbb{N} \):

\[
P(W_{0,i(\ell + t)-t}) - P(W_{0,i(\ell + t)}) \leq t[P(A) + \|1_A\|c\gamma(t)]P(W_{0,i(\ell + t)-2t}),
\]

which allows us to write

\[
P(W_{0,i(\ell + t)-t}) \leq P(W_{0,i(\ell + t)}) [1 + t[P(A) + \|1_A\|c\gamma(t)]t_A], \tag{4.5}
\]

Inequality (4.4) follows now easily from Lemmata 4.1 4.2 and (4.5). In fact,

\[
\left| P(W_{0,i(\ell + t)}) - (1 - \ell P(A))P(W_{0,(i-1)(\ell + t)}) \right| \leq \left| P(W_{0,i(\ell + t)}) - P(W_{0,\ell t}w_{(\ell + t),(i-1)(\ell + t)}) \right| \\
+ \left| P(W_{0,\ell t}w_{(\ell + t),(i-1)(\ell + t)}) - (1 - \ell P(A))P(W_{0,(i-1)(\ell + t)}) \right| \\
\leq t\left[ P(A) + \|1_A\|c\gamma(t) \right]P(W_{0,(i-1)(\ell + t)-t}) + \|1_A\|c\ell\gamma(t)[P(W_{0,(i-1)(\ell + t)}) + (\ell - R(A))^2 [P(A)+\|1_A\|c\gamma(t)]P(W_{0,(i-1)(\ell + t)}) \\
+ \|1_A\|c(s - R(A)) [P(A)+\|1_A\|c\gamma(t)]P(W_{0,(i-1)(\ell + t)}) \right] [s-1] \sum_{q=R(A)}^\gamma(q) \\
\leq \gamma_A P(W_{0,(i-1)(\ell + t)}),
\]

Since \( \ell P(A) < 2 \), then it is clear that \( |L| < 1 \). Also, note that \( |P(W_{0,\ell t}) - L| \leq \gamma_A \). Now, we use (4.4) recursively to estimate \( |P(W_{0,k(\ell + t)}) - L^k| \). In fact, setting \( W_{0,0} := \mathcal{X} \), we have

\[
|P(W_{0,k(\ell + t)}) - L^k| \leq \sum_{i=0}^{k-1} L^{k-1-i} |P(W_{0,(i+1)(\ell + t)}) - L P(W_{0,i(\ell + t)})| \\
\leq \sum_{i=0}^{k-1} L^{k-1-i} \gamma_A P(W_{0,i(\ell + t)}). \tag{4.6}
\]

Based on (4.6) we will show that

\[
P(W_{0,i(\ell + t)}) \leq (L + \gamma)^i, \quad \text{for all} \ i = 0, 1, \ldots \tag{4.7}
\]
It is easy to check that $P(\mathcal{W}_{0,(\ell+t)}) = L + (P(\mathcal{W}_{0,(\ell+t)}) - L) \leq L + \Upsilon_A$. Now, assume that (4.7) holds for all $i = 1$ up to $i = s$. Then, using (4.6), we have

$$P(\mathcal{W}_{0,(s+1)(\ell+t)}) = L^{s+1} + |P(\mathcal{W}_{0,(s+1)(\ell+t)}) - L^{s+1}| \leq L^{s+1} + \sum_{i=0}^{s} L^{s-i} \Upsilon_A P(\mathcal{W}_{0,i(\ell+t)})$$

$$\leq L^s(L + \Upsilon_A) + \sum_{i=1}^{s} L^{s-i} \Upsilon_A (L + \Upsilon_A)^{i} \quad \text{by inductive assumption}$$

$$= (L + \Upsilon)^{s+1}.$$

Plugging (4.7) into (4.6), we get

$$|P(\mathcal{W}_{0,k(\ell+t)}) - L^k| \leq \sum_{i=0}^{k-1} L^{k-1-i} \Upsilon_A (L + \Upsilon_A)^{i} = \sum_{i=0}^{k-1} \sum_{j=0}^{i} \binom{i}{j} L^{k-1-j} \Upsilon_A^{j+1}$$

$$= \sum_{j=0}^{k-1} L^{k-1-j} \Upsilon_A^{j+1} \sum_{i=j}^{k-1} \binom{i}{j} = \sum_{j=0}^{k} L^{k-1-j} \Upsilon_A^{j+1} \binom{k}{j+1} = \sum_{q=1}^{k} L^{k-q} \Upsilon_A^{q} \binom{k}{q}$$

$$= (L + \Upsilon_A)^{k} - L^k \leq k \Upsilon_A (L + \Upsilon_A)^{k-1}, \quad (4.8)$$

where the last inequality follows by the mean value theorem. In the same way, noting that $b \leq k$ and plugging (4.5) and (4.7) into (4.3), we get that

$$|P(\mathcal{W}_{0,n}) - P(\mathcal{W}_{0,k(\ell+t)})| \leq k \Upsilon_A P(\mathcal{W}_{0,k(\ell+t)}) \leq k \Upsilon_A (L + \Upsilon)^k.$$

Hence $|P(\mathcal{W}_{0,n}) - L^k| \leq k \Upsilon A (L + \Upsilon_A)^{k-1}(1 + L + \Upsilon_A)$. \hfill \Box

**Corollary 4.4.** Under the same assumptions of Lemma 4.3 if $k \Upsilon A < L/2$ then

$$|P(\mathcal{W}_{0,n}(A)) - L^k| \leq 5k \Upsilon_A L^{k-1}.$$

**Proof.** Under the assumption that $k > 2$ and $k \Upsilon < L/2$, we have

$$|P(\mathcal{W}_{0,k(\ell+t)}) - L^k| \leq k \Upsilon A (L + \Upsilon_A)^{k-1}(1 + L + \Upsilon_A)$$

$$\leq k \Upsilon A L^{k-1} \left(1 + \frac{1}{2k}\right)^{k-1} (1 + L + \Upsilon_A) \leq 5k \Upsilon A L^{k-1}. \quad \Box$$

The following results will be needed for the sharper estimates in the HTS setting.

**Corollary 4.5.** Under the same assumptions of Lemma 4.3, let $\tau > 0$, note that $\tau n = \tau k(\ell + t) + \tau b$ and set $\beta = \tau k - \lfloor \tau k \rfloor$. Then, when $\lfloor \tau k \rfloor > 0$ we may write:

$$|P(\mathcal{W}_{0,\tau n}) - L^{\lfloor \tau k \rfloor}| \leq (3 + \Upsilon_A) \tau k \Upsilon A (L + \Upsilon_A)^{\lfloor \tau k \rfloor} - 1.$$
In the case $\tau k = 0$, we have
\[ |P(W_{0,\tau n}) - (1 - \tau k(\ell + t)P(A))| \leq \tau k \gamma_A. \]

**Proof.** We start first with case $\tau k > 0$. The idea is to break the time interval $[0, \tau n]$ into $\tau k$ blocks of size $\ell + t$ plus one block of size $\beta(\ell + t)$ plus one last block of size $\tau b$ and then argue as in Lemma 4.3.

Using Lemma 4.1 with $s = 0$, (4.5) and (4.7) we have:
\[ |P(W_{0,\tau n}) - P(W_{0,\tau k(\ell + t)})| \leq |\tau b||P(A) + \|1_A||c\gamma(t)||P(W_{0,\tau k(\ell + t)})| \leq \tau k \gamma_A(L + \gamma_A)^{\tau k}. \tag{4.9} \]

In order to apply the estimates in Lemma 4.3 we need that all blocks have a length larger than $t$. Hence, we consider two cases. If $\beta(\ell + t) \leq t$ then using Lemma 4.1 (4.5) and (4.7) again, we have:
\[ |P(W_{0,\tau k(\ell + t)}) - P(W_{0,\tau k(\ell + t)})| \leq t[P(A) + \|1_A||c\gamma(t)||P(W_{0,\tau k(\ell + t)} - t) \leq \gamma_A(L + \gamma_A)^{\tau k}. \]

Using (4.8) and the estimates just above we obtain:
\[ |P(W_{0,\tau n}) - L^{\tau k}| \leq (3 + \gamma_A)\tau k \gamma_A (L + \gamma_A)^{\tau k}. \]

If $\beta(\ell + t) > t$, then proceeding as in (4.6), it follows that
\[ |P(W_{0,\tau k(\ell + t)}) - L^{\tau k}(1 - \beta tP(A))| \leq \sum_{i=1}^{[\tau k]} L^{[\tau k] - 1} |P(W_{0,(i+1)\beta(\ell + t)}) - P(W_{0,(i+1)\beta(\ell + t)})| + L^{[\tau k]} |P(W_{0,\beta(\ell + t)} - (1 - \beta(\ell + t)P(A))| \]
\[ \leq \sum_{i=1}^{[\tau k]} L^{[\tau k] - 1} \gamma_A \|P(W_{0,(i+1)\beta(\ell + t)}) + L^{[\tau k]} - 1 \gamma_A + L^{[\tau k]} \beta \gamma_A. \]

Since $P(W_{0,(i+1)\beta(\ell + t)}) \leq P(W_{0,i\beta(\ell + t)})$, using (4.7) as in (4.8), we get
\[ |P(W_{0,\tau k(\ell + t)}) - L^{\tau k}(1 - \beta(\ell + t)P(A))| \leq [\tau k]\gamma_A(L + \gamma_A)^{\tau k} - 1 + \beta \gamma_A L^{\tau k} \]
\[ \leq \tau k \gamma_A(L + \gamma_A)^{\tau k} - 1. \]

Using the estimate above and (4.9) we have
\[ |P(W_{0,\tau n}) - L^{\tau k}| \leq \beta(\ell + t)P(A)L^{\tau k} + \tau k \gamma_A(L + \gamma_A)^{\tau k} - 1 + L + \gamma_A \]
\[ \leq (3 + \gamma_A)\tau k \gamma_A (L + \gamma_A)^{\tau k}. \]

The case $[\tau k > 0$ follows easily from (4.9) and the fact that
\[ |P(W_{0,\tau k(\ell + t)}) - (1 - \tau k(\ell + t)P(A))| \leq \tau k \gamma_A, \]

which can be easily derived from Lemma 4.2. 

$\square$
Corollary 4.6. Under the same assumptions of Lemma 4.3, fix $A, B \in \mathcal{B}$ and $n \in \mathbb{N}$. Also let $\ell, k, b, t \in \mathbb{N}$ be again as in Lemma 4.3 and take $j \geq (\lceil \tau k \rceil - 1)(\ell + t) + t$. Then we have:

$$
P(\mathcal{W}_{0,\tau n}(A) \cap f^{-j}(B)) \leq (L + T_A)^{\lceil \tau k \rceil - 1}P(B).
$$

Proof. Since $P(\mathcal{W}_{0,\tau n} \cap f^{-j}(B)) \leq P(\mathcal{W}_{0,(\lceil \tau k \rceil - 1)(\ell + t)} \cap f^{-j}(B))$, we only need to provide an upper bound for the latter.

Now, observe that, in Lemma 4.2, we may replace $\mathcal{W}_{0,m}$ by some event $B$ and $\mathcal{W}_{s+t,m}$ by $f^{-(s+t)}(B)$ and the statement remains true. Using this observation and Lemma 4.4 as for (4.7) in the proof of Lemma 4.3 we obtain that

$$
P(\mathcal{W}_{0,(\lceil \tau k \rceil - 1)(\ell + t)} \cap f^{-j}(B)) \leq (L + T_A)^{\lceil \tau k \rceil - 1}P(B).
$$

\[\square\]

5. Improved error estimates for EVLs

Proof of Theorem 5.1. Let $L_n = (1 - \ell_n P(A_n))$. First we consider the case in which $\zeta$ is not a periodic point. We have

$$
|P(M_n \leq u_n) - e^{-\theta \tau}| \leq |P(M_n \leq u_n) - P(\mathcal{W}_{0,n}(A_n))| + |P(\mathcal{W}_{0,n}(A_n)) - L_n^k|
$$

$$
+ \left|L_n^k - e^{-\theta \tau}\right|,
$$

(5.1)

For the third term on the right of previous equation we start by noting that, by (2.8), we have that there exists $C$ such that

$$
|e^{-k_n \ell_n P(A_n)} - e^{-\theta \tau}| \leq e^{-\theta \tau} |P(\mathcal{W}_{0,n}(A_n))| + o\left(|\theta \tau - k_n \ell_n P(A_n)|\right)
$$

$$
\leq C e^{-\theta \tau} \left(\theta \tau - n P(A_n) + k_n \ell_n P(A_n)\right) = C e^{-\theta \tau} \alpha_n,
$$

(5.2)

where $\alpha_n := |\theta \tau - n P(A_n) + k_n \ell_n P(A_n)|$. Using (2.7) and (5.2), we have that there exist $C', C'' > 0$ such that

$$
\left|(1 - \ell_n P(A_n))^k_n - e^{-k_n \ell_n P(A_n)}\right| = e^{-k_n \ell_n P(A_n)} \left(\frac{(k_n \ell_n P(A_n))^2}{2k_n} + o\left(\frac{1}{k_n}\right)\right)
$$

$$
\leq C' e^{-k_n \ell_n P(A_n)} \left(\frac{\theta \tau}{k_n}\right)^2
$$

$$
\leq C'' e^{-\theta \tau} \left(\frac{\theta \tau}{k_n} + \frac{\theta \tau}{k_n} \alpha_n\right).
$$

Hence, there exists $C_1 > 0$, depending on $n$, such that

$$
\left|L_n^k - e^{-\theta \tau}\right| \leq C_1 e^{-\theta \tau} \left(\alpha_n + \left(\frac{\theta \tau}{k_n}\right)^2\right).
$$

(5.3)

Observe that for all positive integer $i$, we have $P(\mathcal{W}_{0,i(\ell_n - t_n)} \sim L_n^{i-2} (1 - (\ell_n - t_n) P(A_n))^2$ and $P(\mathcal{W}_{0,i(\ell_n + t_n)}) \sim L_n^i$. Since $t_n = o(\ell_n)$, $\|1_{A_n}\| c \leq M$, $t_n P(A_n) \to 0$ and $t_n \gamma(t_n)$ as $n \to \infty$, then for $n$ large enough we have $1 + \ell_{A_n} t_n \|P(A_n)\| + \|1_{A_n}\| c \gamma(t_n) < 2$. 
Recall that, by definition, $R_n = R(A_n) \to \infty$, as $n \to \infty$. This implies that $\sum_{j=R_n}^{\infty} \gamma(j) \to 0$, as $n \to \infty$. Consequently we may write for $n$ sufficiently large:

$$\mathcal{Y}_n := \mathcal{Y}_{A_n} \leq 2t_n(\mathbb{P}(A_n) + M\gamma(t_n)) + M\ell_n \gamma(t_n) + \ell_n^2(\mathbb{P}(A_n)^2 + M\gamma(t_n)) + M^2(\ell_n\mathbb{P}(A_n) + \ell_n \gamma(t_n)) \sum_{j=R_n}^{\ell_n-1} \gamma(j)$$

$$\leq 2t_n\mathbb{P}(A_n) + 4M\ell_n^2\gamma(t_n) + \ell_n^2\mathbb{P}(A_n)^2 + M^2\ell_n\mathbb{P}(A_n) \sum_{j=R_n}^{\ell_n-1} \gamma(j).$$

Hence, there exists $C > 0$ such that for all $n$ sufficiently large we have

$$k_n\mathcal{Y}_n \leq C \left( k_n t_n \mathbb{P}(A_n) + \frac{n^2 \gamma(t_n)}{k_n} + \frac{(n\mathbb{P}(A_n))^2}{k_n} + n\mathbb{P}(A_n) \sum_{j=R_n}^{\ell_n-1} \gamma(j) \right). \quad (5.4)$$

By the properties of the sequences $(k_n)_{n \in \mathbb{N}}, (t_n)_{n \in \mathbb{N}}$, we have that $k_n \mathcal{Y}_n \to 0$, as $n \to \infty$. Moreover, since $L_n = (1 - \ell_n\mathbb{P}(A_n)) \to 1$, as $n \to \infty$ then it is clear that, for $n$ sufficiently large all assumptions of Corollary 4.4 are satisfied. Consequently, we can use Corollary 4.4 [5.3] and (5.4) to obtain the following estimate for the second term in (5.1)

$$|\mathbb{P}(M_n \leq u_n) - L_n^{k_n}| \leq C\epsilon_n^{k_n-1} \left( k_n t_n \mathbb{P}(A_n) + \frac{n^2 \gamma(t_n)}{k_n} + \frac{(n\mathbb{P}(A_n))^2}{k_n} + \frac{n\mathbb{P}(A_n)}{k_n} \sum_{j=R_n}^{\ell_n-1} \gamma(j) \right)$$

$$\leq C'' e^{-\theta \tau} \left( k_n t_n \mathbb{P}(A_n) + \frac{n^2 \gamma(t_n)}{k_n} + \frac{(\theta \tau)^2}{k_n} + \theta \tau \sum_{j=R_n}^{\ell_n-1} \gamma(j) \right).$$

For the first term, we use first Proposition 2.4 then the statement of Corollary 4.6 (with $\tau = 1$) and finally the facts that $\lim_{n \to \infty} L_n = 1, \mathbb{P}(U_n \setminus A_n) \sim (1 - \theta) \tau/n$ and $\lim_{n \to \infty} k_n \mathcal{Y}_n = 0$ in order to obtain that there exists $C''' > 0$ such that

$$|\mathbb{P}(M_n \leq u_n) - \mathbb{P}(\mathcal{W}_{0,n}(A_n))| \leq \sum_{j=1}^{q} \mathbb{P}(\mathcal{W}_{0,n}(A_n) \cap f^{-n+j}(U_n \setminus A_n))$$

$$\leq q(L_n + \mathcal{Y}_n)^{k_n-1}\mathbb{P}(U_n \setminus A_n)$$

$$\leq C''' e^{-\theta \tau} (1 - \theta) \frac{\tau}{n}.$$
6. Improved Error Estimates for HTS

Proof of Theorem 5.6. Assume first that \( \tau \geq k_\varepsilon^{-1} \), so that \( |\tau k_\varepsilon| > 0 \).

By (2.8), we have that there exists \( C \) depending on \( \varepsilon \) but not on \( \tau \) such that
\[
|e^{-[\tau k_\varepsilon]t_\varepsilon p(A_\varepsilon)} - e^{-\theta \tau}| \leq e^{-\theta \tau} \left[ |\theta \tau - |\tau k_\varepsilon| t_\varepsilon p(A_\varepsilon)| + \alpha(|\theta \tau - |\tau k_\varepsilon| t_\varepsilon p(A_\varepsilon)|) \right]
\leq C\tau e^{-\theta \tau} \left| \theta - \frac{p(A_\varepsilon)}{p(B_\varepsilon)} + k_\varepsilon t_\varepsilon p(A_\varepsilon) \right| = C\tau e^{-\theta \tau} \alpha_\varepsilon. \tag{6.1}
\]

Using (2.7) and (6.1), we have that there exist \( C', C'' > 0 \) depending on \( \varepsilon \) but not on \( \tau \) such that
\[
\left| (1 - t_\varepsilon p(A_\varepsilon))^{[\tau k_\varepsilon]} - e^{-[\tau k_\varepsilon]t_\varepsilon p(A_\varepsilon)} \right| = e^{-[\tau k_\varepsilon]t_\varepsilon p(A_\varepsilon)} \left( \frac{|\tau k_\varepsilon| t_\varepsilon p(A_\varepsilon)^2}{2|\tau k_\varepsilon|} + o \left( \frac{1}{|\tau k_\varepsilon|} \right) \right)
\leq C'' e^{-[\tau k_\varepsilon]t_\varepsilon p(A_\varepsilon)} \frac{\tau}{k_\varepsilon^2}
\leq C'' e^{-\theta \tau} \left( \frac{\tau}{k_\varepsilon} + \frac{\tau^2}{k_\varepsilon \alpha_\varepsilon} \right).
\]

Hence, there exists \( C_1 > 0 \), depending on \( \varepsilon \) but not on \( \tau \), such that
\[
\left| (1 - t_\varepsilon p(A_\varepsilon))^{[\tau k_\varepsilon]} - e^{-\theta \tau} \right| \leq C_1 e^{-\theta \tau} \left( \tau \alpha_\varepsilon + \frac{\tau}{k_\varepsilon} + \frac{\tau^2}{k_\varepsilon \alpha_\varepsilon} \right). \tag{6.2}
\]

For the next term we need to use Corollary 4.5. For that purpose, we will check first its assumptions which include the assumptions of Lemma 4.3. First of all, we let \( n = [p(B_\varepsilon)^{-1}] \). Recall that \( p(A_\varepsilon) \sim \theta p(B_\varepsilon) \) and \( k_\varepsilon \to \infty \), as \( \varepsilon \to 0 \). Hence, for \( \varepsilon > 0 \) sufficiently small, we have that \( n/k_\varepsilon p(A_\varepsilon) < 2 \). Since \( k_\varepsilon t_\varepsilon p(B_\varepsilon)^{-1} \to 0 \), as \( \varepsilon \to 0 \), then for \( \varepsilon \) sufficiently small we have \( \ell_\varepsilon = \lfloor n/k_\varepsilon \rfloor / t_\varepsilon > 0 \) and \( \ell_\varepsilon - t_\varepsilon > 0 \). Note that \( L_\varepsilon = 1 - t_\varepsilon p(A_\varepsilon) \sim 1 - \theta / k_\varepsilon \to 1 \), as \( \varepsilon \to 0 \). We will now estimate \( k_\varepsilon Y_{A_\varepsilon} \) and show that \( k_\varepsilon Y_{A_\varepsilon} \to 0 \), as \( \varepsilon \to 0 \).

Observe that for all positive integer \( i \), we have \( p(\mathcal{W}_{\ell_\varepsilon t_\varepsilon + t_\varepsilon}) \sim L_\varepsilon^{i-2}(1 - (t_\varepsilon - t_\varepsilon)p(A_\varepsilon))^2 \) and \( p(\mathcal{W}_{0,0}) \sim L_\varepsilon \). Since \( t_\varepsilon = o(\ell_\varepsilon) \), \( |1_{A_\varepsilon}| \leq M \), \( t_\varepsilon p(A_\varepsilon) \to 0 \) and \( t_\varepsilon \gamma(t_\varepsilon) \) as \( \varepsilon \to 0 \), then for \( \varepsilon > 0 \) large enough we have \( 1 + t_{A_\varepsilon} t_\varepsilon p(A_\varepsilon) + ||1_{A_\varepsilon}||^2 \gamma(t_\varepsilon) < 2 \). Also recall that, by definition, we have that \( R_\varepsilon \to \infty \), as \( \varepsilon \to 0 \). This implies that \( \sum_{j=R_\varepsilon}^\infty \gamma(j) \to 0 \), as \( \varepsilon \to 0 \). Consequently, we may write for \( \varepsilon > 0 \) sufficiently small:
\[
Y_{A_\varepsilon} \leq 2t_\varepsilon p(A_\varepsilon) + M \gamma(t_\varepsilon) + M t_\varepsilon \gamma(t_\varepsilon) + t_\varepsilon^2 p(A_\varepsilon)^2 + M \gamma(t_\varepsilon)
+ M^2 (t_\varepsilon p(A_\varepsilon) + t_\varepsilon \gamma(t_\varepsilon)) \sum_{j=R_\varepsilon}^{\ell_\varepsilon-1} \gamma(j)
\leq 2t_\varepsilon p(A_\varepsilon) + 4M t_\varepsilon^2 \gamma(t_\varepsilon) + t_\varepsilon^2 p(A_\varepsilon)^2 + M^2 t_\varepsilon p(A_\varepsilon) \sum_{j=R_\varepsilon}^{\ell_\varepsilon-1} \gamma(j).
\]
Hence, there exists $C > 0$ such that for all $\varepsilon > 0$ sufficiently large we have

$$k_\varepsilon T_{A_\varepsilon} \leq C \left( k_\varepsilon t_\varepsilon \mathbb{P}(A_\varepsilon) + \frac{\mathbb{P}(B_\varepsilon)^{-2} \gamma(t_\varepsilon)}{k_\varepsilon} + \frac{1}{k_\varepsilon} + \sum_{j=\ell\varepsilon}^{\ell\varepsilon-1} \gamma(j) \right) = C \Gamma_\varepsilon. \quad (6.3)$$

By the choices of $k_\varepsilon, t_\varepsilon$ it is clear that $k_\varepsilon T_{A_\varepsilon} \to 0$, as $\varepsilon \to 0$. Applying Corollary 4.5 we have for all $\varepsilon$ small enough:

$$\left| \mathbb{P}(\mathcal{W}_{0,\tau n}(A_\varepsilon)) - L^{\lfloor k_\varepsilon \rfloor} \right| \leq 4\tau k_\varepsilon T_{A_\varepsilon} (L + T_{A_\varepsilon})^{\lfloor k_\varepsilon \rfloor} = 4\tau k_\varepsilon T_{A_\varepsilon} L^{\lfloor k_\varepsilon \rfloor} \left( 1 + \frac{\lfloor k_\varepsilon \rfloor T_{A_\varepsilon} L^{-1}}{\tau k_\varepsilon} \right)^{\lfloor k_\varepsilon \rfloor}
\leq 5\tau k_\varepsilon T_{A_\varepsilon} L^{\lfloor k_\varepsilon \rfloor} e^{\tau k_\varepsilon T_{A_\varepsilon}}.$$

Using now (6.2), we have that there exists $C_2 > 0$ depending on $\varepsilon$ (but not on $\tau$) such that

$$\left| \mathbb{P}(\mathcal{W}_{0,\tau n}(A_\varepsilon)) - L^{\lfloor k_\varepsilon \rfloor} \right| \leq C_2 \left( \tau^2 \alpha_\varepsilon \Gamma_\varepsilon + \frac{\tau^2}{k_\varepsilon} \Gamma_\varepsilon + \frac{\tau^3}{k_\varepsilon} \alpha_\varepsilon \Gamma_\varepsilon \right) e^{-\left( \theta - k_\varepsilon T_{A_\varepsilon} \right) \tau}. \quad (6.4)$$

By Proposition 2.4, Corollary 4.6 and (4.7)

$$\left| \mathbb{P}(\mathcal{W}_{0,\tau n}(B_\varepsilon)) - \mathbb{P}(\mathcal{W}_{0,\tau n}(A_\varepsilon)) \right| \leq \sum_{j=1}^{q} \mathbb{P}(\mathcal{W}_{0,\tau n}(A_\varepsilon) \cap f^{-n+j}(B_\varepsilon \setminus A_\varepsilon)) \leq q(1 - \theta)\mathbb{P}(B_\varepsilon)\mathbb{P}(\mathcal{W}_{0,\tau n}(A_\varepsilon)) \leq q(1 - \theta)\mathbb{P}(B_\varepsilon)(L + T_{A_\varepsilon})^{\lfloor k_\varepsilon \rfloor} - 1.$$
Note that if \( k_\varepsilon^{-1} \leq \tau < 1 \) then \((6.4)\) can be rewritten in the following way:

\[
|\mathbb{P}(\mathcal{W}_{0,\tau n}(A_\varepsilon)) - L^{\tau k_\varepsilon}| \leq C\varepsilon e^{-\theta \tau},
\]

for some \( C > 0 \) depending on \( \varepsilon \). Hence, in this case we may write that there exists some constant \( C > 0 \) depending on \( \varepsilon \) such that

\[
|\mathbb{P} \left( r_{B_\varepsilon(\xi)} > \frac{\tau}{\mathbb{P}(B_\varepsilon)} \right) - e^{-\theta \tau}| \leq C(\alpha_\varepsilon + \Gamma_\varepsilon)e^{-\theta \tau}.
\]

Now, we consider the case when \( \tau < k_\varepsilon^{-1} \). By Corollary 4.5, we have that, in this case, \( |\mathbb{P}(\mathcal{W}_{0,\tau n}(A_\varepsilon)) - (1 - \tau k_\varepsilon(\ell_\varepsilon + t_\varepsilon)\mathbb{P}(A_\varepsilon))| \leq \tau k_\varepsilon \Upsilon_\varepsilon \). Also note that there exist \( 0 < \xi < \theta \tau \) and \( C > 0 \) depending on \( \varepsilon \) such that

\[
\left| e^{-\theta \tau} - (1 - \tau k_\varepsilon(\ell_\varepsilon + t_\varepsilon)\mathbb{P}(A_\varepsilon)) \right| = \left| 1 - \theta \tau + \frac{e^{-\xi}}{2}((\theta \tau)^2 - (1 - \tau k_\varepsilon(\ell_\varepsilon + t_\varepsilon)\mathbb{P}(A_\varepsilon)) \right|
\]

\[
\leq C_\alpha_\varepsilon + \frac{e^{-\xi}}{2}(\theta \tau)^2.
\]

Since \( \tau < k_\varepsilon^{-1} < \Gamma_\varepsilon \) and, consequently, for \( \varepsilon > 0 \) sufficiently large we have \( e^{\theta \tau} \leq 1 \), then we can write, again, that there exists \( C > 0 \) depending on \( \varepsilon \) such that

\[
\left| \mathbb{P} \left( r_{B_\varepsilon(\xi)} > \frac{\tau}{\mathbb{P}(B_\varepsilon)} \right) - e^{-\theta \tau} \right| \leq C(\alpha_\varepsilon + \Gamma_\varepsilon)e^{-\theta \tau}.
\]

\( \square \)

7. An application: Rychlik systems

To apply the most powerful theorems in this paper, i.e., Theorems 3.4 and 3.6, we would like a dynamical system \( f : \mathcal{X} \rightarrow \mathcal{X} \) with a measure \( \mu \) which satisfies (R1) and (R2) and moreover has decay of correlations at more than quadratic rate against \( L^1 \) observables. In this section we give a natural class of examples of such a system, particular examples to keep in mind here are piecewise smooth, uniformly expanding, full-branched interval maps with an absolutely continuous invariant measure. The main idea is to use maps to which the seminal paper [Ryc83] applies, along with some extra information on the smoothness of potentials which will give us continuity of measure. The maps we define below are certainly not in the most general form possible, but we will make some restrictions for expository reasons. Note that the setup of [Kel12] also applies to this class of examples. We also recall that our theory applies to higher dimensional examples, such as those studied in [Sau00].

7.1. Interval maps modelled by a full shift. We will consider our maps \( F : \bigcup_i C^i \rightarrow X \) where \( X \) is an interval, \( \bigcup_i C^i \subset X \) is an at most countable union of open intervals and \( F : C^i \rightarrow X \) is a bijection. Let \( X^\infty = \{ x \in X : F^n(x) \text{ is defined for all } n \in \mathbb{N} \} \). Given a sequence of natural numbers \( i_0, \ldots, i_{n-1} \), the collection of points \( x \) which have \( F^{k}(x) \in C^k \) for \( k = 0, \ldots, n-1 \) is the corresponding \( n \)-cylinder. Let \( P_n \) denote the collection of all such cylinders. We will further assume that for any \( x, y \in X^\infty \), if \( x \neq y \) then there exists \( n \in \mathbb{N} \) such that \( x \) and \( y \) are in different \( n \)-cylinders. Since such maps are nicely modelled by the full shift, we denote the class of these maps by \( FS \).
7.2. Thermodynamic formalism for $FS$ maps. Given a potential $\Phi : X \to \mathbb{R}$, we define the $n$-th variation as

$$V_n(\Phi) := \sup_{C_n \in \mathcal{P}_n} \sup_{x,y \in C_n} \{|\Phi(x) - \Phi(y)|\}.$$ 

The potential is said to have summmable variations if $\sum_{n \geq 1} V_n(\Phi) < \infty$, and be locally Hölder if there exists $\alpha > 0$ such that $V_n(\Phi) = O(e^{-\alpha n})$.

Define

$$Z_n(\Phi) := \sum_{F^n x = x} e^{S_n \Phi(x)}$$

where $S_n \Phi$ is the ergodic sum $\Phi + \Phi \circ F + \cdots + \Phi \circ F^{n-1}$. Assuming that $\Phi$ has summable variations then we define the pressure as

$$P(\Phi) := \lim_{n \to \infty} \frac{1}{n} \log Z_n(\Phi).$$

Theorem 7.1. Suppose that $\Phi$ is a locally Hölder potential and $P(\Phi) < \infty$. Then there exists a $(\Phi - P(\Phi))$-conformal probability measure $m_\Phi$ and a density $\rho_\Phi$ such that $d\mu_\Phi = \rho_\Phi dm_\Phi$ defines an invariant probability measure. Moreover, $\log \rho_\Phi$ is locally Hölder; in particular $\rho_\Phi$ is bounded away from zero and infinity on cylinders. If $-\int \Phi d\mu < \infty$ then $\mu_\Phi$ is an equilibrium state for $\Phi$.

This theorem is part of [Sar01, Theorem 1] with the smoothness of $\rho_\Phi$ discussed in Remark 2 of that paper. Specifically for the full shift case we consider here, see also [MU01] and [Sar03]. Note that as proved in Sarig’s thesis, this theorem also holds under summable variations.

7.3. Continuity of measures: conditions [R1] and [R2]. Theorem 7.1 shows that $\rho_\Phi$ is fairly smooth in a symbolic sense (i.e., in terms of the metric on $X^\infty$ induced from the cylinder structure). However, this type of property also passes to the usual metric on the interval: namely, for $x \in X^\infty$

$$\lim_{\delta \to 0} \frac{\mu_\Phi(B_\delta(x))}{m_\Phi(B_\delta(x))} = \rho_\Phi(x).$$

(7.1)

This follows since by the Hölder continuity of $\log \rho_\Phi$, for any $\varepsilon > 0$ there exists $n$ such that for $x, y$ in the same $n$-cylinder, $e^{-\varepsilon} \rho_\Phi(x) \leq \rho_\Phi(y) \leq e^\varepsilon \rho_\Phi(x)$. Hence, once $\delta$ is so small that $B_\delta(x)$ is contained in the $n$-cylinder at $x$,

$$e^{-\varepsilon} \rho_\Phi(x) \leq \frac{\mu_\Phi(B_\delta(x))}{m_\Phi(B_\delta(x))} \leq e^\varepsilon \rho_\Phi(x).$$

A further remark is that since $m_\Phi$ is non-atomic and gives any open set positive measure, for $z \in X^\infty$, $\delta \mapsto m_\Phi(B_\delta(z))$ is continuous at 0. Moreover, by (7.1), $\delta \mapsto \mu_\Phi(B_\delta(z))$ is continuous at 0. So clearly [R1] holds. For [R2] we can use the conformality of $m_\Phi$ to prove that in fact $\theta = 1 - e^{S_{\Phi(z)} - P(\Phi)}$. 
7.4. **Rychlik conditions.** In order to get the nice decay of correlations for maps in $\mathcal{F}\mathcal{S}$ with good potentials, we use [Ryc83]. This requires the further assumption that $\text{Var}(e^{\Phi}) < \infty$, where here we mean $\text{Var}$ in the classical sense of variations. It is easy to show that this implies that $P(\Phi) < \infty$. In fact, as we see in the theorem below, we require more: if we define $\Phi : \bigcup_i C_i \to [-\infty, \infty)$ by

$$\Phi(x) = \begin{cases} \Phi(x) & \text{if } x \in \bigcup_i C_i, \\
-\infty & \text{if } x \in \bigcup_i \partial C_i. \end{cases}$$

**Theorem 7.2** ([Ryc83]). Suppose that $\Phi$ is a locally Hölder potential with $\text{Var}(e^{\Phi}) < \infty$ and $\Phi < 0$. Then the measure $\mu_\Phi$ obtained in Theorem 7.1 has exponential decay of correlations for $BV$ against $L^1$ observables.

We say that a system $(X, F, \Phi)$ satisfying the conditions in the theorem is a **Rychlik system**. Notice that a characteristic function on an annulus has $BV$ norm bounded by 4, so in Theorem 3.4 (Theorem 3.6) we can take $M$ to be 4 for $n$ large enough (for $\varepsilon$ small enough).

A simple example of an application of this theorem is to the case that there exist $\lambda > 1$ and $K \geq 1$ such that $F$ is $C^{1+Lip}$ and for each $i$, $F : C_i \to X$ has bounded distortion: $|DF(x)|/|DF(y)| \leq K$ for $x, y \in C_i$. Then set $\Phi(x) = -\log |DF(x)|$. Our conditions guarantee local Hölder continuity for $\Phi$ and $P(\Phi) < \infty$ as well as $\text{Var}(e^{\Phi}) \leq (C+K)m(X) < \infty$ (where $C$ is the Lipschitz constant) and $\Phi < 0$. If moreover, the union of the domains $C_i$ is equal to $X$ up to sets of Lebesgue measure zero, then $m_\Phi$ is Lebesgue and the resulting measure $\mu_\Phi$ is an equilibrium state which is also probability measure absolutely continuous w.r.t. Lebesgue (an acip).

7.5. **Discussion of $R_\varepsilon$.** Given a Rychlik system $(X, f, \mu)$ as above, we know from [STV02] that if $\mu$ is an ergodic $F$-invariant measure with positive entropy, then for a typical point $\zeta$, $R_\varepsilon = R(A_\varepsilon(\zeta))$ grows like $\varepsilon^{-d}$ for some $d \in (0, 1]$ (in fact $d$ is the dimension of the measure $\mu$). Also it is easy to see that for $\zeta$ a repelling periodic point, $R(A_\varepsilon(\zeta))$ is at least of order $\varepsilon^{-d'}$ where $d'$ depends on the strength of repulsion at $\zeta$. An argument showing that indeed $R(A_\varepsilon(\zeta)) \to \infty$ for all points $\zeta$ was given in [FFT12, Section 6].

**References**

[Abad01] Miguel Abadi, *Exponential approximation for hitting times in mixing processes*, Math. Phys. Electron. J. 7 (2001), Paper 2, 19 pp. (electronic). MR 1871384 (2002h:60069) [1.1]

[Abad04] , *Sharp error terms and necessary conditions for exponential hitting times in mixing processes*, Ann. Probab. 32 (2004), no. 1A, 243–264. MR MR2040782 (2004m:60042) [1.1]

[ACS00] Valentin Afraimovich, Jean-Rene Chazottes, and Benoît Saussol, *Local dimensions for Poincaré recurrences*, Electron. Res. Announc. Amer. Math. Soc. 6 (2000), 64–74 (electronic). MR 1777857 (2001d:37021) [3.10]
[AFLV11] José F. Alves, Jorge M. Freitas, Stefano Luzzatto, and Sandro Vaienti, From rates of mixing to recurrence times via large deviations, Adv. Math. 228 (2011), no. 2, 1203–1236. MR 2822221

[AVF12] Hale Aytaç, Jorge Milhazes Freitas, and Sandro Vaienti, Laws of rare events for deterministic and random dynamical systems, To appear in Transactions of the American Mathematical Society, (Preprint arXiv:1207.5188), 2012.

[AG01] M. Abadi and A. Galves, Inequalities for the occurrence times of rare events in mixing processes. The state of the art, Markov Process. Related Fields 7 (2001), no. 1, 97–112. Inhomogeneous random systems (Cergy-Pontoise, 2000). MR 1835750

[AL13] Miguel Abadi and Rodrigo Lambert, The distribution of the short-return function, Nonlinearity 26 (2013), no. 5, 1143–1162. MR 3043376

[Ald82] David J. Aldous, Markov chains with almost exponential hitting times, Stochastic Process. Appl. 13 (1982), no. 3, 305–310. MR 671039

[AS11] Miguel Abadi and Benoit Saussol, Hitting and returning to rare events for all α-mixing processes, Stochastic Process. Appl. 121 (2011), no. 2, 314–323. MR 2746177

[AV08] Miguel Abadi and Sandro Vaienti, Large deviations for short return, Discrete Contin. Dyn. Syst. 21 (2008), no. 3, 729–747. MR 2399435 (2009j:37015)

[AV09] Miguel Abadi and Nicolas Vergne, Sharp error terms for return time statistics under mixing conditions, J. Theoret. Probab. 22 (2009), no. 1, 18–37. MR 2472003 (2010f:60068)

[CHM91] Michael R. Chernick, Tailen Hsing, and William P. McCormick, Calculating the extremal index for a class of stationary sequences, Adv. in Appl. Probab. 23 (1991), no. 4, 835–850. MR MR1133731 (93c:60073)

[Col01] P. Collet, Statistics of closest return for some non-uniformly hyperbolic systems, Ergodic Theory Dynam. Systems 21 (2001), no. 2, 401–420. MR 1827111

[Dol98] Dmitry Dolgopyat, On decay of correlations in Anosov flows, Ann. of Math. (2) 147 (1998), no. 2, 357–390. MR 1626749 (99g:58073)

[DY06] Mark F. Demers and Lai-Sang Young, Escape rates and conditionally invariant measures, Nonlinearity 19 (2006), no. 2, 377–397. MR 2199394 (2006j:37051)

[FFT10] Ana Cristina Moreira Freitas, Jorge Milhazes Freitas, and Mike Todd, Hitting time statistics and extreme value theory for dynamical systems, Statist. Probab. Lett. 78 (2008), no. 9, 1088–1093. MR 2422964 (2009e:37062)

[FFT11] Ana Cristina Moreira Freitas, Jorge Milhazes Freitas, and Mike Todd, Extreme value laws in dynamical systems for non-smooth observations, J. Stat. Phys. 142 (2011), no. 1, 108–126. MR 2749711 (2012a:60149)

[FFT12] Ana Cristina Moreira Freitas, Jorge Milhazes Freitas, and Mike Todd, The extremal index, hitting time statistics and periodicity, Adv. Math. 231 (2012), no. 5, 2626 – 2665.

[FP12] Andrew Ferguson and Mark Pollicott, Escape rates for gibbs measures, Ergod. Theory Dynam. Systems 32 (2012), 961–988.

[Fre13] Jorge Milhazes Freitas, Extremal behaviour of chaotic dynamics, Dyn. Syst. 28 (2013), no. 3, 302–332.

[GS97] A. Galves and B. Schmitt, Inequalities for hitting times in mixing dynamical systems, Random Comput. Dynam. 5 (1997), no. 4, 337–347. MR 1483874 (98i:60017)

[HSV99] Masaki Hirata, Benoît Saussol, and Sandro Vaienti, Statistics of return times: a general framework and new applications, Comm. Math. Phys. 206 (1999), no. 1, 33–55. MR 1736991 (2001c:37007)

[HV10] Nicolai Haydn and Sandro Vaienti, The Rényi entropy function and the large deviation of short return times, Ergodic Theory Dynam. Systems 30 (2010), no. 1, 159–179. MR 2586350 (2011a:37017)
[HW79] W. J. Hall and Jon A. Wellner, *The rate of convergence in law of the maximum of an exponential sample*, Statist. Neerlandica 33 (1979), no. 3, 151–154. MR 552259 (80k:60029)

[Kel12] Gerhard Keller, *Rare events, exponential hitting times and extremal indices via spectral perturbation*, Dynamical Systems 27 (2012), no. 1, 11–27.

[KL09] Gerhard Keller and Carlangelo Liverani, *Rare events, escape rates and quasistationarity: some exact formulae*, J. Stat. Phys. 135 (2009), no. 3, 519–534. MR 2535206 (2011a:37012)

[Lea74] M. R. Leadbetter, *On extreme values in stationary sequences*, Z. Wahrscheinlichkeitstheorie und Verw. Gebiete 28 (1973/74), 289–303. MR MR0362465 (50 #14906)

[LLR83] M. R. Leadbetter, Georg Lindgren, and Holger Rootzén, *Extremes and related properties of random sequences and processes*, Springer Series in Statistics, Springer-Verlag, New York, 1983.

[LN89] M. R. Leadbetter and S. Nandagopalan, *On exceedance point processes for stationary sequences under mild oscillation restrictions*, Extreme value theory (Oberwolfach, 1987), Lecture Notes in Statist., vol. 51, Springer, New York, 1989, pp. 69–80. MR MR0992049 (90k:60065)

[MS01] William P. McCormick and Lynne Seymour, *Rates of convergence and approximations to the distribution of the maximum of chain-dependent sequences*, Extremes 4 (2001), no. 1, 23–52. MR 1876178 (2002i:60106)

[MU01] R. Daniel Mauldin and Mariusz Urbański, *Gibbs states on the symbolic space over an infinite alphabet*, Israel J. Math. 125 (2001), 93–130. MR 1853808 (2002k:37048)

[Res08] Sidney I. Resnick, *Extreme values, regular variation and point processes*, Springer Series in Operations Research and Financial Engineering, Springer, New York, 2008, Reprint of the 1987 original. MR 2364939 (2008h:60002)

[Ryc83] Marek Rychlik, *Bounded variation and invariant measures*, Studia Math. 76 (1983), no. 1, 69–80. MR MR728198 (85h:28019)

[Sar01] Omri M. Sarig, *Thermodynamic formalism for null recurrent potentials*, Israel J. Math. 121 (2001), 285–311. MR 1818392 (2001m:37056)

[Sar03] Omri Sarig, *Existence of Gibbs measures for countable Markov shifts*, Proc. Amer. Math. Soc. 131 (2003), no. 6, 1751–1758 (electronic). MR 1955261 (2004b:37056)

[Sau00] Benoît Saussol, *Absolutely continuous invariant measures for multidimensional expanding maps*, Israel J. Math. 116 (2000), 223–248. MR 1759406 (2001c:37037)

[Smi82] Richard L. Smith, *Uniform rates of convergence in extreme-value theory*, Adv. in Appl. Probab. 14 (1982), no. 3, 600–622. MR 665296 (84h:60054)

[STV02] B. Saussol, S. Troubetzkoy, and S. Vaienti, *Recurrence, dimensions, and Lyapunov exponents*, J. Statist. Phys. 106 (2002), no. 3-4, 623–634. MR MR1884547 (2003a:37007)
