Habilitation à diriger des recherches

Spécialité: **Mathématiques**

Document de synthèse

Étude quantitative du comportement à long terme des orbites: récurrence de Poincaré et plus petite distance

Présentée par

Jérôme Rousseau

Soutenue le 14 Avril 2021 à l'Université de Rennes 1, devant le jury composé de

Jorge FREITAS	Professeur, Universidade do Porto
Sébastien GOUËZEL	Directeur de Recherche, Université de Rennes 1
Carlos MATHEUS	Directeur de recherche, École Polytechnique
Florence MERLEVÈDE	Professeure, Université Paris-Est Marne-la-Vallée
Barbara SCHAPIRA	Maître de Conférences, Université de Rennes 1
Sandro VAIENTI	Professeur, Université de Toulon
Hong-Kun ZHANG	Professeure, University of Massachusetts

Remerciements

Je voudrais tout d'abord remercier Benoît Saussol qui depuis le début de ma thèse et jusqu'à aujourd'hui m'a soutenu, conseillé et avec qui j'ai énormément appris.

I would like to thank Hong-Kun Zhang who accepted to be a reviewer for my habilitation and for the interest she had for my work.

Je voudrais remercier Florence Merlevède et Carlos Matheus d'avoir accepter d'être rapporteurs et d'avoir lu attentivement mon manuscrit.

Un grand merci à Barbara Schapira et Sébastien Gouëzel pour avoir fait partie de mon jury et avoir bravé le covid pour être présent lors de ma soutenance.

Eu gostaria de agradecer imensamente Jorge e Ana Cristina Freitas por todas as coisas que fizeram pra eu (e pra Vanessa). Um agradecimento particular pra Jorge por ter participado do jury e por sempre aceitar de reler meus artigos.

Je voudrais remercier Sandro Vaienti pour avoir fait partie du jury et pour tous les échanges que nous avons pu avoir depuis le début de ma licence jusqu'à aujourd'hui.

Je tiens à remercier tous les collègues et amis avec qui j'ai pu discuté de maths et autres et qui ont contribué de manière plus ou moins directe à l'avancement de ma recherche, par ordre alphabétique: Miguel Abadi, Romain Aimino, José Alves, Jayadev Athreya, Davide Azevedo, Oliver Butterley, Adriana Coutinho, Davide Faranda, Stefano Galatolo, Nicolai Haydn, Rodrigo Lambert, Lingmin Liao, Philippe Marie, Matt Nicol, Isaia Nisoli, Françoise Pène, Manuel Stadlbauer, Mike Todd, Paulo Varandas, Fan Yang, Yun Zhao.

Je remercie bien sur ma famille pour leur soutien.

Finalemente, eu gostaria de agradecer a Luca e Vanessa pelo amor, pelo carinho e pelo apoio.

Contents

Introduction

Pour essayer d'étudier la réalité, les scientifiques expérimentaux ont souvent besoin d'approcher leur système ou d'utiliser un système simplifié pour obtenir un système de dimension plus petite, plus adapté à une analyse. Avec la même philosophie, les scientifiques travaillant avec des systèmes de grande dimension sont généralement juste intéressés par la mesure de certaines quantités (température, pression, vitesse du vent, hauteur des vagues,...) ou essayent d'utiliser une mesure ou une observation du système pour obtenir des informations sur le système entier.

En suivant ces idées, depuis le début de ma thèse de doctorat, l'un de mes principaux objectifs a été d'étudier certaines propriétés statistiques de systèmes dynamiques observés. Plus précisément, pour un système dynamique (X, T, μ), au lieu d'étudier l'orbite d'un point x (x, $T(x)$,..., $T^n(x)$, ...), je m'intéresse à l'observation de cet orbite ($f(x)$, $f(T(x))$,..., $f(T^n(x))$,...), où $f : X \to Y$ est une fonction à valeur dans un espace Y (a priori de dimension plus petite que X).

D'un point de vue théorique, nous verrons comment obtenir des résultats sur les systèmes dynamiques aléatoires grâce à l'analyse des systèmes dynamiques observés.

Les systèmes dynamiques aléatoires, qui généralisent les systèmes dynamiques déterministes, permettent de modéliser plus précisément les phénomènes naturels (comme par exemple l'existence de petites perturbations ou d'erreurs d'approximation). Contrairement aux systèmes déterministes qui ne considèrent que l'itération d'une unique fonction, avec les systèmes aléatoires nous pouvons composer des transformations différentes (en ajoutant par exemple un bruit ou des perturbations aléatoires). Nous noterons que cela difficulte l'analyse de leurs propriétés statistiques, en particulier lorsque ces différentes transformations n'ont pas de mesure invariante commune.

Dans mes travaux, je me suis concentré sur l'étude de deux objets: les temps de retour (et d'entrée) et la plus petite distance entre des orbites.

Pour les temps de retour/entrée, ma principale contribution dans le domaine a été de généraliser des résultats connus (sur les taux de récurrence et les fluctuations des temps de retour/entrée) aux systèmes dynamiques observés et aux systèmes dynamiques aléatoires.

Ainsi, dans le Chapitre 2 (dont les résultats proviennent en grande partie de la thèse de doctorat de l'auteur), nous nous intéressons aux taux de récurrence. Plus précisément, nous présenterons des résultats sur le comportement asymptotique du temps de retour pour l'observation, défini par:

$$\tau^f_{B(f(x),r)}(x) := \inf \left\{ k \in \mathbb{N}^* : \, f(T^k x) \in B\left(f(x), r\right) \right\}.$$

Pour des systèmes mélangeant rapidement, le comportement de τ_r^f (quand $r \to 0$) est du type r^{-d} avec un exposant dépendant de la dimension locale de la mesure image $f_*\mu$ [RS10]. Nous expliquerons comment utiliser ces résultats pour étudier le comportement des temps de retour pour les systèmes dynamiques aléatoires [MR11]. Finalement, nous montrerons aussi que pour certains flots hyperboliques (et en particulier pour le flot géodésique), le comportement du temps de retour est lié à la dimension locale de la mesure invariante [R12].

Dans le Chapitre 3, nous nous concentrons sur les fluctuations des temps d'entrée et de retour. En particulier, en considérant $\tau^f_{B(f(x_0),r)}(.)$ comme une variable aléatoire, nous présenterons des résultats de convergence en loi [R14].

Nous noterons ici que l'étude des distributions des temps d'entrée et de retour s'est grandement développée ces dix dernières années. En effet, une connection a été établie entre la distribution des temps d'entrée/retour et la Théorie des Valeurs Extrêmes [36]. Ainsi, cela a apporté une nouvelle vision sur le sujet et de nouvelles techniques pour étudier les évènements rares. En particulier comme ces évènements correspondent souvent à une déviation du comportement moyen (qui, la plupart du temps, sont liés à des événements indésirables), il y a un intérêt pratique à les étudier et ils ont une importance cruciale dans des domaines comme la finance, l'assurance et l'écologie, entre autres.

Nous analyserons aussi la distribution des temps d'entrée et de retour pour les systèmes dynamiques aléatoires. Dans ce cas, deux types de résultats seront présentés: des résultats intégrés (annealed) et fibrés (quenched). Dans le cas intégré, nous obtiendrons (en appliquant nos résultats pour les observations de systèmes dynamiques) une convergence exponentielle pour des systèmes dynamiques aléatoires mélangeant rapidement. Le cas fibré est plus subtile, ainsi les résultats les plus complets seront donnés pour les sous-shifts aléatoires de type fini [RSV14, RT15] et une convergence exponentielle sera aussi présentée pour des systèmes aléatoires mélangeant vérifiant certaines hypothèses géométriques [HRY20].

La dernière partie du chapitre sera consacrée aux larges déviations pour les temps de retour [CRS18].

Après les temps d'entrée/retour, le deuxième objet d'étude principal que nous présentons ici est la plus petite distance entre les orbites. Plus précisément, pour deux points x, y nous étudions le comportement asymptotique de la plus petite distance entre l'orbite de x et l'orbite de y, définie par:

$$m_n(x,y) = \min_{i,j=0,\dots,n-1} \left(d(T^i(x), T^j(y)\right).$$

Nous soulignerons ici que nous avons été la première équipe à définir et étudier cet objet dans [BLR19]. Nous verrons donc dans le Chapitre 4 que m_n décroit de façon polynomiale avec un exposant dépendant de la dimension fractale généralisée (aussi connu sous le nom de dimension L^q ou dimension HP).

Les dimensions fractales généralisées ont été introduites initialement pour caractériser et mesurer l'étrangeté des attracteurs chaotiques et plus généralement pour décrire la structure fractale des ensembles invariants [43, 44]. Comme l'estimation de ces dimensions joue un role important dans la description des systèmes dynamiques, différentes approches numériques et procédés ont été développés pour les calculer (e.g. [34]) et l'étude de la plus petite distance entre les orbites apporte une nouvelle technique pour estimer ces dimensions.

Nous décrirons aussi dans ce chapitre, le comportement de la plus petite distance pour les observations de systèmes dynamiques et pour les systèmes dynamiques aléatoires [BLR19, CLR20, BR21].

Il est intéressant de noter (et cela sera développé plus amplement dans le Chapitre 5) que lorsque nous travaillons avec un système dynamique symbolique, étudier la plus petite distance entre les orbites revient à étudier la taille de la plus grande sous-chaine commune entre deux suites.

Le problème de la plus grande sous-chaine commune est un problème bien connu et bien étudié avec des applications en biologie (e.g. pour l'étude des suites d'ADN), en informatique (e.g. pour les algorithmes de compression) ou encore en linguistique (e.g. pour des comparaison de textes ou de languages). La technique développée dans [BLR19] pour étudier la plus petite distance nous a donc permis de généraliser des résultats sur la plus long sous-chaine commune (en particulier les résultats de [8]) à des processus stochastiques mélangeant rapidement mais aussi à des processus stochastiques en milieux aléatoires [R21].

Tout au long de ce document, nous verrons que les dimensions fractales sont liées au temps de retour mais aussi à la plus petite distance entre les orbites. Nous rappellerons donc brièvement dans le Chapitre 1, quelques notions de théorie de la dimension. Nous donnerons aussi la définition générale de systèmes dynamiques aléatoires et de sous-shift aléatoire de type fini que nous utiliserons dans ce document.

Pour distinguer mes travaux des autres références, j'utiliserai des références alpha-numériques pour mes articles et des références numériques pour le reste.

Par souci d'homogénéité, les résultats obtenus dans les articles [RVZ12] (sur la récurrence dans des boules dynamiques avec erreurs) et [AR16] (sur les inégalités de concentration dans les systèmes séquentiels) ne seront pas abordés dans ce document.

Chapter 1

A short introduction to dimension theory and random dynamical systems

In this chapter, we will recall some definitions and properties of dimension theory and random dynamical systems.

Local dimensions and generalized fractal dimensions play a major role when studying statistical properties of dynamical systems and will appear in our results throughout this document. We will only give here the basic information needed in the next chapters, we refer the reader to e.g. [31, 65] for extensive studies on dimension theory.

Random dynamical systems will be one of the main subject of our analysis, thus we will give in this chapter the general definition of random dynamical systems and random subshift of finite type that we will use in this document. One can see e.g. the review [58] for a detailed introduction to this theory.

1.1 Local dimension and generalized fractal dimensions

Let X be a metric space and denote $B(x,r)$ the ball centered in x and of radius r.

The lower and upper pointwise or local dimension of a Borel probability measure μ on X at a point $x \in X$ are defined by

$$\underline{d}_\mu(x) = \varliminf_{r\to 0} \frac{\log \mu\left(B\left(x,r\right)\right)}{\log r} \qquad \text{and} \qquad \overline{d}_\mu(x) = \varlimsup_{r\to 0} \frac{\log \mu\left(B\left(x,r\right)\right)}{\log r}.$$

The measure μ is called exact dimensional if there exists a constant d_μ such that

$$\underline{d}_\mu(x) = \overline{d}_\mu(x) = d_\mu \quad \text{for} \quad \mu\text{-almost every} \quad x \in X.$$

We recall that the Hausdorff dimension of a probability measure μ on X is given by

$$\dim_H \mu = \inf\{\dim_H Z : \mu(Z) = 1\},$$

where $\dim_H Z$ denotes the Hausdorff dimension of Z.

For an exact dimensional measure, the Hausdorff dimension and the local dimension coincide:

Proposition 1.1 ([79]). *If μ is exact dimensional, then*

$$d_\mu = \dim_H \mu.$$

For $k \neq 1$, the lower and upper generalized fractal dimensions (also known as L^q or HP dimensions) of a measure μ are defined by:

$$\underline{D}_k(\mu) = \varliminf_{r\to 0} \frac{\log \int_X \mu\left(B\left(x,r\right)\right)^{k-1} d\mu(x)}{(k-1)\log r}$$

and

$$\overline{D}_k(\mu) = \varlimsup_{r\to 0} \frac{\log \int_X \mu\left(B\left(x,r\right)\right)^{k-1} d\mu(x)}{(k-1)\log r}.$$

When the limit exists we will denote the common value of $\underline{D}_k(\mu)$ and $\overline{D}_k(\mu)$ by $D_k(\mu)$.

When the measure is exact dimensional, we have:

Proposition 1.2 ([17]). *If μ is exact dimensional, then*

$$\lim_{\substack{k\to 1\\ k>1}} \overline{D}_k(\mu) \leq d_\mu = \dim_H \mu \leq \lim_{\substack{k\to 1\\ k<1}} \underline{D}_k(\mu).$$

For other properties of the generalized fractal dimensions, their existence and relations with other dimensions, one can see e.g. [17, 32, 65, 66].

1.2 Random dynamical systems

Let Ω be a metric space and $\mathcal{B}(\Omega)$ its Borelian σ-algebra. Let $\vartheta : \Omega \to \Omega$ be a measurable transformation on Ω preserving some probability measure $\mathbb{P}$. Given a compact metric space X and a family $\mathcal{T} = (T_\omega)_{\omega\in\Omega}$ of transformations $T_\omega : X \to X$, we say that it defines a *random dynamical system* over $(\Omega, \mathcal{B}(\Omega), \mathbb{P}, \vartheta)$ via $T^n_\omega = T_{\vartheta^{n-1}(\omega)} \circ \cdots \circ T_{\vartheta(\omega)} \circ T_\omega$ for every $n \geq 1$ and $T^0_\omega = Id$.

The dynamics of the random dynamical systems generated by $\mathcal{T}$ over$(\Omega, \mathcal{B}(\Omega), \mathbb{P}, \vartheta)$ is given by the skew-product:

$$\begin{aligned} S : \Omega \times X &\to \Omega \times X \\ (\omega, x) &\to (\vartheta(\omega), T_\omega(x)). \end{aligned}$$

A probability measure μ is *invariant* by the random dynamical system if it is S-invariant and $\pi_*\mu = \mathbb{P}$, where $\pi : \Omega \times X \to \Omega$ is the canonical projection.

Henceforth, we denote by ν the marginal of μ on X, i.e. $\nu = \int \mu_\omega \, d\mathbb{P}$ where $(\mu_\omega)_\omega$ denote the decomposition of μ on X, that is, $d\mu(\omega, x) = d\mu_\omega(x)d\mathbb{P}(\omega)$. The measures μ_ω are called the *sample measures*.

1.3 Random subshift of finite type

We now give the definition of a particular random dynamical system: random subshift of finite type.

Let $(\Omega, \theta, \mathbb{P})$ be an invertible ergodic measure preserving system, set $X = \mathbb{N}^{\mathbb{N}}$ and let $\sigma : X \to X$ denote the shift. Let $b : \Omega \to \mathbb{N}$ be a random variable. Let $A = \{A(\omega) = (a_{ij}(\omega)) : \omega \in \Omega\}$ be a random transition matrix, i.e. for any $\omega \in \Omega$, $A(\omega)$ is a $b(\omega) \times b(\theta\omega)$-matrix with entries in $\{0, 1\}$, at least one non-zero entry in each row and each column and such that $\omega \mapsto a_{ij}(\omega)$ is measurable for any $i \in \mathbb{N}$ and $j \in \mathbb{N}$. For any $\omega \in \Omega$ define the subset of the integers $X_\omega = \{1, \ldots, b(\omega)\}$ and

$$\mathcal{E}_\omega = \{x = (x_0, x_1, \ldots)\colon x_i \in X_{\theta^i\omega} \text{ and } a_{x_i x_{i+1}}(\theta^i\omega) = 1 \text{ for all } i \in \mathbb{N}\} \subset X,$$

$$\mathcal{E} = \{(\omega, x)\colon \omega \in \Omega, x \in \mathcal{E}_\omega\} \subset \Omega \times X.$$

We consider the random dynamical system coded by the skew-product $S : \mathcal{E} \to \mathcal{E}$ given by $S(\omega, x) = (\theta\omega, \sigma x)$. Let μ be an S-invariant probability measure with marginal $\mathbb{P}$ on Ω and let $(\mu_\omega)_\omega$ denote its decomposition on $\mathcal{E}_\omega$, that is, $d\mu(\omega, x) = d\mu_\omega(x)d\mathbb{P}(\omega)$. The measures μ_ω are called the *sample measures*. Note $\mu_\omega(A) = 0$ if $A \cap \mathcal{E}_\omega = \emptyset$. We denote by $\nu = \int \mu_\omega \, d\mathbb{P}$ the marginal of μ on X.

We emphasize that the sample measures are not invariant. However, since θ is invertible, by S-invariance of μ and almost everywhere uniqueness of the decomposition $d\mu = d\mu_\omega \, d\mathbb{P}$, we get for $\mathbb{P}$-almost every $\omega \in \Omega$,

$$(\sigma^i)_*\mu_\omega = \mu_{\theta^i\omega} \qquad \text{for all } i \in \mathbb{N}.$$

For $y \in X$ we denote by $C_n(y) = \{z \in X : y_i = z_i \text{ for all } 0 \leq i \leq n-1\}$ the *n-cylinder* that contains y and we set $\mathcal{F}_0^n(X)$ as the sigma-algebra in X generated by all the n-cylinders.

Chapter 2

Recurrence rates

Poincaré recurrence Theorem is one of the fundamental theorems at the origin of dynamical systems. This theorem states that the orbit of almost every point will come back as close as you want from its starting point. Unfortunately, it does not give any information on the time needed to come back (called the return time). This result was refined by Kač [56] who proved that for an ergodic dynamical system, in a set of positive measure, the mean of the return time in this set is equal to the inverse of the measure of this set.

Recent results showed relations between quantitative indicators representing the scaling behavior of return times in small targets, decay of correlations and dimension theory. If we consider the behavior, as $r \to 0$, of the return time $\tau_r(x)$ of a point x in $B(x,r)$, in many interesting cases this is a power law $\tau_r(x) \sim r^R$. This exponent R gives a quantitative measure of the speed of recurrence of an orbit near to its starting point, and this will be a quantitative recurrence indicator.

A general philosophy is that in "chaotic" systems this exponent is equal to the local dimension of the invariant measure (see e.g. [19, 18, 73, 74, 40]).

However, considering observations of the system (for example, temperature or pressure while studying climate) could be more significant than considering the whole dynamical system. Thus, in [RS10], we wondered if similar results could be obtained if we study the return time of the image (or observation) of the orbit. More precisely, for a measurable function $f : X \to Y$, we defined and studied $\tau_r^f(x) := \inf\left\{k \in \mathbb{N}^* : f(T^k x) \in B\left(f(x), r\right)\right\}$. These results will be presented in Section 2.1 and we will explained how to use these results to study the return time in random dynamical systems in Section 2.2.

In continuous time, results on return times are scarcer than in discrete time. Barreira and Saussol [18] proved that for a suspension flow over an Anosov diffeomorphism such that the invariant measure is an equilibrium state of an Hölder potential, the return time of ν-almost every point y in the ball $B(y,r)$ behaves like $r^{-\dim_H \nu + 1}$ when r goes to zero (similar results have been proved for hitting time of Lorenz like flows [41]).

Pène and Saussol [67] studied the billiard flow in the plane with periodic configuration of scatterers, they proved that, almost everywhere, the return time of a point (p,v) in the ball $B((p,v),r)$ is of the order $\exp(\frac{1}{r^2})$ and that the return time of the position of a point in $B(p,r)$, i.e. the return time of the projection of the flow on the billiard, is of the order $\exp(\frac{1}{r})$ almost everywhere.

Following these works and the idea presented above, we studied in [R12] the recurrence rates for flows and observations of flows and these results will be presented in Section 2.3.

2.1 Recurrence rates for observations

Let us assume that X is a metric space and $\mathcal{A}$ is its Borel σ-algebra.

Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving system (m.p.s.) i.e. μ is a measure on $(X, \mathcal{A})$ with $\mu(X) = 1$ and μ is invariant by T (i.e $\mu(T^{-1}A) = \mu(A)$ for all $A \in \mathcal{A}$) where $T : X \to X$.

Let $f : X \to Y$ be a function, called observable (we will specify the space X and Y later).

We introduce the hitting and return time for the observation and its associated recurrence rates.

Definition 2.1. *Let $f : X \to \mathbb{R}^N$ be a measurable function, called observation, and $A \subset \mathbb{R}^N$, we define for $x \in X$ the hitting time for the observation of x in A:*

$$\tau_A^f(x) := \inf\left\{k \in \mathbb{N}^* : \, f(T^k x) \in A\right\}.$$

Also, we define for $x \in X$ the return time for the observation:

$$\tau_r^f(x) := \inf\left\{k \in \mathbb{N}^* : \, f(T^k x) \in B\left(f(x), r\right)\right\}.$$

We then define the lower and upper recurrence rate for the observation:

$$\underline{R}^f(x) := \liminf_{r\to 0} \frac{\log \tau_r^f(x)}{-\log r} \qquad \overline{R}^f(x) := \limsup_{r\to 0} \frac{\log \tau_r^f(x)}{-\log r}.$$

We are now able to give our first result, linking recurrence rates and pointwise dimensions:

Theorem 1 ([RS10])**.** *Let $(X, \mathcal{A}, \mu, T)$ be a m.p.s. Consider a measurable observable $f : X \to Y = \mathbb{R}^N$. Then*

$$\underline{R}^f(x) \le \underline{d}_{f_*\mu}(f(x)) \qquad \textit{and} \qquad \overline{R}^f(x) \le \overline{d}_{f_*\mu}(f(x))$$

for μ-almost every $x \in X$.

This result is satisfactory in the sense that it holds for any dynamical system and observation. Moreover, under natural assumptions we will show that the equality is true.

Remark 2.2. *We observe that these inequalities may be strict, the caricatural example is when T is the identity map. Some more interesting examples were treated in [GRS15], where it was shown that even with a polynomial decay of correlations, one could obtain strict inequalities.*

We then can introduce the decay of correlations:

Definition 2.3. (X, T, μ) *has a super-polynomial decay of correlations if, for all ψ Lipschitz function from X to $\mathbb{R}$, for all ϕ measurable bounded function from X to $\mathbb{R}$ and for all $n \in \mathbb{N}^*$, we have:*

$$\left| \int_X \psi.\phi \circ T^n \, d\mu - \int_X \psi d\mu \int_X \phi d\mu \right| \le \|\psi\|_{Lip} \|\phi\|_\infty \theta_n$$

with $\lim_{n\to\infty} \theta_n n^p = 0$ for all $p > 0$.

To obtain optimal results on the return time for the observation we need to assume that the system presents some kind of aperiodicity:

Definition 2.4. *A m.p.s. $(X, \mathcal{A}, \mu, T)$ is called μ-almost aperiodic for the observation f if*

$$\mu\left(x \in X : \exists n \in \mathbb{N}^* \text{ such that } f(T^n x) = f(x)\right) = 0.$$

We emphasize that this condition can be remove introducing non-instantaneous return times [RS10].

With these conditions, we can improve the previous theorem.

Theorem 2 ([RS10])**.** *Let $(X, \mathcal{A}, \mu, T)$ be a m.p.s μ-almost aperiodic for the observation f and with a super-polynomial decay of correlations. Consider a Lipschitz observable $f : X \to Y = \mathbb{R}^N$. Then, we have*

$$\underline{R}^f(x) = \underline{d}_{f_*\mu}(f(x)) \qquad \text{and} \qquad \overline{R}^f(x) = \overline{d}_{f_*\mu}(f(x))$$

for μ-almost every x such that $\underline{d}^f_\mu(x) > 0$.

Moreover, if $f_*\mu$ is exact dimensional, we have

$$\underline{R}^f(x) = \overline{R}^f(x) = \dim_H f_*\mu \qquad \text{for } \mu\text{-almost every } x \in X.$$

Taking the identity function for f, we recover the result of [18] and [73] under weaker assumptions. The main assumption of the theorem about decay of correlations is satisfied in a variety of systems with some hyperbolic behavior and studied in an abundant literature (e.g. [15, 80]).

Theorem 2 does not apply to those points where $\underline{d}_{f_*\mu}(f(x)) = 0$. When $\overline{d}_{f_*\mu}(f(x)) = 0$ also, this is not a restriction because Theorem 1 applies and gives $\overline{R}^f(x) = \underline{R}^f(x) = 0$. However, the question remains when $\overline{d}_{f_*\mu}(f(x)) \neq \underline{d}_{f_*\mu}(f(x)) = 0$ on a positive measure set. Indeed, the assumptions of Theorem 2 are not strong enough to ensure the almost everywhere existence of the pointwise dimension for the observations.

In [RS10], a short example was given to remark that one can obtain results on the quantitative study of Poincaré recurrence for random dynamical systems using the study of recurrence for observations of dynamical systems. This idea was developed in [MR11] to study recurrence rates for random dynamical systems and will be explained in the next section.

2.2 Recurrence rates for random dynamical systems

We consider in this section $\mathcal{T}$ a random dynamical system on X over $(\Omega, \mathcal{B}(\Omega), \mathbb{P}, \vartheta)$ with an invariant measure μ, as defined in Section 1.2.

For a fixed $\omega \in \Omega$, the quenched random hitting time in a measurable subset $A \subset X$ of the random orbit starting from a point $x \in X$ is:

$$\tau_A^\omega(x) = \inf\{n > 0 \, : \, T_\omega^n x \in A\}.$$

We also define the quenched random return time of a point $x \in X$ into the open ball $B(x, r)$:

$$\tau_r^\omega(x) \;\; := \;\; \inf\left\{k > 0 : T_\omega^n x \in B(x, r)\right\}$$

and the quenched random lower and upper recurrence rates:

$$\underline{R}^\omega(x) := \liminf_{r\to 0} \frac{\log \tau_r^\omega(x)}{-\log r} \qquad \text{and} \qquad \overline{R}^\omega(x) := \limsup_{r\to 0} \frac{\log \tau_r^\omega(x)}{-\log r}.$$

We proved in [MR11] that the recurrence rates are linked with the pointwise dimensions of the marginal measure ν.

Theorem 3 ([MR11])**.** *Let $\mathcal{T}$ be a random dynamical system on X over $(\Omega, \mathcal{B}(\Omega), \mathbb{P}, \vartheta)$ with an invariant measure μ. For μ-almost every $(\omega, x) \in \Omega \times X$*

$$\underline{R}^\omega(x) \leq \underline{d}_\nu(x) \qquad \textit{and} \qquad \overline{R}^\omega(x) \leq \overline{d}_\nu(x).$$

Proof. This theorem is proved using Theorem 1 applied to the dynamical system $(\Omega \times X, \mathcal{B}(\Omega \times X), \mu, S)$ with the observation f defined by

$$\begin{array}{rcl} f\ :\Omega \times X & \longrightarrow & X \\ (\omega, x) & \longmapsto & x. \end{array}$$

With this observation, for all $(\omega, x) \in \Omega \times X$ and for all $r > 0$ we identify the return time for the observation

$$\tau_r^f(\omega, x) = \tau_r^\omega(x),$$

the pushforward measure

$$f_*\mu = \nu,$$

and the pointwise dimensions

$$\underline{d}_{f_*\mu}(f(\omega, x)) = \underline{d}_\nu(x) \qquad \text{and} \qquad \overline{d}_{f_*\mu}(f(\omega, x)) = \overline{d}_\nu(x).$$

□

Even if the inequalities in Theorem 3 can be strict, with more assumptions on the random dynamical system one can prove that the equalities hold. This drives us to introduce the decay of correlations for a random dynamical system:

Definition 2.5. *A random dynamical system has a super-polynomial decay of correlations if for all $n \in \mathbb{N}^*$, all ψ Lipschitz observables from X to $\mathbb{R}$ and all ϕ measurable bounded functions from X to $\mathbb{R}$*

$$\left|\int_X \int_\Omega \psi(x)\phi(T_{\vartheta^{n-1}\omega} \circ \ldots \circ T_\omega x) d\mu(\omega, x) - \int_X \psi d\nu \int_X \phi d\nu\right| \leq \|\psi\|_{Lip} \|\phi\|_\infty \theta_n$$

with $\lim_{n\to\infty} \theta_n n^p = 0$ for any $p > 0$.

As in Section 2.1 we will assume some kind of aperiodicity condition:

Definition 2.6. *The random dynamical system $\mathcal{T}$ on X over $(\Omega, \mathcal{B}(\Omega), \mathbb{P}, \vartheta)$ with invariant measure μ is called random-aperiodic if*

$$\mu\{(\omega, x) \in \Omega \times X : \exists n \in \mathbb{N},\, T_{\vartheta^{n-1}\omega} \circ \ldots \circ T_\omega x = x\} = 0.$$

Let $T_0 = \mathcal{R}_\alpha$ an irrational rotation of the circle for an irrational number α and T_1 the identity map of the circle. The i.i.d. random dynamical system constructed with this two maps chosen with the same probability $P(0) = P(1) = \frac{1}{2}$ is not random-aperiodic and we have $\mathbb{P}\{\omega \in \Omega \ : \ \tau_r^\omega(x) = 1\} = P(1) = \frac{1}{2}$ for all $r > 0$. This means that after one iteration half of the points did not rotate and $\left(\frac{1}{2}\right)^n$ of them after n iterations. Anyway, almost every point will eventually rotate and their dynamics will be quite interesting providing that we wait long enough. To avoid this kind of problem with the first return time for non-random-aperiodic system, non-instantaneous return times were introduced in [MR11](more details are also presented on the Section 2.3 of [RS10]).

Theorem 4 ([MR11])**.** *Let $\mathcal{T}$ be a random dynamical system on X over $(\Omega, \mathcal{B}(\Omega), \mathbb{P}, \vartheta)$ with an invariant measure μ. If the random dynamical system has a super-polynomial decay of correlations and is random-aperiodic then*

$$\underline{R}^\omega(x) = \underline{d}_\nu(x) \qquad \textit{and} \qquad \overline{R}^\omega(x) = \overline{d}_\nu(x)$$

for μ-almost every $(\omega, x) \in \Omega \times X$ such that $\underline{d}_\nu(x) > 0$.

We will now applied our results to some non-i.i.d. random expanding maps. Other examples such as random hyperbolic toral automorphisms, random perturbations of expanding maps of the circle and random perturbations of piecewise expanding maps of the interval where also given in [MR11].

Example 2.7 (Non-i.i.d. random expanding maps)**.** *Let T_1 and T_2 be the two following maps defined on the one-dimensional torus $X = \mathbb{T}^1$:*

$$\begin{array}{rcl} T_1 : X & \longrightarrow & X \\ x & \longmapsto & 2x \end{array} \qquad \textit{and} \qquad \begin{array}{rcl} T_2 : X & \longrightarrow & X \\ x & \longmapsto & 3x. \end{array}$$

The dynamic of the random dynamical system is given by the following skew product

$$\begin{array}{rcl} S : \Omega \times X & \longrightarrow & \Omega \times X \\ (\omega, x) & \longmapsto & (\vartheta(\omega), T_\omega x) \end{array}$$

with $\Omega = [0,1]$, $T_\omega = T_1$ if $\omega \in [0, 2/5)$, $T_\omega = T_2$ if $\omega \in [2/5, 1]$ and where ϑ is the following piecewise linear map

$$\vartheta(\omega) = \begin{cases} 2\omega & \textit{if} \ \ \omega \in [0, 1/5) \\ 3\omega - 1/5 & \textit{if} \ \ \omega \in [1/5, 2/5) \\ 2\omega - 4/5 & \textit{if} \ \ \omega \in [2/5, 3/5) \\ 3\omega/2 - 1/2 & \textit{if} \ \ \omega \in [3/5, 1]. \end{cases}$$

In fact, one can observe that the random orbit is constructed by choosing the map T_1 and T_2 following a Markov process with the stochastic matrix

$$A = \begin{pmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{pmatrix}.$$

It was proved in [MR11] that S is $Leb \otimes Leb$-invariant and that this random system satisfies the assumptions of Theorem 4, thus for $Leb \otimes Leb$-almost every $(\omega, x) \in [0,1] \times \mathbb{T}^1$

$$\lim_{r \to 0} \frac{\log \tau_r^\omega(x)}{-\log r} = 1.$$

2.3 Recurrence rates for flows

It is natural to wonder if the results of Section 2.1 can be extend to continuous time and more precisely if one can obtain quantitative results of recurrence for flows and more generally for observations of flows. Thus, we studied in [R12] the recurrence rates for flows and observations of flows.

To avoid the introduction of some complex definitions and notions, we will only treat here a special observation (the projection on the manifold for the geodesic flow), but we refer the reader to Section 2 and 3 of [R12] for more general results on the recurrence rate for observation of flows presenting some hyperbolic behaviour.

Let M be a compact Riemannian manifold and d the Riemannian metric. Let Ψ be a flow on M. Let ν be a probability measure on M invariant for the flow Ψ. We introduce the notion of return time and recurrence rates for flows:

Definition 2.8. *We define for $x \in M$ the return time of the flow Ψ:*

$$\tau_r^{\Psi}(x) := \inf\{t > \eta_r(x) \,:\, \Psi_t(x) \in B(x,r)\}$$

where $B(x,r)$ is the ball centered in x and of radius r and $\eta_r(x)$ is the first escape time of the ball $B(x,r)$, i.e. $\eta_r(x) = \inf\{t > 0, \Psi_t x \notin B(x,r)\}$. We define also the lower and upper recurrence rates:

$$\underline{R}^{\Psi}(x) := \liminf_{r\to 0} \frac{\log \tau_r^{\Psi}(x)}{-\log r} \qquad \text{and} \qquad \overline{R}^{\Psi}(x) := \limsup_{r\to 0} \frac{\log \tau_r^{\Psi}(x)}{-\log r}.$$

We showed that these recurrence rates are linked to the local dimension of the invariant measure.

Firstly, we will prove a theorem satisfied for any flow:

Theorem 5 ([R12])**.** *Let Ψ be a differentiable flow on M and ν an invariant probability measure for Ψ. For ν-almost every $x \in M$ which is not a fixed point*

$$\underline{R}^{\Psi}(x) \le \underline{d}_{\nu}(x) - 1 \qquad \text{and} \qquad \overline{R}^{\Psi}(x) \le \overline{d}_{\nu}(x) - 1.$$

To obtain an equality between recurrence rates and dimensions, we need more assumptions on the system:

Theorem 6 ([R12])**.** *Let Ψ be an Anosov flow on M. If ν is an equilibrium state of an Hölder potential, then*

$$\underline{R}^{\Psi}(x) = \underline{d}_{\nu}(x) - 1 \qquad \text{and} \qquad \overline{R}^{\Psi}(x) = \overline{d}_{\nu}(x) - 1$$

for ν-almost every $x \in M$.

We can apply, for example, the previous theorem to the geodesic flow on a smooth manifold with striclty negative curvature. Since the geodesic flow is defined on the unit tangent bundle T^1M, we can also considered a particular observation of this flow: the position on the manifold M. Let Π be the canonical projection:

$$\begin{aligned} \Pi \,: T^1M &\longrightarrow M \\ (p,v) &\longmapsto p. \end{aligned}$$

We study the return time for the canonical projection on the manifold M:

$$\tau_r^{\Psi,\Pi}(p,v) := \inf\{t > r \,:\, \Pi(\Psi_t(p,v)) \in B(p,r)\}.$$

Since Ψ is the geodesic flow on T^1M, the first escape time of the projection of the flow on the manifold of the ball $B(p,r)$ is equal to r for r small enough. We define the recurrence rates for the canonical projection:

$$\underline{R}^{\Psi,\Pi}(p,v) := \liminf_{r\to 0} \frac{\log \tau_r^{\Psi,\Pi}(p,v)}{-\log r} \qquad \text{and} \qquad \overline{R}^{\Psi,\Pi}(p,v) := \limsup_{r\to 0} \frac{\log \tau_r^{\Psi,\Pi}(p,v)}{-\log r}.$$

Theorem 7 ([R12])**.** *Let Ψ be the geodesic flow defined on T^1M and ν an invariant probability measure for Ψ. Then for ν-almost every $(p,v) \in T^1M$*

$$\underline{R}^{\Psi,\Pi}(p,v) \leq \underline{d}_{\Pi_*\nu}(p) - 1 \qquad \textit{and} \qquad \overline{R}^{\Psi,\Pi}(p,v) \leq \overline{d}_{\Pi_*\nu}(p) - 1.$$

Moreover, if M has a strictly negative curvature and if ν is an equilibrium state of an Hölder potential then

$$\underline{R}^{\Psi,\Pi}(p,v) = \underline{d}_{\Pi_*\nu}(p) - 1 \qquad \textit{and} \qquad \overline{R}^{\Psi,\Pi} = \overline{d}_{\Pi_*\nu}(p) - 1$$

for ν-almost every $(p,v) \in T^1M$ non-multiple such that $\underline{d}_{\Pi_*\nu}(p) > 1$.

Since the geodesic flow preserves the Lebesgue measure on T^1M, we can apply Theorem 6 and Theorem 7 to obtain the following noteworthy result:

Corollary 2.9. *Let M be a n-dimensional manifold with strictly negative curvature. Let Ψ be the geodesic flow defined on T^1M. Then for Lebesgue-almost every $(p,v) \in T^1M$*

$$R^{\Psi}(p,v) = 2n - 2$$

and

$$R^{\Psi,\Pi}(p,v) = n - 1.$$

Chapter 3

Fluctuations of the return time and hitting time

After obtaining almost sure results on return times in the previous chapter, we will concentrate here on the fluctuations of the return times and hitting times.

Firstly, we will consider the distribution of return times statistics (RTS) and hitting time statistics (HTS) (we refer the reader to the reviews [26, 3, 74, 48] for detailed information and bibliography on this subject). More precisely, we define the distribution of normalized hitting time by

$$F_A^{hit}(t) = \mu\left(\left\{x \in X : \tau_A(x) > \frac{t}{\mu(A)}\right\}\right)$$

and the distribution of normalized return times by

$$F_A^{ret}(t) = \frac{1}{\mu(A)}\mu\left(\left\{x \in A : \tau_A(x) > \frac{t}{\mu(A)}\right\}\right).$$

We are interested in the convergence in law of the distribution of normalized hitting and return times when $\mu(A) \to 0$ for sets A well-chosen (for example cylinders of a partition).

Haydn, Lacroix and Vaienti [49] proved that the limit of the distribution of the return times exists if and only if the limit of the distribution of the hitting times exists. Moreover, an exponential distribution was proved for various families of dynamical systems: Axiom A diffeomorphisms [52], Markov chains [70], some rational transformations [47], uniformly expanding transformations of the interval [27], and some non-uniformly hyperbolic systems [53, 75]. Recently, Freitas, Freitas and Todd [36, 37] linked hitting time statistics to extreme value theory.

In this chapter, we will concentrate on the distribution of return times and hitting time statistics for observation of dynamical systems and random dynamical systems.

In the last part of the chapter, we will concentrate on large deviation for return time.

Several works already addressed large deviations for return time. Abadi and Vaienti in [5] proved large deviation properties of $\tau(C_n)/n$, where $\tau(C_n)$ is the first return of a n-cylinder to itself.

For the n-th return time τ_A^n into a fixed set A, a large deviation result was considered by Chazottes and Leplaideur [24] (see also [60]) for Axiom A diffeomorphisms with equilibrium states.

In [CRS18], we study the limiting behavior as $r \to 0$ of $\mu\left(\tau_r \geq r^{-d_\mu-\epsilon}\right)$ and $\mu\left(\tau_r \leq r^{-d_\mu+\epsilon}\right)$. This characterization is via asymptotic exponential bound and may be seen as a differentiable version [16].

3.1 Exponential law for observations of dynamical systems

In this section, we will use the setting of Section 2.1 and study the distributions of hitting and return times for observations of dynamical systems.

First of all, we remind that the conditional measure is

$$\nu_A(B) = \frac{\nu(A \cap B)}{\nu(A)}.$$

To obtain an information on the fluctuation of the return time we will need that the system fulfills the assumptions of the Theorem 2 and we will also need an hypothesis on the measure:

I) For $f_*\mu$-almost every $y \in \mathbb{R}^n$, there exist $a > 0$ and $b \geq 0$ such that

$$f_*\mu\left(B\left(y,r\right)\backslash B\left(y,r-\rho\right)\right) \leq r^{-b}\rho^a$$

for any $r > 0$ sufficiently small and any $0 < \rho < r$.

Following the same idea of Section 2.2, we will apply, in Section 3.2, our results to random dynamical systems and give some examples where all the assumptions are fulfilled. Some examples of measure which fulfilled the hypothesis (I) are given in Lemma 44 of [74].

Our theorem on the fluctuations of the return time for the observations is the following:

Theorem 8 ([R14])**.** *Let $(X, \mathcal{A}, \mu, T)$ be a m.p.s with a super-polynomial decay of correlations and $f : X \to \mathbb{R}^N$ a Lipschitz observation such that the system is μ-almost aperiodic for the observation f. If hypothesis* (I) *is satisfied, then for every $t \geq 0$ and μ-almost every $x_0 \in X$ such that $\underline{d}_{f_*\mu}(f(x_0)) > 0$:*

$$\lim_{r\to 0} \mu\left(x \in X, \tau^f_{B(f(x_0),r)}(x) > \frac{t}{f_*\mu\left(B\left(f(x_0),r\right)\right)}\right) = e^{-t}$$

and

$$\lim_{r\to 0} \mu_{f^{-1}B(f(x_0),r)}\left(x \in X, \tau^f_{B(f(x_0),r)}(x) > \frac{t}{f_*\mu\left(B\left(f(x_0),r\right)\right)}\right) = e^{-t}.$$

One can observe that, for rapidly mixing dynamical system, we can applied this theorem to the observation $f = id$ and we obtain some of the results cited in the introduction of this chapter and, in particular, a generalization of Theorem 40 of [74]:

Corollary 3.1. *Let $X \subset \mathbb{R}^N$ and let $(X, \mathcal{A}, \mu, T)$ be a m.p.s with a super-polynomial decay of correlations. Let us suppose that for μ-almost every $y \in X$, there exist $a > 0$ and $b \geq 0$ such that $\mu\left(B\left(y,r\right)\backslash B\left(y,r-\rho\right)\right) \leq r^{-b}\rho^a$ for any $r > 0$ sufficiently small and any $0 < \rho < r$.*

Then for every $t \geq 0$ and for μ-almost every $x_0 \in X$ such that $\underline{d}_\mu(x_0) > 0$:

$$\lim_{r\to 0} \mu\left(x \in X, \tau_{B(x_0,r)}(x) > \frac{t}{\mu\left(B\left(x_0,r\right)\right)}\right) = e^{-t}$$

and

$$\lim_{r\to 0} \mu_{B(x_0,r)}\left(x \in X, \tau_{B(x_0,r)}(x) > \frac{t}{\mu\left(B\left(x_0,r\right)\right)}\right) = e^{-t}.$$

.

One can remark that, in this corollary, we do not need the assumption on the μ-almost aperiodicity since the system is mixing.

Remark 3.2. *We observe that if a super-polynomial decay of correlations implies the convergence to an exponential law, a polynomial decay may not be enough. Indeed, in [GRS15], a family of systems with polynomial mixing were studied but it was shown that for these systems the hitting time and return time distributions to balls do not converge to the exponential law.*

3.2 Annealed exponential law for random dynamical systems

In this section, we will use our results for observations of dynamical systems to prove an exponential law for random dynamical systems presenting some rapidly mixing conditions. We will follow the setting of Section 2.2.

We emphasize that in this section we are interested in obtaining an annealed exponential law: we fixed a point z, studied the return/hitting time around the target z and proved an exponential law with respect to the invariant measure μ. In the next section, we will be interested in quenched exponential law: both the target z and the realization ω are fixed, and the exponential law is proved with respect to the sample measures μ_ω.

As in the previous section, we will need an assumption on the measure and an assumption on the decay of correlations:

a) For ν-almost every $x \in X$, there exist $a > 0$ and $b \geq 0$ such that

$$\nu\left(B\left(x,r\right)\backslash B\left(x,r-\rho\right)\right) \leq r^{-b}\rho^{a}$$

for any $r > 0$ sufficiently small and any $0 < \rho < r$.

b) For all $n \in \mathbb{N}^*$, ψ Lipschitz observables from X to $\mathbb{R}$ and φ measurable bounded from $\Omega \times X$ to $\mathbb{R}$

$$\left|\int_{\Omega\times X} \psi(x)\varphi(S^n(\omega,x))d\mu - \int_X \psi d\nu \int_{\Omega\times X} \varphi d\mu\right| \leq \|\psi\|_{Lip}.\|\varphi\|_\infty.\theta_n$$

with $\lim_{n\to\infty}\theta_n n^p = 0$ for any $p > 0$.

We observe that this last assumption is weaker than assuming super-polynomial decay of correlations for the skew-product, however it is stronger than the assumption used for the recurrence rates in Theorem 4.

Theorem 9 ([R14])**.** *Let $\mathcal{T}$ be a random dynamical system on X over $(\Omega, \mathcal{B}(\Omega), \mathbb{P}, \vartheta)$ with an invariant measure μ. If the random dynamical system is random-aperiodic and satisfied hypothesis* (a) *and* (b) *then for every $t \geq 0$ and for ν-almost every $z \in X$ such that $\underline{d}_\nu(z) > 0$:*

$$\lim_{r\to 0}\mu\left((\omega,x)\in\Omega\times X, \tau^{\omega}_{B(z,r)}(x) > \frac{t}{\nu\left(B\left(z,r\right)\right)}\right) = e^{-t}$$

and

$$\lim_{r\to 0}\mu_{\Omega\times B(z,r)}\left((\omega,x)\in\Omega\times X,\tau^{\omega}_{B(z,r)}(x)>\frac{t}{\nu\left(B\left(z,r\right)\right)}\right)=e^{-t}.$$

The basic idea to prove this theorem is the same as the one explained in the proof of Theorem 3.

For i.i.d. random dynamical systems, to obtain an exponential law, we just need to assume a super-polynomial decay of correlations for the random dynamical system, i.e. our observables are from X to $\mathbb{R}$, which is a more natural assumption than hypothesis (b).

More precisely, let $\{T_\lambda\}_{\lambda\in\Lambda}$ be a family of transformations defined on a compact Riemannian manifold X and let $\mathcal{P}$ be a probability measure on a metric space Λ. We will consider $\mathcal{T}$ a random dynamical system on X over $(\Lambda^{\mathbb{N}},\mathcal{P}^{\mathbb{N}},\sigma)$ with a stationary measure ν, where σ is the shift. That is, for an *i.i.d.* stochastic process $\underline{\lambda}=(\lambda_n)_{n\geq 1}\in\Lambda^{\mathbb{N}}$ with common distribution $\mathcal{P}$, a random evolution of an initial state $x\in X$ will be:

$$T^n_{\underline{\lambda}}x=T_{\lambda_n}\circ\ldots T_{\lambda_1}x$$

for every $n\geq 0$.

Definition 3.3. *The i.i.d. random dynamical system has a super-polynomial decay of correlations if, for all $n\in\mathbb{N}^*$, ψ Lipschitz observables from X to $\mathbb{R}$ and φ measurable bounded from X to $\mathbb{R}$*

$$|\int_{\Lambda^{\mathbb{N}}\times X}\psi(x)\varphi(T^n_{\underline{\lambda}}x)d\mathcal{P}^{\mathbb{N}}d\nu-\int_X\psi d\nu\int_X\varphi d\nu|\leq\|\psi\|_{Lip}.\|\varphi\|_\infty.\theta_n$$

with $\lim_{n\to\infty}\theta_n n^p=0$ *for any* $p>0$*.*

Theorem 10 ([R14])**.** *Let $\mathcal{T}$ be an i.i.d. random dynamical system on X over $(\Lambda^{\mathbb{N}},\mathcal{P}^{\mathbb{N}},\sigma)$ with a stationary measure ν. If the random dynamical system is random-aperiodic, satisfied hypothesis* (a) *and has a super-polynomial decay of correlations then for every $t\geq 0$ and for ν-almost every $z\in X$ such that $\underline{d}_\nu(z)>0$:*

$$\lim_{r\to 0}\mathcal{P}^{\mathbb{N}}\otimes\nu\left((\underline{\lambda},x)\in\Lambda^{\mathbb{N}}\times X,\tau^{\underline{\lambda}}_{B(z,r)}(x)>\frac{t}{\nu\left(B\left(z,r\right)\right)}\right)=e^{-t}$$

and

$$\lim_{r\to 0}\mathcal{P}^{\mathbb{N}}\otimes\nu_{\Lambda^{\mathbb{N}}\times B(z,r)}\left((\underline{\lambda},x)\in\Lambda^{\mathbb{N}}\times X,\tau^{\underline{\lambda}}_{B(z,r)}(x)>\frac{t}{\nu\left(B\left(z,r\right)\right)}\right)=e^{-t}.$$

Remark 3.4. *We emphasize that this result extends the result of [13] for randomly perturbed dynamical systems. The principal generalization lies in the decay of correlations.*

First of all, they need polynomial decay of correlations against L^1 observables when here we just need super-polynomial decay of correlations against L^∞ observables. Besides, for the observables ψ, we do not assume that indicator functions of balls are bounded in the Banach space.

Moreover, they study randomly perturbed dynamical systems, more precisely, they perturbed an original map with random additive noise when in our setting we can study more general random dynamical systems, as shown in the following examples.

As in Section 2.2, we will apply our results to the non-i.i.d. random expanding maps defined in Example 2.7. Other examples, such as random circle maps expanding in average and randomly perturbed dynamical systems were given in [R14, Section 3].

Example 3.5 (Non-i.i.d. random expanding maps). *As in Example 2.7, let T_1 and T_2 be the two following maps defined on the one-dimensional torus $X = \mathbb{T}^1$:*

$$\begin{array}{rcl} T_1 : X & \longrightarrow & X \\ x & \longmapsto & 2x \end{array} \quad \textit{and} \quad \begin{array}{rcl} T_2 : X & \longrightarrow & X \\ x & \longmapsto & 3x \end{array}$$

chosen following a Markov process with the stochastic matrix

$$A = \begin{pmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{pmatrix}.$$

For every $t \geq 0$ and for Leb-almost every $z \in \mathbb{T}^1$:

$$\lim_{r\to 0} Leb \otimes Leb \left(\tau^{\omega}_{B(z,r)}(x) > \frac{t}{r} \right) = e^{-t}$$

and

$$\lim_{r\to 0} Leb \otimes Leb_{[0,1]\times B(z,r)} \left(\tau^{\omega}_{B(z,r)}(x) > \frac{t}{r} \right) = e^{-t}.$$

3.3 Quenched exponential law for random dynamical systems

In this section, we will give results on quenched HTS/RTS for random dynamical systems. The more complete results are for random subshifts of finite type and will be treated in Section 3.3.1. In Section 3.3.2 we will consider random dynamical systems on manifolds modelled by a skew product which have certain geometric properties and whose measures satisfy quenched decay of correlations at a sufficient rate.

3.3.1 Random subshifts of finite type

We will follow in this section the definition of random subshifts of finite type given in Section 1.3

First of all we will assume that the random variable b is such that $\mathbb{E}(\log b) < \infty$. This hypothesis on b guarantees that the metric entropy $h_\mu(S, \Omega \times \mathcal{F}_0^1)$ is finite and we will denote it by h.

We assume the following: there are constants $h \geq h_0 > 0$, $c > 0$, a random variable $C \in L^p(\Omega, \mathbb{P})$ for some $p \in (0, 1]$ and a summable function $\alpha(g)$ such that for all m, n, $A \in \mathcal{F}_0^n$ and $B \in \mathcal{F}_0^m$:

(I) the marginal measure ν satisfies

$$\left| \nu(A \cap \sigma^{-g-n} B) - \nu(A)\nu(B) \right| \leq \alpha(g);$$

(II) $\mu_\omega(C^n(y)) \le ce^{-h_0 n}$ for any $y \in X$ and $n \ge 1$, for $\mathbb{P}$-almost every $\omega \in \Omega$;

(III) For $\mathbb{P}$-almost every $\omega \in \Omega$

$$\left|\mu_\omega(A \cap \sigma^{-g-n}B) - \mu_\omega(A)\mu_{\theta^{n+g}\omega}(B)\right| \le C(\omega)\alpha(g).$$

Theorem 11 ([RSV14]). *We assume that hypothesis (I), (II) and (III) hold and that there exists a constant $q > \frac{h}{h_0}(1+\frac{3}{p})$ such that α satisfies $\alpha(g)g^q \to 0$ when $g \to +\infty$. For ν-almost every z, $\mathbb{P}$-almost every ω and all $t \ge 0$ we have*

$$\lim_{n\to\infty} \mu_\omega\left(\tau_{C_n(z)}(\cdot) > \frac{t}{\nu(C_n(z))}\right) = e^{-t}$$

This can be view as a *quenched* exponential law for hitting time. The later convergence together with integration over Ω and dominated convergence theorem yields the following *annealed* version:

Corollary 3.6. *Under the same hypothesis of Theorem 11, for ν-almost every z and $t \ge 0$,*

$$\lim_{n\to\infty} \nu\left(\tau_{C_n(z)}(\cdot) > \frac{t}{\nu(C_n(z))}\right) = e^{-t}.$$

The limit law in Theorem 11 is only obtain for typical point z, a natural question that arises here is: what about non-typical points? In the deterministic setting, it has been shown in [52, 35, 38] that if the point z is periodic, then the distribution is of the form $e^{-\Theta t}$ where $\Theta \in (0,1)$ is a parameter which takes into account the amount of repulsion at z. Indeed, in a general subshift of finite type setting in [35], and a more restricted setting in [38], a dichotomy was proved: it was shown that the limit exists for any z, and is $e^{-\Theta t}$ for $\Theta \in (0,1)$ in the case that z is periodic, and $\Theta = 1$ in the case when z is non-periodic. Some of these results were motivated by the connection of HTS laws to Extreme Value Laws (EVL), see [28, 36], one reason why Θ can be referred to as the *extremal index*.

Thus, in [RT15], we obtain a limit for every point and proved a dichotomy similar to the one in the discrete case. To do so, we will need stronger assumptions. We assume that there are constants $h_1 > 0$ and $c_0 > 0$ such that for all $m, n, g \in \mathbb{N}$, $A \in \mathcal{F}_0^n$ and $B \in \mathcal{F}_0^m$:

(III-a) for $\mathbb{P}$-almost every $\omega \in \Omega$,

$$\left|\mu_\omega(A \cap \sigma^{-g-n}B) - \mu_\omega(A)\mu_{\theta^{n+g}\omega}(B)\right| \le \alpha(g)\mu_\omega(A)\mu_{\theta^{n+g}\omega}(B);$$

(IV) for $\mathbb{P}$-almost every $\omega \in \Omega$, if $y \in X_\omega$ and $n \ge 1$ then $c_0^{-1}e^{-h_1 n} \le \nu(C_n(y))$;

(V) the sample measures satisfy

$$\operatorname*{ess-sup}_{\omega\in\Omega} \sup_{x\in X} \mu_\omega(C_1(x)) < 1.$$

With these stronger assumptions, we can obtain the following dichotomy:

Theorem 12 ([RT15]). *Assume (I), (III-a), (IV) and (V) hold and that there exists a constant $q > 2\frac{h_0}{h_1}$ such that α satisfies $\alpha(g)g^q \to 0$ as $g \to +\infty$. Let $z \in X$. Then for $\mathbb{P}$-almost every ω, either*

(a) z is a periodic point of period p and if the limit $\Theta := \lim_{n\to\infty} \frac{\nu(C_n(z)\setminus C_{n+p}(z))}{\nu(C_n(z))}$ exists, then for all $t \geq 0$ we have

$$\lim_{n\to\infty} \mu_\omega \left(\tau_{C_n(z)}(\cdot) > \frac{t}{\nu(C_n(z))} \right) = e^{-\Theta t};$$

or

(b) for all $t \geq 0$ we have

$$\lim_{n\to\infty} \mu_\omega \left(\tau_{C_n(z)}(\cdot) > \frac{t}{\nu(C_n(z))} \right) = e^{-t}.$$

This dichotomy can be compared with the dichotomy for deterministic systems, see for example [1] or [13, Theorem A]. Analogously to Corollary 3.6 we easily obtain the following annealed law.

Corollary 3.7. *Under the same hypothesis of Theorem 12, for $z \in X$, either*

(a) z is a periodic point of period p and if the limit $\Theta := \lim_{n\to\infty} \frac{\nu(C_n(z)\setminus C_{n+p}(z))}{\nu(C_n(z))}$ exists, then for all $t \geq 0$ we have

$$\lim_{n\to\infty} \nu \left(\tau_{C_n(z)}(\cdot) > \frac{t}{\nu(C_n(z))} \right) = e^{-\Theta t};$$

or

(b) for all $t \geq 0$ we have

$$\lim_{n\to\infty} \nu \left(\tau_{C_n(z)}(\cdot) > \frac{t}{\nu(C_n(z))} \right) = e^{-t}.$$

Remark 3.8. *1. In both of the above results, for the standard examples the limit Θ does indeed exist [RT15][Section 5].*

2. Theorem 12 works only in the case of finite alphabet since assumption (II) cannot be fulfilled otherwise. A stronger mixing assumption for the marginal measure μ allowed us to treat the case of infinite alphabets in [RT15].

3. One can prove [RT15][Lemma 2.1] that assuming (III-a) and (V) implies that (II) holds.

Here, we will apply our results to random Bernoulli shifts. Random Gibbs Measures were also treated in [RT15].

Example 3.9 (Random Bernoulli shifts)**.** *Let $s \geq 1$ and (Ω, θ) be a subshift of finite type on the symbolic space $\{0, 1, \dots, s\}^{\mathbb{Z}}$ and let $\mathbb{P}$ be a Gibbs measure from a Hölder potential.*

Let $b \geq 1$ and make the shift $\{0, 1, \dots, b\}^{\mathbb{N}}$ a random subshift by putting on it the random Bernoulli measures constructed as follows. Let $W = (w_{ij})$ be a $s \times b$ stochastic matrix with entries in $(0, 1)$. Set $p_j(\omega) = w_{\omega_0, j}$. The random Bernoulli measure μ_ω is defined by

$$\mu_\omega([x_0 \dots x_n]) = p_{x_0}(\omega) p_{x_1}(\theta\omega) \dots p_{x_n}(\theta^n \omega).$$

We proved in [RSV14] and [RT15] that one can apply Theorem 12.

We also observe that, for example, when the base is i.i.d., for z is a periodic point of period p, we can compute the extremal index:

$$\Theta = \int (1 - p_{z_0}(\omega)) d\mathbb{P} \cdots \int (1 - p_{z_{p-1}}(\omega)) d\mathbb{P}.$$

3.3.2 Random maps

We will follow in this section the definition of random dynamical system given in Section 1.2.

Moreover, we will assume that $\vartheta : \Omega \to \Omega$ is the two-sided shift map on a full shift space Ω and that the measure $\mathbb{P}$ is ergodic. We also assume that X is a compact manifold and for every $\omega \in \Omega$, $T_\omega = T_{\omega_0}$.

For every realisation $\omega \in \Omega$, let $\Gamma^u(\omega)$ be a collection of unstable leaves $\gamma^u(\omega)$ and $\Gamma^s(\omega)$ a collection of stable leaves $\gamma^s(\omega)$. We assume that $\gamma^u \cap \gamma^s$ consists of a single point for all $(\gamma^u, \gamma^s) \in \Gamma^u \times \Gamma^s$. The map T_ω contracts along the stable leaves and similarly the inverse branches of T_ω contract along the unstable leaves.

For an unstable leaf $\gamma^u(\omega)$ denote by $\mu_\omega^{\gamma^u}$ the disintegration of μ_ω with respect to the γ^u. We assume that μ_ω has a product like decomposition $d\mu_\omega = d\mu_\omega^{\gamma^u} d\upsilon_\omega(\gamma^u)$, where υ_ω is a transversal measure. That is, if f is a function on X then

$$\int f(x)\, d\mu_\omega(x) = \int_{\Gamma^u(\omega)} \int_{\gamma^u} f(x)\, d\mu_\omega^{\gamma^u}(x)\, d\upsilon_\omega(\gamma^u).$$

If $\gamma^u, \hat{\gamma}^u \in \Gamma^u(\omega)$ are two unstable leaves then the holonomy map $\mathcal{H} : \gamma^u \to \hat{\gamma}^u$ is defined such that by $\mathcal{H}(x)$ is the unique point of intersection between $\hat{\gamma}^u$ and $\gamma^s(x)$ for $x \in \gamma^u$, where $\gamma^s(x)$ is the local stable leaf through x.

Let us denote by $J_n^\omega = \frac{dT_\omega^n \mu_\omega^{\gamma^u}}{d\mu_\omega^{\gamma^u}}$ the Jacobian of the map T_ω^n with respect to the measure μ_ω in the unstable direction.

Fix ω and let γ^u be a local unstable leaf. Assume there exist $R > 0$ and for every $n \in \mathbb{N}$ finitely many $y_k \in T_\omega^n \gamma^u$ so that $T_\omega^n \gamma^u \subset \bigcup_k B_{R,\gamma^u}(y_k)$, where $B_{R,\gamma^u}(y)$ is the embedded R-disk centered at y in the unstable leaf γ^u. Denote by $\zeta_{\varphi,k} = \varphi(B_{R,\gamma^u}(y_k))$ where $\varphi \in \mathcal{I}_n^\omega$ and $\mathcal{I}_n^\omega$ denotes the inverse branches of T_ω^n. We call ζ an n-cylinder. Then there exists a constant L so that the number of overlaps $N_{\varphi,k} = |\{\zeta_{\varphi',k'} : \zeta_{\varphi,k} \cap \zeta_{\varphi',k'} \neq \varnothing, \varphi' \in \mathcal{I}_n^\omega\}|$ is bounded by L for all $\varphi \in \mathcal{I}_n^\omega$ and for all k

and n. This follows from the fact that $N_{\varphi,k}$ equals $|\{k' : B_{R,\gamma^u}(y_k) \cap B_{R,\gamma^u}(y_{k'}) \neq \varnothing\}|$ which is uniformly bounded by some constant L.

To obtain an exponential law for the distribution of hitting time and return time, we need a few assumptions. First of all, we need information on the annealed and quenched decay of correlations:
(I) There exists a decay function $\lambda(k)$ so that

$$\left|\int_\Omega \int_X \psi(x)\phi(T_\omega^k x)\, d\mu(\omega, x) - \int_X \psi d\nu \int_X \phi d\nu\right| \leq \lambda(k)\|\psi\|_{Lip}\|\phi\|_\infty \qquad \forall k \in \mathbb{N}$$

for every $\psi \in Lip(X, \mathbb{R})$ and $\phi \in L^\infty(X, \mathbb{R})$.
(II) For $\mathbb{P}$-almost every ω, the individual measure μ_ω has the following decay of correlations

$$\left|\int_X \psi(x)\phi(T_\omega^k x)\, d\mu_\omega(x) - \int_X \psi d\mu_\omega \int_X \phi d\mu_{\theta^k\omega}\right| \leq \lambda(k)\|\psi\|_{Lip}\|\phi\|_\infty \qquad \forall k \in \mathbb{N},$$

for every $\phi \in L^\infty(X, \mathbb{R})$ which are constant on local stable leaves γ^s of T_ω and for every $\psi \in Lip(X, \mathbb{R})$.
Then, we need some geometric assumptions:
(III) (Distortion) For $\mathbb{P}$-almost every ω, we require that $\frac{J_n^\omega(x)}{J_n^\omega(y)} = \mathcal{O}(\Theta(n))$ for all $x, y \in \zeta$ and n, where ζ are n-cylinders in unstable leaves γ^u and Θ is a non-decreasing function which below we assume to be $\Theta(n) = \mathcal{O}(n^{\kappa'})$ for some $\kappa' \geq 0$.
(IV) (Contraction) There exists a function $\delta(n) \to 0$ which decays at least summably polynomially, i. e. $\delta(n) = \mathcal{O}(n^{-\kappa})$ with $\kappa > 1$, so that $\operatorname{diam}\zeta \leq \delta(n)$ for all n-cylinder ζ and all n and ω.
Finally, we need some information on the measures:
(V) There exist $0 < d_0 < d_1$ and K such that $r^{d_0} \geq \nu(B(z, r)) \geq r^{d_1}$ and

$$\frac{1}{K} \leq \frac{\nu(B(z, r))}{\mu_\omega(B(z, r))} \leq K$$

for all $r > 0$ small enough, for μ_ω-almost every $z \in X$ and $\mathbb{P}$-almost every $\omega \in \Omega$.
(VI) (Annulus condition) Assume that for some $\xi \geq \beta > 0$:

$$\sup_\omega \frac{\mu_\omega(B(z, r+\rho) \setminus B(z, r-\rho))}{\nu(B(z, r))} = \mathcal{O}(\frac{\rho^\xi}{r^\beta})$$

for every $\rho < r$.

We can now state the main results of [HRY20]. Here $\mu_\omega|_B$ is the conditional measure of μ_ω restricted to the set $B \subset X$. Under the previous assumption we obtain an exponential law for the distribution of hitting times and return times.

Theorem 13 ([HRY20])**.** *Let a random dynamical system satisfy the above requirements (I)–(VI) where δ and λ both decay super-polynomially fast.*

Then

$$\lim_{r\to 0} \mu_\omega\left(x \in X : \tau_{B(z,r)}^\omega(x) > \frac{t}{\nu(B(z, r))}\right) = e^{-t}$$

and

$$\lim_{r\to 0} \mu_\omega|_{B(z,r)}\left(x \in X : \tau^\omega_{B(z,r)}(x) > \frac{t}{\nu(B(z,r))}\right) = e^{-t}$$

for all $t > 0$, for μ_ω-almost every $z \in X$ and $\mathbb{P}$-almost every $\omega \in \Omega$.

It is important to notice that this was the first paper where a quenched law was proved for the return times. Indeed, in [RSV14, RT15, 39], the law was only obtained for the hitting times. This is a significant difference with the deterministic setting where if a limiting distribution exists for the hitting times, then it also exists for the return times (and the other way around) [49].

One can see in [HRY20, Theorem 2.3] that even if δ and λ do not decay super-polynomially fast, one can still obtain an exponential distribution assuming some technical conditions on the constants present in the hypothesis (I)–(VI).

This theorem was applied in [HRY20] to random C^2 maps of the interval, random parabolic maps on the unit interval and random perturbation of partially hyperbolic attractors on a compact Riemannian manifold.

3.4 Large deviation estimates for return times

In this section, we will obtain large deviation estimates for return times, in the deterministic setting (i.e. we follow the setting of Section 2.1).

We define the rate functions which will appear in our large deviations estimates. The first one is related to the deviations in the pointwise dimension; it has been computed in [66] in the case of conformal repellers.

Definition 3.10. *The exponential rate for dimension is defined for $\epsilon > 0$ by:*

$$\underline{\psi}(\pm\epsilon) = \varliminf_{r\to 0} \frac{1}{\log r} \log \mu\left(\left\{\frac{\log \mu(B(x,r))}{-\log r} \in I_{\pm\epsilon}\right\}\right),$$

where $I_\epsilon = (-\infty, -d_\mu - \epsilon)$ and $I_{-\epsilon} = (-d_\mu + \epsilon, +\infty)$.

The second quantifies the probability of quick returns near the origin.

Definition 3.11. *The exponential rate for fast return times is defined for $\epsilon, a > 0$ by:*

$$\underline{\varphi}(a,\epsilon) = \varliminf_{r\to 0} \frac{1}{\log r} \log \mu\left(\left\{x_0 : \mu_{B(x_0,2r)}\left(\tau_{B(x_0,2r)} \le r^{-d_\mu+\epsilon}\right) \ge Cr^a\right\}\right), \tag{3.1}$$

for some constant $C > 0$.

We may now state the main result proved in [CRS18]. We emphasize that the value of C in (3.1) is irrelevant in the theorem.

Theorem 14 ([CRS18]). *Let $(X, \mathcal{A}, \mu, T)$ be a m.p.s. Suppose that μ is an exact dimensional measure. Given $\epsilon > 0$, we have:*

$$\varliminf_{r\to 0}\frac{1}{\log r}\log\mu\left(\tau_r \geq r^{-d_\mu-\epsilon}\right) \geq \max_{\gamma\in(0,1)}\min\left\{(1-\gamma)\epsilon, \underline{\psi}(\gamma\epsilon)\right\} \tag{3.2}$$

and

$$\varliminf_{r\to 0}\frac{1}{\log r}\log\mu\left(\tau_r \leq r^{-d_\mu+\epsilon}\right) \geq \max_{\substack{\gamma\in(0,1)\\ a,\epsilon''>0}}\min\left\{-\gamma\epsilon-\epsilon''+a, \underline{\psi}(\gamma\epsilon), \underline{\varphi}(a,\epsilon), \underline{\psi}(-\epsilon'')\right\}.$$

This result is satisfactory in the sense that it can be applied to a broad class of dynamical systems, provided one can estimate the rate functions $\underline{\psi}$ and $\underline{\varphi}$.

The rate function for dimension $\underline{\psi}$ is rather classical. We can observe that in (3.2) if the rate function for dimension $\underline{\psi}$ is positive in some interval $(0, \epsilon)$, it readily implies that $\mu\left(\tau_r \geq r^{-d_\mu-\epsilon}\right)$ has a fast decay.

The rate function $\underline{\varphi}$ is not so well known. However, for several dynamical systems an estimation of the error in the approximation to the exponential law for return time has been computed. In many cases, including Hénon maps [23, Theorem 3.1], it is possible to show that for some $a, b > 0$, and any sufficiently small $r > 0$,

E1 there exists a set $\Omega_r \subset X$ such that $\mu(\Omega_r^c) < r^b$;

E2 for all $x \in \Omega_r$,

$$\sup_{t\geq 0}\left|\mu_{B(x,r)}\left(\tau_{B(x,r)} > \frac{t}{\mu(B(x,r))}\right) - e^{-t}\right| \leq r^a.$$

The conditions E1-E2 imply that $\underline{\varphi}(a,\epsilon) \geq \min\{\underline{\psi}(a-\epsilon), b\}$ [CRS18][Proposition 4.2].

Finally, we observe that in [CRS18] we apply our result to $C^{1+\alpha}$ conformal repeller with an equilibrium state of a Hölder potential. Then, we compute the rate functions and obtain large deviation estimates for return times for repeller.

Chapter 4

Shortest distance between orbits

In this chapter we study the behaviour of the shortest distance between k orbits, i.e. for $(x_1,\ldots,x_k)\in X^k$:

$$m_n(x_1,\ldots,x_k)=\min_{i_1,\ldots,i_k=0,\ldots,n-1}\left(d(T^{i_1}x_1,\ldots,T^{i_k}x_k)\right),$$

where $d(x_1,\ldots,x_k)=\max_{i\neq j} d(x_i,x_j)$, and show a relation between this shortest distance and the generalized fractal dimensions.

Even if the shortest distance between orbits seems to be something natural to define and study, to the best of our knowledge, it was done for the first time in our article [BLR19]. One can observe that this quantity shares some similarities with the correlation sum and the correlation integral of the Grassberger-Procaccia algorithm [44, 45] and the nearest neighbour analysis [29], with the synchronization of coupled map lattices [33], with dynamical extremal index [34], with the connectivity, proximality and recurrence gauges defined by Boshernitzan and Chaika [20] and also with logarithm laws and shrinking target properties (see e.g. the survey [12]). One can also remark that information on the hitting time can give information on the shortest distance. Indeed, if $\tau_{B(y,r)}(x)\leq n$, we have $m_n(x,y)<r$. Moreover, we will explain in the next chapter the connection between the shortest distance and a well-known probabilistic problem: the longest common substring problem.

4.1 Shortest distance between k orbits

In this section, we will denote by μ^k the product measure $\mu\otimes\cdots\otimes\mu$.

First of all, we observe that defining a distance between more than two points is not classical. Here, we will use the following distance $d(x_1,\ldots,x_k)=\max_{i\neq j} d(x_i,x_j)$. However, other definitions could have been chosen for $d(x_1,\ldots,x_k)$ without altering our results (see e.g. [77] and references therein for examples of generalizations of the usual two-way distance). For example, we could have used $d_1(x_1,\ldots,x_k)=\min_{z\in X}\max_i d(x_i,z)$, or $d_2(x_1,\ldots,x_k)=\sqrt{\sum_{i\neq j} d(x_i,x_j)^2}$ but our results would have been the same since d, d_1, and d_2 are equivalent.

We will show that the behaviour of m_n as $n\to\infty$ is linked with the generalized fractal dimension. The case $k=2$, that is the study of the shortest distance between 2 orbits, was treated in [BLR19] while the case $k\geq 2$ was treated in [BR21].

Theorem 15 ([BLR19, BR21]). *Let $(X,\mathcal{A},\mu,T)$ be a measure preserving system such that $\underline{D}_k(\mu)>0$. Then for μ^k-almost every $(x_1,\ldots,x_k)\in X^k$,*

$$\varlimsup_{n\to+\infty}\frac{\log m_n(x_1,\ldots,x_k)}{-\log n}\leq\frac{k}{(k-1)\underline{D}_k(\mu)}.$$

This general result can be applied to any dynamical system such that $\underline{D}_k(\mu)>0$. Even if the inequality in Theorem 15 can be strict (noting for example the trivial case when T is the identity), we will prove that an equality holds under some rapidly mixing conditions:

(H1) There exists a Banach space $\mathcal{C}$, such that for all $\psi,\ \phi \in \mathcal{C}$ and for all $n \in \mathbb{N}^*$, we have

$$\left|\int_X \psi.\phi \circ T^n \, d\mu - \int_X \psi d\mu \int_X \phi d\mu\right| \leq \|\psi\|_{\mathcal{C}} \|\phi\|_{\mathcal{C}} \theta_n,$$

with $\theta_n = a^n$ $(0 \leq a < 1)$ and where $\|\cdot\|_{\mathcal{C}}$ is the norm in the Banach space $\mathcal{C}$.

(H2) There exist $0 < r_0 < 1$, $c \geq 0$ and $\xi \geq 0$ such that for every $p \in \{1, \ldots, k\}$, for μ^{k-p}-almost every $x_{p+1}, \ldots, x_k \in X$ and any $0 < r < r_0$, the function $\psi_p : X \to \mathbb{R}$, defined below, belongs to the Banach space $\mathcal{C}$ and verify

$$\|\psi_p\|_{\mathcal{C}} \leq c r^{-\xi}.$$

Fixed $x_2, \ldots, x_k \in X$, we define

$$\psi_1(x) = \prod_{j=2}^{k} \mathbb{1}_{B(x_j, r)}(x).$$

For $p > 1$, we fix $x_{p+1}, \ldots, x_k \in X$, and set

$$\psi_p(x) = \bar{\psi}(x, x_{p+1}, \ldots, x_k), \text{ where}$$

$$\begin{aligned}&\bar{\psi}(x_p, x_{p+1}, \ldots, x_k) \\ &= \prod_{l=p+1}^{k} \mathbb{1}_{B(x_l, r)}(x_p) \int_{X^{p-1}} \left[\prod_{j=1}^{p-1} \prod_{l=j+1}^{k} \mathbb{1}_{B(x_j, r)}(x_l)\right] d\mu^{p-1}(x_1, \ldots, x_{p-1}).\end{aligned}$$

When the Banach space $\mathcal{C}$ is the space of Hölder functions $\mathcal{H}^\alpha(X, \mathbb{R})$, we will replace our assumption (H2) by an assumption easier to interpret in Theorem 17.

We will also need some topological information on the space X.

Definition 4.1. *A separable metric space (X, d) is called tight if there exist $r_0 > 0$ and $N_0 \in \mathbb{N}$, such that for any $0 < r < r_0$ and any $x \in X$ one can cover $B(x, 2r)$ by at most N_0 balls of radius r.*

We emphasize that any subset of $\mathbb{R}^n$ with the Euclidian metric is tight, any subset of a Riemannian manifold of bounded curvature is tight and that if (X, d) admits a doubling measure then it is tight [46].

Theorem 16 ([BLR19, BR21])**.** *Let $(X, \mathcal{A}, \mu, T)$ be a measure preserving system, such that (X, d) is tight, satisfying (H1) and (H2) and such that $\underline{D}_k(\mu) > 0$. Then for μ^k-almost every $(x_1, \ldots, x_k) \in X^k$,*

$$\varliminf_{n \to +\infty} \frac{\log m_n(x_1, \ldots, x_k)}{-\log n} \geq \frac{k}{(k-1)\overline{D}_k(\mu)}.$$

Moreover, if $D_k(\mu)$ exists then for μ^k-almost every $(x_1, \ldots, x_k) \in X^k$,

$$\lim_{n \to +\infty} \frac{\log m_n(x_1, \ldots, x_k)}{-\log n} = \frac{k}{(k-1)D_k(\mu)}.$$

When the Banach space $\mathcal{C}$ is the space of Hölder functions $\mathcal{H}^\alpha(X,\mathbb{R})$ we can adapt our proof and (H2) can be replaced by the following condition:

(HA) There exist $r_0>0$, $\xi\geq 0$ and $\beta>0$ such that for μ-almost every $x\in X$ and any $r_0>r>\rho>0$,

$$\mu(B(x,r+\rho)\backslash B(x,r-\rho))\leq r^{-\xi}\rho^\beta.$$

This assumption is satisfied, for example, if the measure is Lebesgue or absolutely continuous with respect to Lebesgue with a bounded density.

Theorem 17 ([BLR19, BR21])**.** *Let $(X,\mathcal{A},\mu,T)$ be a measure preserving system, such that $\underline{D}_k(\mu)>0$ and such that (X,d) is tight, satisfying (H1) with $\mathcal{C}=\mathcal{H}^\alpha(X,\mathbb{R})$ and (HA). Then for μ^k-almost every $(x_1,\ldots,x_k)\in X^k$,*

$$\varliminf_{n\to+\infty}\frac{\log m_n(x_1,\ldots,x_k)}{-\log n}\geq\frac{k}{(k-1)\overline{D}_k(\mu)}.$$

For example, one can apply this theorem to expanding maps of the interval with a Gibbs measure associated to a Hölder potential (see e.g. [74]) and C^2 endomorphism (of a d-dimensional compact Riemannian manifold) admitting a Young tower with exponential tail (see [36, Section 6] and [28]).

We now will apply Theorem 16 to a short list of simple examples and to a more complex family of examples (multidimensional piecewise expanding maps). Finally, for irrational rotations (which do not satisfy the mixing assumption of Theorem 16), we will show a relation between the shortest distance and the irrationality exponent.

Example 4.2. *Theorem 16 can be applied to the following systems:*

1. *For $m\in\{2,3,\ldots\}$, let $T:[0,1]\to[0,1]$ be such that $x\mapsto mx \mod 1$ and $\mu=Leb$.*

2. *Let $T:(0,1]\to(0,1]$ be such that $T(x)=2^n(x-2^{-n})$ for $x\in(2^{-n},2^{-n+1}]$ and $\mu=Leb$.*

3. *(β-transformations) For $\beta>1$, let $T:[0,1]\to[0,1]$ be such that $x\mapsto\beta x \mod 1$ and μ be the Parry measure (see [64]), which is an absolutely continuous probability measure with density ρ satisfying $1-\frac{1}{\beta}\leq\rho(x)\leq(1-\frac{1}{\beta})^{-1}$ for all $x\in[0,1]$.*

4. *(Gauss map) Let $T:(0,1]\to(0,1]$ be such that $T(x)=\left\{\frac{1}{x}\right\}$ and $d\mu=\frac{1}{\log 2}\frac{dx}{1+x}$.*

In these examples it is easy to see that $D_k(\mu)=1$. Moreover, (H1) and (H2) are satisfied with the Banach space $\mathcal{C}=BV$, the space of functions of bounded variation (see e.g. [38] Section 4.1 and [54, 68, 69]).

Example 4.3 (Multidimensional piecewise expanding systems)**.** *In this example, we apply Theorem 16 to a family of maps defined by Saussol [72]: multidimensional* *piecewise uniformly expanding maps. It was observed in [6] that these maps generalize Markov maps which also contain one-dimensional piecewise uniformly expanding maps.*

Let $N \geq 1$ be an integer. We will work in the Euclidean space $\mathbb{R}^N$. We denote by $B_\epsilon(x)$ the ball with center x and radius ϵ. For a set $E \subset \mathbb{R}^N$, we write

$$B_\epsilon(E) := \{y \in \mathbb{R}^N : d(y, E) \leq \epsilon\}.$$

Let X be a compact subset of $\mathbb{R}^N$ with $\overline{X^\circ} = X$ and $T : X \to X$. The system (X, T) is a multidimensional piecewise expanding system if there exists a family of at most countably many disjoint open sets $U_i \subset X$ and V_i such that $\overline{U_i} \subset V_i$ and maps $T_i : V_i \to \mathbb{R}^N$ satisfying for some $0 < \alpha \leq 1$, for some small enough $\epsilon_0 > 0$, and for all i:

1. *$T|_{U_i} = T_i|_{U_i}$ and $B_{\epsilon_0}(TU_i) \subset T_i(V_i)$;*
2. *$T_i \in C^1(V_i), T_i$ is injective and $T_i^{-1} \in C^1(T_iV_i)$. Moreover, there exists a constant c, such that for all $\epsilon \leq \epsilon_0, z \in T_iV_i$ and $x, y \in B_\epsilon(z) \cap T_iV_i$ we have*
$$|\det D_xT_i^{-1} - \det D_yT_i^{-1}| \leq c\epsilon^\alpha |\det D_zT_i^{-1}|;$$
3. *$Leb(X \setminus \bigcup_i U_i) = 0$;*
4. *there exists $s = s(T) < 1$ such that for all $u, v \in TV_i$ with $d(u, v) \leq \epsilon_0$ we have $d(T_i^{-1}u, T_i^{-1}v) \leq sd(u, v)$;*
5. *let $G(\epsilon, \epsilon_0) := \sup_x G(x, \epsilon, \epsilon_0)$ where*
$$G(x, \epsilon, \epsilon_0) = \sum_i \frac{Leb(T_i^{-1}B_\epsilon(\partial TU_i) \cap B_{(1-s)\epsilon_0}(x))}{m(B_{(1-s)\epsilon_0}(x))},$$
then the number $\eta = \eta(\delta) := s^\alpha + 2\sup_{\epsilon \leq \delta} \frac{G(\epsilon)}{\epsilon^\alpha}\delta^\alpha$ satisfies $\sup_{\delta \leq \epsilon_0} \eta(\delta) < 1$.

If (X, T) is a topologically mixing multidimensional piecewise expanding map and μ be its absolutely continuous invariant probability measure, then for μ^k-almost every $(x_1, \ldots, x_k) \in X^k$,

$$\lim_{n \to +\infty} \frac{\log m_n(x_1, \ldots, x_k)}{-\log n} = \frac{k}{(k-1)N}.$$

Example 4.4 (Irrational rotations)**.** *For $\theta \in \mathbb{R} \setminus \mathbb{Q}$, let T_θ be the irrational rotation on the unit circle $\mathbb{T} = \mathbb{R}/\mathbb{Z}$ defined by*

$$T_\theta x = x + \theta.$$

Then for any $n \in \mathbb{Z}$, we have $T_\theta^n x = x + n\theta$ and the shortest distance becomes

$$m_n(x, y) = \min_{-n \leq j \leq n} \|(x - y) + j\theta\|.$$

The limit behavior of $m_n(x,y)$ is thus linked to the inhomogeneous Diophantine approximation.

Let

$$\eta = \eta(\theta) := \sup\{\beta \geq 1 : \liminf_{j\to\infty} j^\beta \|j\theta\| = 0\}$$

be the irrationality exponent of θ.

In [BLR19], we showed that the result of Theorem 16 does not hold for T_θ. More precisely, we proved that for Lebesgue almost all $(x,y) \in \mathbb{T}^2$, we have

$$\liminf_{n\to\infty} \frac{\log m_n(x,y)}{-\log n} = \frac{1}{\eta} \quad \textit{and} \quad \limsup_{n\to\infty} \frac{\log m_n(x,y)}{-\log n} = 1.$$

4.2 Shortest distance between observed orbits

Following the ideas of Sections 2.1 and 3.1, in this section, we extend our analysis to the study of observation of orbits.

Let (Y,d) be a metric space and $f : X \to Y$ be a measurable function.

We will study the behaviour of the shortest distance between k observed orbits:

$$m_n^f(x_1,\ldots,x_k) = \min_{i_1,\ldots,i_k=0,\ldots,n-1} \left(d(f(T^{i_1}x_1),\ldots,f(T^{i_k}x_k)) \right).$$

The case $k = 2$ was treated in [CLR20] while the case $k \geq 2$ was treated in [BR21].

Theorem 18 ([CLR20, BR21])**.** *Let $(X,\mathcal{A},\mu,T)$ be a measure preserving system such that $\underline{D}_k(f_*\mu) > 0$. Then for μ^k-almost every $(x_1,\ldots,x_k) \in X^k$,*

$$\overline{\lim_{n\to+\infty}} \frac{\log m_n^f(x_1,\ldots,x_k)}{-\log n} \leq \frac{k}{(k-1)\underline{D}_k(f_*\mu)}.$$

We will assume that f is Lipschitz and as in Section 4.1, we prove that the equality holds under some rapidly mixing conditions:

(H1') For all $\psi,\ \phi \in \mathcal{H}^\alpha(Y,\mathbb{R})$ and for all $n \in \mathbb{N}^*$, we have

$$\left| \int_X \psi(f(x)).\phi(f(T^n x))\, d\mu(x) - \int_X \psi(f(x))d\mu(x) \int_X \phi(f(x))d\mu(x) \right| \leq \|\psi\circ f\|_{\mathcal{H}^\alpha} \|\phi\circ f\|_{\mathcal{H}^\alpha} \theta_n,$$

with $\theta_n = a^n$ $(0 \leq a < 1)$.

For simplicity, we only treat the case when the mixing property is satisfied for Hölder observables. However, we observe that on can adapt the assumptions (H1) and (H2) of the previous section to this setting to work with other Banach spaces.

Now we can state our version of Theorem 17 for observed orbits.

Theorem 19 ([CLR20, BR21])**.** *Let $(X,\mathcal{A},\mu,T)$ be a measure preserving system and f a Lipschitz observation, such that $\underline{D}_k(f_*\mu) > 0$ and such that (Y,d) is* *tight, satisfying (H1') and such that $f_*\mu$ satisfies (HA). Then for μ^k-almost every $(x_1,\dots,x_k)\in X^k$,*

$$\varliminf_{n\to+\infty}\frac{\log m_n^f(x_1,\dots,x_k)}{-\log n}\geq\frac{k}{(k-1)\overline{D}_k(f_*\mu)}.$$

Moreover, if $D_k(f_*\mu)$ exists, then for μ^k-almost every $(x_1,\dots,x_k)\in X^k$,

$$\lim_{n\to+\infty}\frac{\log m_n^f(x_1,\dots,x_k)}{-\log n}=\frac{k}{(k-1)D_k(f_*\mu)}.$$

4.3 Shortest distance between multiple random orbits

In this section, we follow the ideas and setting of Sections 2.2 and 3.2 and use the results on the shortest distance between observe orbits to study the shortest distance between multiple orbits of a random dynamical system.

For $(\omega_1,x_1),\dots,(\omega_k,x_k)$, we define the shortest distance between k random orbits by

$$m_n^{\omega_1,\dots,\omega_k}(x_1,\dots,x_k)=\min_{i_1,\dots,i_k=0,\dots,n-1}\left(d\left(T_{\omega_1}^{i_1}(x_1),\dots,T_{\omega_k}^{i_k}(x_k)\right)\right).$$

Remark 4.5. *We observe that the technic developed here only allows us to obtain annealed results. Another object worth studying would be the quenched shortest distance*

$$m_n^{\omega}(x_1,\dots,x_k)=\min_{i_1,\dots,i_k=0,\dots,n-1}\left(d\left(T_{\omega}^{i_1}(x_1),\dots,T_{\omega}^{i_k}(x_k)\right)\right).$$

In this direction, the only known results are when the system is a random subshift of finite type and will be explain in Section 5.3.

As in the deterministic case, we will assume an exponential decay of correlations for the random dynamical system:

(H1R) (Annealed decay of correlations) For every $n\in\mathbb{N}^*$, and every ψ, $\phi\in\mathcal{H}^\alpha(X,\mathbb{R})$,

$$\left|\int_{\Omega\times X}\psi(T_\omega^n(x))\phi(x)\;d\mu(\omega,x)-\int_{\Omega\times X}\psi\;d\mu\int_{\Omega\times X}\phi\;d\mu\right|\leq\|\psi\|_{\mathcal{H}^\alpha}\|\phi\|_{\mathcal{H}^\alpha}\theta_n,$$

with $\theta_n=a^n$ $(0\leq a<1)$.

Theorem 20 ([CLR20, BR21])**.** *Let $\mathcal{T}$ be a random dynamical system on X over $(\Omega,B(\Omega),\mathbb{P},\vartheta)$ with an invariant measure μ such that $\underline{D}_k(\nu)>0$. Then for μ^k-almost every $(\omega_1,x_1,\dots,\omega_k,x_k)\in(\Omega\times X)^k$,*

$$\varlimsup_{n\to\infty}\frac{\log m_n^{\omega_1,\dots,\omega_k}(x_1,\dots,x_k)}{-\log n}\leq\frac{k}{(k-1)\underline{D}_k(\nu)}.$$

Moreover, if the random dynamical system satisfies assumptions $(H1R)$ *and* ν *satisfies (HA), then*

$$\underline{\lim}_{n\to\infty} \frac{\log m_n^{\omega_1,\ldots,\omega_k}(x_1,\ldots,x_k)}{-\log n} \geq \frac{k}{(k-1)\overline{D}_k(\nu)},$$

and if $D_k(\nu)$ *exists, then*

$$\lim_{n\to\infty} \frac{\log m_n^{\omega_1,\ldots,\omega_k}(x_1,\ldots,x_k)}{-\log n} = \frac{k}{(k-1)D_k(\nu)} .$$

The proof of this theorem follows exactly the same ideas explained in the proof of Theorem 3.

We now apply the above result to the non-i.i.d. random dynamical system defined in Example 2.7. We observe that, in [CLR20], Theorem 20 was also applied to randomly perturbed dynamical systems and random hyperbolic toral automorphisms.

Example 4.6 (Non-i.i.d. random expanding maps)**.** *As in Example 2.7, let* T_1 *and* T_2 *be the two following maps defined on the one-dimensional torus* $X = \mathbb{T}^1$*:*

$$\begin{array}{rcl} T_1 : X & \longrightarrow & X \\ x & \longmapsto & 2x \end{array} \quad \text{and} \quad \begin{array}{rcl} T_2 : X & \longrightarrow & X \\ x & \longmapsto & 3x \end{array}$$

chosen following a Markov process with the stochastic matrix

$$A = \begin{pmatrix} 1/2 & 1/2 \\ 1/3 & 2/3 \end{pmatrix}.$$

Since in this example $\nu = Leb$*, we have* $D_k(\nu) = 1$*, thus Theorem 20 implies that for* Leb^{2k}*-almost every* $(\omega_1, x_1, \ldots, \omega_k, x_k) \in ([0,1]\times\mathbb{T}^1)^k$*,*

$$\lim_{n\to\infty} \frac{\log m_n^{\omega_1,\ldots,\omega_k}(x_1,\ldots,x_k)}{-\log n} = \frac{k}{k-1}.$$

Chapter 5

The longest common substring problem

Motivations to study sequence matching or sequence alignment can be found in various fields of research (e.g. computer science, biology, bioinformatics, geology and linguistics, etc.)

One particularly relevant object in DNA comparison is the longest common substring, i.e. the longest string of DNA which appears in two (or more) strands. For example, for the following two strands

$$ACAATGAGAGGATGACCTTG$$

$$TGACTGTAACTGACACAAGC$$

a longest common substring is ACAA (TGAC is also a longest common substring) and is of length 4 when the total length of the strands is 20.

In this chapter, we will concentrate on the behaviour of the length of the longest common substring when the length of the strings grows, more precisely, for two sequences x and y, the behaviour, when n goes to infinity, of

$$M_n(x, y) = \max\{m : x_{i+k} = y_{j+k} \text{ for } k = 1, \dots, m \text{ and for some } 0 \leq i, j \leq n - m\}.$$

For sequences drawn randomly from the same alphabet, this problem was studied by Arratia and Waterman in [8]. More precisely, if each term of the sequences is drawn independently within some alphabet $\mathcal{A}$ with respect to some probability $\mathcal{P}$, then they proved that for $\mathcal{P}^{\mathbb{N}} \otimes \mathcal{P}^{\mathbb{N}}$-almost every $(x, y) \in \mathcal{A}^{\mathbb{N}} \times \mathcal{A}^{\mathbb{N}}$

$$\lim_{n\to\infty} \frac{M_n(x, y)}{\log n} = \frac{2}{-\log p}$$

where $p = \sum_{a\in\mathcal{A}} \mathcal{P}(a)^2$.

They also proved the same result for independent irreducible and aperiodic Markov chains on a finite alphabet, and in this case p is the largest eigenvalue of the matrix $[(p_{ij})^2]$ (where $[p_{ij}]$ is the transition matrix).

In fact, one can observe that in both case, $-\log p$ corresponds to the Rényi entropy of μ (see Definition 5.1). Generalizations of the work [8] to sequences of different lengths, different distributions, more than two sequences, extreme value theory for sequence matching and distributional results can be found in e.g. [9, 7, 10, 11, 57, 30, 63, 62].

In [BLR19], we showed that a generalization of the longest common substring problem is to study the behaviour of the shortest distance between two orbits (we recall that $m_n(x, y) = \min_{i,j=0,\dots,n-1} (d(T^i x, T^j y))$). Indeed, when $X = \mathcal{A}^{\mathbb{N}}$ for some alphabet $\mathcal{A}$ and T is the shift on X, we can consider the distance between two sequences $x, y \in X$ defined by $d(x, y) = e^{-k}$ where $k = \inf\{i \geq 0, x_i \neq y_i\}$.

Then, assuming that m_n is not too small, that is $-\log m_n(x, y) \leq n$ (we saw in Theorem 15 that this condition is satisfied for almost all couples (x, y) if n is large enough), one can observe that almost surely

$$M_n(x, y) \leq -\log m_n(x, y) \leq M_{2n}(x, y).$$

Thus $M_n(x, y)$ and $-\log m_n(x, y)$ have the same asymptotic behaviour.

In this chapter, we will show how, in [BLR19, CLR20, R21, R21b], we generalized the results of Arratia and Waterman to α-mixing systems with exponential decay, run-length encoded sequences and random subshift of finite type.

5.1 Longest common substring between random sequences

We will consider in this section the symbolic dynamical systems $(\Omega, \mathbb{P}, \sigma)$, where $\Omega = \mathcal{A}^{\mathbb{N}}$ for some alphabet $\mathcal{A}$, σ is the (left) shift on Ω and $\mathbb{P}$ is a σ-invariant probability measure. For k sequences $x^1, \ldots, x^k \in \Omega$, we are interested in the behaviour of

$$M_n(x^1, ..., x^k)$$
$$= \max\{m : x^1_{i_1+j} = ... = x^k_{i_k+j} \text{ for } j = 0, ..., m-1 \text{ and for some } 0 \leq i_1, ..., i_k \leq n-m\}.$$

We will show that the behaviour of M_n is linked with the generalized Rényi entropy of the system.

For $y \in \Omega$ we denote by $C_n(y) = \{z \in \Omega : z_i = y_i \text{ for all } 0 \leq i \leq n-1\}$ the *n-cylinder* containing y. Set $\mathcal{F}_0^n$ as the sigma-algebra over Ω generated by all n-cylinders.

Definition 5.1. *For $k > 1$, we recall the definition of the lower and upper generalized Rényi entropy:*

$$\underline{H}_k(\mathbb{P}) = \varliminf_{n\to+\infty} \frac{\log \sum \mathbb{P}(C_n)^k}{-(k-1)n} \qquad \textit{and} \qquad \overline{H}_k(\mathbb{P}) = \varlimsup_{n\to+\infty} \frac{\log \sum \mathbb{P}(C_n)^k}{-(k-1)n},$$

where the notation $\sum \mathbb{P}(C_n)^k$ means $\sum\limits_{y\in\mathcal{A}^n} \mathbb{P}(C_n(y))^k$. When the limit exists, we will denote it by $H_k(\mathbb{P})$.

Even if the existence of the Rényi entropy is not known in general, it was computed in some particular cases: Bernoulli shift, finite state Markov chains, Gibbs measure of a Hölder-continuous potential [50] and infinite state Markov chains [25]. The existence was also proved for ϕ-mixing measures [61], for weakly ψ-mixing processes [50] and for ψ_g-regular processes [2, 4].

Remark 5.2. *If for a symbolic dynamical system, we consider the distance between two sequences $x, y \in X$ defined by $d(x, y) = e^{-k}$ where $k = \inf\{i \geq 0, x_i \neq y_i\}$, then the Renyi entropy of order k can be seen as the symbolic equivalent of the generalized fractal dimension of order k defined in Section 1.1.*

We say that a system $(\Omega, \mathbb{P}, \sigma)$ is *α-mixing* if there exists a function $\alpha : \mathbb{N} \to \mathbb{R}$ satisfying $\alpha(g) \to 0$ when $g \to +\infty$ and such that for all $m, n \in \mathbb{N}$, $A \in \mathcal{F}_0^n$ and $B \in \mathcal{F}_0^m$:

$$\left|\mathbb{P}(A \cap \sigma^{-g-n}B) - \mathbb{P}(A)\mathbb{P}(B)\right| \leq \alpha(g).$$

It is said to be *α-mixing with an exponential decay* if the function $\alpha(g)$ decreases exponentially fast to 0.

We say that our system is *ψ-mixing* if there exists a function $\psi : \mathbb{N} \to \mathbb{R}$ satisfying $\psi(g) \to 0$ when $g \to +\infty$ and such that for all $m, n \in \mathbb{N}$, $A \in \mathcal{F}_0^n$ and $B \in \mathcal{F}_0^m$:

$$\left|\mathbb{P}(A \cap \sigma^{-g-n}B) - \mathbb{P}(A)\mathbb{P}(B)\right| \leq \psi(g)\mathbb{P}(A)\mathbb{P}(B).$$

Now we are ready to state our next result. The case $k = 2$ was treated in [BLR19] while the case $k \geq 2$ was treated in [BR21].

Theorem 21 ([BLR19, BR21]). *If $\underline{H}_k(\mathbb{P}) > 0$, then for $\mathbb{P}^k$-almost every $(x^1, \ldots, x^k) \in \Omega^k$,*

$$\overline{\lim_{n\to+\infty}} \frac{M_n(x^1,\ldots,x^k)}{\log n} \leq \frac{k}{(k-1)\underline{H}_k(\mathbb{P})}.$$

Moreover, if the system is α-mixing with an exponential decay or if it is ψ-mixing with $\psi(g) = g^{-a}$ for some $a > 0$ then, for $\mathbb{P}^k$-almost every $(x^1, \ldots, x^k) \in \Omega^k$,

$$\underline{\lim_{n\to+\infty}} \frac{M_n(x^1,\ldots,x^k)}{\log n} \geq \frac{k}{(k-1)\overline{H}_k(\mathbb{P})}.$$

Therefore, if the generalized Rényi entropy exists, then for $\mathbb{P}^k$-almost every $(x^1, \ldots, x^k) \in \Omega^k$,

$$\lim_{n\to+\infty} \frac{M_n(x^1,\ldots,x^k)}{\log n} = \frac{k}{(k-1)H_k(\mathbb{P})}.$$

This theorem can be applied, for example, to Markov chains and Gibbs states:

Example 5.3 (Markov chains). *If $(\Omega, \mathbb{P}, \sigma)$ is an irreducible and aperiodic Markov chain on a finite alphabet $\mathcal{A}$, then it is ψ-mixing with an exponential decay (see e.g. [22]). If we denote by P the associated stochastic matrix (with entries P_{ij}), then the matrix $P(k)$ whose entries are $P_{ij}(k) = P_{ij}^k$ has, by the Perron-Frobenius theorem, a single largest eigenvalue λ_k. Moreover, the generalized Rényi entropy exists and $H_k(\mathbb{P}) = -\log \lambda_k/(k-1)$ [50]. Thus, for $\mathbb{P}^k$-almost every $(x^1, \ldots, x^k) \in \Omega^k$,*

$$\lim_{n\to+\infty} \frac{M_n(x^1,\ldots,x^k)}{\log n} = \frac{k}{-\log \lambda_k}.$$

Example 5.4 (Gibbs states). *Let $\mathbb{P}$ be a Gibbs state of a Hölder-continuous potential ϕ. Then, the system is ψ-mixing with an exponential decay [21, 76]. Moreover, the generalized Rényi entropy exists and $H_k(\mathbb{P}) = (1/(k-1))\,(kP(\phi) - P(k\phi))$ where $P(\phi)$ is the pressure of the potential ϕ [50]. Thus, for $\mathbb{P}^k$-almost every $(x^1, \ldots, x^k) \in \Omega^k$,*

$$\lim_{n\to+\infty} \frac{M_n(x^1,\ldots,x^k)}{\log n} = \frac{k}{kP(\phi) - P(k\phi)}.$$

5.2 Longest common substring for run-length encoded sequences

One could wondered if the results of the previous section hold if the sequences are transformed following certain rules of modification. More precisely, if f is a measurable function (called an encoder) transforming a sequence x into another sequence $f(x)$, one could like to study the behaviour of $M_n(f(x), f(y))$ and try to obtain a relation with the Rényi entropy of the pushforward measure $f_*\mathbb{P}$. In this section, we will focus on this problem when the encoder is a compression algorithm: the run-length encoder.

For sequences with long runs of the same value, Run-Length Encoding (RLE) is a simple and efficient lossless data compression method. More precisely, for a run of the same value, the algorithm stored the value and the length of the run. For example, the following binary sequence

$$0000111000000001100111111111100000000$$

will be compressed as

$$(0,4)(1,3)(0,8)(1,2)(0,2)(1,9)(0,8).$$

Thus, this sequence of 37 characters will be represented after compression by a sequence of 14 characters.

RLE is typically used for image compression but has also application in image analysis [51], texture analysis of volumetric data [78] and has also been used for data compression of television signals [71] and fax transmission [55].

As in the previous section, we consider the symbolic dynamical systems $(\Omega, \mathbb{P}, \sigma)$. We assume that alphabet $\mathcal{A}$ is finite.

We now define properly the run-length encoder:

Definition 5.5. *Let $\mathcal{B} = \{(\alpha, k)\}_{\alpha\in\mathcal{A}, k\in\mathbb{N}}$. We define the run-length encoder $f : \mathcal{A}^{\mathbb{N}} \to \mathcal{B}^{\mathbb{N}}$ by*

$$f(\underbrace{\alpha_1 \dots \alpha_1}_{k_1}\underbrace{\alpha_2 \dots \alpha_2}_{k_2} \dots \underbrace{\alpha_n \dots \alpha_n}_{k_n} \dots) = (\alpha_1, k_1)(\alpha_2, k_2) \dots (\alpha_n, k_n) \dots$$

We observe that for all $i \in \mathbb{N}$, we consider that $\alpha_{i+1} \neq \alpha_i$.

We will focus our analysis on the length of the longest common substring of RLE sequences. Given two sequences $x, y \in \Omega$, we will study the behaviour of the n-length of the longest common substring of the RLE sequences $f(x)$ and $f(y)$

$$M_n^{RLE}(x, y) := M_n(f(x), f(y)).$$

To obtain information on the growth length of the longest common substring for RLE sequences, we will need an assumption on the decay of the measure of cylinders:

(A) There exist $c > 0$ and $h > 0$, such that for any $n \in \mathbb{N}$ and any $a \in \mathcal{A}$

$$\mathbb{P}(\underbrace{a \dots a}_{n}) \leq ce^{-hn}.$$

We observe that in particular this assumption is always satisfied if the process is ψ-mixing with summable decay [42, Lemma 1].

In [R21b], we proved a lower and an upper bound for the growth rate of the length of the longest common substring for RLE sequences:

Theorem 22 ([R21b]). *If $\underline{H}_2(f_*\mathbb{P}) > 0$ and if hypothesis (A) is satisfied, then for $\mathbb{P} \otimes \mathbb{P}$-almost every x, y,*

$$\overline{\lim_{n\to\infty}} \frac{M_n^{RLE}(x, y)}{\log n} \leq \frac{2}{\underline{H}_2(f_*\mathbb{P})}.$$

Moreover, if the system is α-mixing with an exponential decay (or ψ-mixing with $\psi(g) = g^{-a}$ for some $a > 0$), then, for $\mathbb{P} \otimes \mathbb{P}$-almost every x, y,

$$\underline{\lim}_{n\to\infty} \frac{M_n^{RLE}(x,y)}{\log n} \geq \frac{2}{\overline{H}_2(f_*\mathbb{P})}.$$

Thus, if the Rényi entropy exists, we get for $\mathbb{P} \otimes \mathbb{P}$-almost every x, y,

$$\lim_{n\to\infty} \frac{M_n^{RLE}(x,y)}{\log n} = \frac{2}{H_2(f_*\mathbb{P})}.$$

We will now give two examples satisfying our assumptions and where the Rényi entropy of the pushforward measure can be explicitly computed (for more details, we refer the reader to [R21b]).

Example 5.6 (Bernoulli process)**.** *Let us consider the alphabet $\mathcal{A} = \{a, b\}$ and the Bernoulli measure $\mathbb{P}$ such that $\mathbb{P}([a]) = p$ and $\mathbb{P}([b]) = 1 - p$ with $0 < p < 1$.*

For $\mathbb{P} \otimes \mathbb{P}$-almost every realizations x, y, we have

$$\lim_{n\to\infty} \frac{M_n^{RLE}(x,y)}{\log n} = \frac{4}{\log\left(\frac{(1+p)(2-p)}{p(1-p)}\right)}.$$

Example 5.7 (Markov chain with more than 2 states)**.** *Let us consider $(\Omega, \mathbb{P}, \sigma)$ an irreducible and aperiodic Markov chain on the finite alphabet $\mathcal{A} = \{\alpha_i\}_{1\leq i\leq N}$ and with transition matrix $P = (p_{ij})_{1\leq i,j\leq N}$ with $0 < p_{ij} < 1$ for every $1 \leq i, j \leq N$.*

We define the following transition matrix $Q = (q_{(\alpha_i,k)(\alpha_j,\ell)})_{1\leq i,j\leq N, k,\ell\in\mathbb{N}}$ where for all $1 \leq i \leq N$ and $k, \ell \in \mathbb{N}$

$$q_{(\alpha_i,k)(\alpha_i,\ell)} = 0.$$

and for $i \neq j$

$$q_{(\alpha_i,k)(\alpha_j,\ell)} = \frac{p_{ij}p_{jj}^{\ell-1}(1-p_{jj})}{(1-p_{ii})}.$$

Let λ be the largest positive eigenvalue of the matrix $Q_2 = \left(\left(q_{(\alpha_i,k)(\alpha_j,\ell)}\right)^2\right)_{1\leq i,j\leq N, k,\ell\in\mathbb{N}}$.

Then, we have for $\mathbb{P} \otimes \mathbb{P}$-almost every x, y

$$\lim_{n\to\infty} \frac{M_n^{RLE}(x,y)}{\log n} = \frac{2}{-\log\lambda}.$$

5.3 Longest common substring for random subshift of finite type

In this section, we will study the longest common substring for random subshift of finite type (following the setting of Section 1.3). Since, as explain in the introduction of this chapter, studying the longest common substring for a symbolic system is equivalent to studying the shortest distance between orbits, this section can be seen as a symbolic version of Section 4.3. Moreover, we emphasize that for random subshift we managed to obtain quenched results while only the annealed case was treated in Section 4.3.

To obtain our results, we will need information on the decay of the measure of cylinders, thus we define

$$h_2 = \underline{\lim}_{k\to+\infty} \frac{\log \int_\Omega \max_{C_k} \mu_\omega(C_k) d\mathbb{P}}{-k}$$

where the max is taken over all k-cylinders.

We will assume the following: there is a constant $a \in [0,1)$ and a function $\alpha(g)$ satisfying $\alpha(g) = \mathcal{O}(a^g)$ such that for all n, m, $A \in \mathcal{F}_0^n(X)$ and $B \in \mathcal{F}_0^m(X)$:

(I) the marginal measure ν satisfies

$$\left|\nu(A \cap \sigma^{-g-n}A) - \nu(A)^2\right| \le \alpha(g);$$

(II) for $\mathbb{P}$-almost every $\omega \in \Omega$

$$\left|\mu_\omega(A \cap \sigma^{-g-n}B) - \mu_\omega(A)\mu_{\theta^{n+g}\omega}(B)\right| \le \alpha(g).$$

One can observe that assumption (I) is weaker than α-mixing since in the intersection we only deal with the same cylinder A. We recall that the measure ν is α-mixing if:

(I-a) (exponential α-mixing) the marginal measure ν satisfies

$$\left|\nu(A \cap \sigma^{-g-n}B) - \nu(A)\nu(B)\right| \le \alpha(g)$$

for all m, n, $A \in \mathcal{F}_0^n(X)$ and $B \in \mathcal{F}_0^m(X)$.

First of all, we consider the annealed case:

Theorem 23 ([CLR20]). *If $0 < \underline{H}_2(\nu)$, then*

$$\overline{\lim_{n\to\infty}} \frac{M_n(x,y)}{\log n} \le \frac{2}{\underline{H}_2(\nu)} \text{ for } \mu\otimes\mu\text{-almost every } ((\omega,x),(\tilde{\omega},y)) \in \mathcal{E}\times\mathcal{E}.$$

Moreover, if hypothesis (I-a) holds, then

$$\underline{\lim}_{n\to\infty} \frac{M_n(x,y)}{\log n} \ge \frac{2}{\overline{H}_2(\nu)} \text{ for } \mu\otimes\mu\text{-almost every } ((\omega,x),(\tilde{\omega},y)) \in \mathcal{E}\times\mathcal{E}.$$

First of all, we observe that the statement of this theorem is slightly different that the one of Theorem 4.4 in [CLR20] since they consider more general dynamical systems and not only random subshifts of finite type. Nevertheless, one can adapt easily their results and proof to obtain the theorem as stated here.

We observe that the technics used in [CLR20] only give annealed results, thus, in [R21], we used different tools to obtain quenched results.

We present now the first quenched result of this section which gives an upper bound for the growth rate of the longest common substring.

Theorem 24 ([R21]). *If $0 < \underline{H}_2(\nu) \leq 2h_2$ and if hypothesis (I) and (II) hold, then for $\mathbb{P}$-almost every $\omega \in \Omega$,*

$$\overline{\lim_{n\to\infty}} \frac{M_n(x,y)}{\log n} \leq \frac{2}{\underline{H}_2(\nu)} \text{ for } \mu_\omega \otimes \mu_\omega\text{-almost every } (x,y) \in \mathcal{E}_\omega \times \mathcal{E}_\omega.$$

One can notice that in the deterministic case (Theorem 21) and in the annealed case (Theorem 23), no mixing assumptions are needed to obtain the upper bound. The main problem and difference with the deterministic case is that the sample measures are not invariant which is the main reason to use mixing to obtain the upper bound (and the lower).

Moreover, one can observe that assuming $\underline{H}_2(\nu) \leq 2h_2$ is not a too restrictive assumption. Indeed, in the deterministic case this hypothesis is always satisfied (see e.g. [50] in the proof of Theorem 1 (IV)). In the random setting, this assumption prohibits for example to have some sample measures with an extreme behaviour (relatively with the others).

To obtain a lower bound, we will need stronger assumptions: we will need α-mixing for the measure ν and we will require some mixing properties for the base transformation $(\Omega, \theta, \mathbb{P})$.

First of all, we will treat the case when $(\Omega, \theta, \mathbb{P})$ is a ρ-mixing two-sided shift, i.e. $\Omega = \mathcal{A}^{\mathbb{Z}}$ for some alphabet $\mathcal{A}$, θ is the shift and:

(III) (exponential ρ-mixing) For all n and for all $\psi \in L^2(\mathcal{F}_{-\infty}^n(\Omega))$ and $\phi \in L^2(\mathcal{F}_0^\infty(\Omega))$

$$\left| \int_\Omega \psi.\phi \circ \theta^{n+g} \, d\mathbb{P} - \int_\Omega \psi d\mathbb{P} \int_\Omega \phi d\mathbb{P} \right| \leq \rho(g) \|\psi\|_2 \|\phi\|_2$$

with $\rho(n) = \mathcal{O}(a^n)$.

Moreover, we will need that the sample measure μ_ω of a cylinder of size n does not depend on all the terms of ω:

(IV) there exists a function ℓ with $\ell(n) = \mathcal{O}(n)$ such that for $\mathbb{P}$-almost every ω and every cylinder $C \in \mathcal{F}_0^n(X)$, the function $\omega \mapsto \mu_\omega(C)$ belongs to $L^2(\mathcal{F}_{-\ell(n)}^{\ell(n)}(\Omega))$.

One can observe that it is quite simple to check if assumption (IV) is satisfied, however this assumption is restrictive and only enables us to work with some special family of sample measures.

To deal with more general random subshifts (e.g. random Gibbs measures) we need a stronger mixing assumption on the base $(\Omega, \theta, \mathbb{P})$ (satisfied for example for Anosov diffeomorphisms [59]):

(III') (exponential 3-mixing) There exists a Banach space $\mathcal{B}$ such that for all ψ, ϕ, $\varphi \in \mathcal{B}$, for all $n \in \mathbb{N}^*$ and $m \in \mathbb{N}^*$, we have

$$\left| \int_\Omega \psi.\phi \circ \theta^n.\varphi \circ \theta^{n+m} \, d\mathbb{P} - \int_\Omega \psi d\mathbb{P} \int_\Omega \phi d\mathbb{P} \int_\Omega \varphi d\mathbb{P} \right| \leq \|\psi\|_{\mathcal{B}} \|\phi\|_{\mathcal{B}} \|\varphi\|_{\mathcal{B}} \rho(\min(n,m))$$

with $\rho(n) = \mathcal{O}(a^n)$ and $\|.\|_{\mathcal{B}}$ is the norm in the Banach space $\mathcal{B}$.

We are now able to replace assumption (IV) by a less restrictive assumption:

(IV') There exists $\xi \geq 0$ such that for every $n \in \mathbb{N}$ and every cylinder $C \in \mathcal{F}_0^n(X)$, the functions $\psi_1 : \omega \mapsto \mu_\omega(C)$ and $\psi_2 : \omega \mapsto \max_{C_n} \mu_\omega(C_n)$ (where the max is taken over all n-cylinders) belong to the Banach space $\mathcal{B}$ and

$$\|\psi_1\|_{\mathcal{B}} \leq \xi^n \qquad \text{and} \qquad \|\psi_2\|_{\mathcal{B}} \leq \xi^n.$$

We observe that some alternative assumptions where also given in [R21] to work with infinite alphabets for example.

Theorem 25 ([R21])**.** *If $0 < \underline{H}_2(\nu) \leq \overline{H}_2(\nu) < 2h_2$ and if*

- *hypothesis (I-a), (II), (III) and (IV) are satisfied,*

or

- *hypothesis (I-a), (II), (III') and (IV') are satisfied,*

then, for $\mathbb{P}$-almost every $\omega \in \Omega$,

$$\varliminf_{n\to\infty} \frac{M_n(x,y)}{\log n} \geq \frac{2}{\overline{H}_2(\nu)} \text{ for } \mu_\omega \otimes \mu_\omega\text{-almost every } (x,y) \in \mathcal{E}_\omega \times \mathcal{E}_\omega.$$

Moreover, if the Rényi entropy exists, we get for $\mathbb{P}$-almost every $\omega \in \Omega$,

$$\lim_{n\to\infty} \frac{M_n(x,y)}{\log n} = \frac{2}{H_2(\nu)} \text{ for } \mu_\omega \otimes \mu_\omega\text{-almost every } (x,y) \in \mathcal{E}_\omega \times \mathcal{E}_\omega.$$

We will now apply our results to random Bernoulli shifts. We observe that random Gibbs measures where also treated in [R21].

Example 5.8 (Random Bernoulli shifts)**.** *Let $s \geq 1$ and (Ω, θ) be a subshift of finite type on the symbolic space $\{0, 1, \ldots, s\}^{\mathbb{Z}}$ and let $\mathbb{P}$ be a Gibbs measure from a Hölder potential.*

Let $b \geq 1$ and make the shift $\{0, 1, \ldots, b\}^{\mathbb{N}}$ a random subshift by putting on it the random Bernoulli measures constructed as follows. Let $W = (w_{ij})$ be a $s \times b$ stochastic matrix with entries in $(0, 1)$. Set $p_j(\omega) = w_{\omega_0,j}$. The random Bernoulli measure μ_ω is defined by

$$\mu_\omega([x_0 \ldots x_n]) = p_{x_0}(\omega)p_{x_1}(\theta\omega)\ldots p_{x_n}(\theta^n\omega).$$

We proved in [R21] that if $0 < \underline{H}_2(\nu) \leq 2h_2$ one can apply Theorem 24 and if besides that $\overline{H}_2(\nu) < 2h_0$ then one can apply Theorem 25.

For example, when the base is i.i.d., we can compute the Rényi entropy:

$$H_2(\nu) = -\log\left(\sum_{x_0}\left(\int p_{x_0}(\omega)d\mathbb{P}\right)^2\right)$$

and show that

$$h_2 = -\log\left(\int \max_{x_0} p_{x_0}(\omega)d\mathbb{P}\right).$$

So, if $H_2(\nu) < 2h_2$ (which can be easily checked), we have for $\mathbb{P}$-almost every $\omega \in \Omega$,

$$\lim_{n\to\infty} \frac{M_n(x,y)}{\log n} = \frac{2}{-\log\left(\sum_{x_0}\left(\int p_{x_0}(\omega)d\mathbb{P}\right)^2\right)}$$

for $\mu_\omega \otimes \mu_\omega$-almost every $(x,y) \in \mathcal{E}_\omega \times \mathcal{E}_\omega$.

Works of the author

[AR16] R. Aimino and J. Rousseau, *Concentration inequalities for sequential dynamical systems of the unit interval*, Ergodic Theory and Dynamical Systems **36** (2016), no. 8, 2384–2407.

[BLR19] V. Barros, L. Liao and J. Rousseau, *On the shortest distance between orbits and the longest common substring problem,* Advances in Mathematics, **334** (2019), 311-339.

[BR21] V. Barros and J. Rousseau, *Shortest distance between multiple orbits and generalized fractal dimensions,* to appear in Annales Henri Poincaré, arXiv:1912.07516

[CLR20] A. Coutinho, R. Lambert and J. Rousseau, *Matching strings in encoded sequences*, Bernoulli **26** (2020), no. 3, 2021–2050.

[CRS18] A. Coutinho, J. Rousseau, B. Saussol, *Large deviation for return times*, Nonlinearity **31** (11) (2018) 5162–5179.

[GRS15] S. Galatolo, J. Rousseau and B. Saussol, *Skew products, quantitative recurrence, shrinking targets and decay of correlations*, Ergodic Theory and Dynamical Systems **35** (2015), no. 6, 1814–1845

[HRY20] N. Haydn, J. Rousseau, F. Yang, *Exponential Law for Random Maps on Compact Manifolds*, Nonlinearity **33** (2020), no. 12, 6760–6789

[MR11] P. Marie and J. Rousseau, *Recurrence for random dynamical systems,* Discrete and Continuous Dynamical Systems, **30** (2011), 1-16.

[R21] J. Rousseau, *Longest common substring for random subshifts of finite type,* to appear in Annales de l'Institut Henri Poincaré Probabilité et Statistiques, arXiv: 1905.08131.

[R21b] J. Rousseau, *Rényi entropy and pattern matching for run-length encoded sequences*, to appear in ALEA, Latin American Journal of Probability and Mathematical Statistics, arXiv: 2003.05500.

[R12] J. Rousseau, *Recurrence rates for observations of flows*, Ergodic Theory and Dynamical Systems, **32** (2012), 1727–1751.

[R14] J. Rousseau, *Hitting time statistics for observations of dynamical systems,* Nonlinearity **27** (2014), 2377-2392.

[RS10] J. Rousseau and B. Saussol, *Poincaré recurrence for observations*, Transactions of the American Mathematical Society, **362** (2010), 5845–5859.

[RSV14] J. Rousseau, B. Saussol, P. Varandas, *Exponential law for random subshifts of finite type*, Stochastic Processes and their Applications **124** (2014) 3260–3276.

[RT15] J. Rousseau and M. Todd, *Hitting times and periodicity in random dynamics*, Journal of Statistical Physics **161** (2015), no. 1, 131–150.

[RVZ12] J. Rousseau and P. Varandas and Y. Zhao, *Entropy formulas for dynamical systems with mistakes*, Discrete and Continuous Dynamical Systems **32** (2012), no.12, 4391–4407.

Bibliography

[1] M Abadi, *Hitting, returning and the short correlation function,* Bull. Braz. Math. Soc. **37** (2006) 1–17.

[2] M. Abadi and L. Cardeno, *Renyi entropies and large deviations for the first-match function,* IEEE Trans. Inf. Theory(61), 4 (2015), 1629–1639.

[3] M. Abadi and A. Galves, *Inequalities for the ocurrence times of rare events in mixing processes. The state of the art*, Markov Process. Related Fields, **7** (2001),97–112.

[4] M. Abadi and R. Lambert, *From the divergence between two measures to the shortest path between two observables,* Ergod. Theory Dyn., **39** (2019), no. 7, 1729–1744.

[5] M. Abadi and S. Vaienti, *Large deviations for short recurrence,* Discrete and Continuous Dynamical Systems-Series A,**21** (2008), no. 3, 729-747.

[6] J. F. Alves, J. M. Freitas, S. Luzzatto, and S. Vaienti, *From rates of mixing to recurrence times via large deviations*, Adv. Math., **228** (2011), no. 2, 1203–1236.

[7] R. Arratia, L. Gordon and M. Waterman,*An extreme value theory for sequence matching*, Ann. Statist., **14** (1986), no. 3, 971–993.

[8] R. Arratia and M. Waterman, *An Erdös–Rényi law with shifts*, Adv. Math., **55** (1985), 13–23.

[9] R. Arratia and M. Waterman, *Critical phenomena in sequence matching*, Ann. Probab., **13** (1985), no. 4, 1236–1249.

[10] R. Arratia and M. Waterman, *The Erdös-Rényi strong law for pattern matching with a given proportion of mismatches*, Ann. Probab., **17** (1989), no. 3, 1152–1169.

[11] R. Arratia and M. Waterman, *A phase transition for the score in matching random sequences allowing deletions*, Ann. Appl. Probab., **4** (1994), no. 1, 200–225.

[12] J. S. Athreya, *Logarithm laws and shrinking target properties*, Proceedings of the Indian Academy of Sciences-Mathematical Sciences, Volume **119**, Number 4, pages 541–559, 2009.

[13] H. Aytaç, J. M. Freitas and S. Vaienti, Laws of rare events for deterministic and random dynamical systems;Trans. Am. Math. Soc., **367** (2015), 8229–8278.

[14] L. Barreira and B. Saussol, *Hausdorff dimension of measures via Poincaré recurrence*, Comm. Math. Phys, **219** (2001), 443–463.

[15] V. Baladi, *"Positive Transfer Operators and Decay of Correlations,"* Advances Series in Nonlinear Dynamics, World Scientific Publishing Co. Inc., River Edge, NJ, **16**, (2000).

[16] S. Jain and R.K. Bansal, *On large deviation property of recurrence times.* International Symposium on Information Theory Proceedings (ISIT), (2013), 2880-2884.

[17] J.-M. Barbaroux, F. Germinet and S. Tcheremchantsev, *Generalized fractal dimensions: equivalences and basic properties*, J. Math. Pures Appl. **80**, 10 (2001), 977–1012.

[18] L. Barreira and B. Saussol, *Hausdorff dimension of measures via Poincaré recurrence*, Comm. Math. Phys., **219** (2001), no. 2, 443–463.

[19] M. D. Boshernitzan, *Quantitative recurrence results*, Invent. Math., **113** (1993), 617–631.

[20] M. Boshernitzan and J. Chaika, *Diophantine properties of IETs and general systems: quantitative proximality and connectivity*, Invent. Math. **192** (2013), no. 2, 375–412.

[21] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Math, Vol. **470**, Springer, Berlin, 1975.

[22] R. Bradley, *Basic Properties of Strong Mixing Conditions. A Survey and Some Open Questions*, Probability Surveys Vol. **2** (2005) 107–144.

[23] J-R. Chazottes and P. Collet, *Poisson approximation for the number of visits to balls in non-uniformly hyperbolic dynamical systems*, Ergodic Theory Dynam. Systems, **33** (2013), 49–80.

[24] J.R. Chazottes and R. Leplaideur, *Fluctuations of the Nth return time for Axiom A diffeomorphisms*, Discrete Contin. Dyn. Syst., **13** (2005), 399-411.

[25] G. Ciuperca, V. Girardin, L. Lhote, *Computation and Estimation of Generalized Entropy Rates for Denumerable Markov Chains*, IEEE Trans. Inf. Theory, **57** (2011), 4026 – 4034.

[26] Z. Coelho, *Asymptotic laws for symbolic dynamical systems*, in Topics in symbolic dynamics and applications (Temuco, 1997), vol. 279 of London Math. Soc. Lecture Note Ser., Cambridge Univ. Press, Cambridge, 2000, 123–165.

[27] P. Collet, *Some ergodic properties of maps of the interval*, in Dynamical systems (Temuco, 1991/1992), vol. 52 of Travaux en Cours, Hermann, Paris, 1996, 55-91.

[28] P. Collet, *Statistics of closest return for some non-uniformly hyperbolic systems*, Ergodic Theory Dynam. Systems, **21** (2001), 401–420.

[29] C. D. Cutler and D. A. Dawson, *Estimation of dimension for spatially distributed data and related limit theorems*, J. Multivariate Anal., **28** (1989), no. 1, 115–148.

[30] A. Dembo, S. Karlin and O. Zeitouni, *Critical phenomena for sequence matching with scoring*, Ann. Probab., **22** (1994) 1993–2021.

[31] K. Falconer, *Techniques in fractal geometry*, John Wiley & Sons Ltd., Chichester, 1997.

[32] A.H. Fan, K.S. Lau and H. Rao, *Relationships between different dimensions of a measure*, Monatsh. Math., **135** (2002), no. 3, 191–201.

[33] D. Faranda, H. Ghoudi, P. Guiraud and S. Vaienti, *Extreme value theory for synchronization of coupled map lattices*, Nonlinearity, **31** (2018), 3326–3358.

[34] D. Faranda and S. Vaienti, *Correlation dimension and phase space contraction via extreme value theory*, Chaos, **28** (2018), 041103.

[35] A. Ferguson, M. Pollicott, *Escape rates for Gibbs measures,* Ergodic Theory Dynam. Systems **32** (2012) 961–988.

[36] A. C. M. Freitas, J. M. Freitas and M. Todd, *Hitting time statistics and extreme value theory.* Probab. Theory Related Fields, **147** (2010), 675–710.

[37] A.C.M. Freitas, J.M. Freitas and M. Todd, *Extreme Value Laws in Dynamical Systems for Non-smooth Observations*, J. Stat. Phys., **142** (2011), no. 1, 108-126.

[38] A. C. M. Freitas, J. M. Freitas and M. Todd, *The extremal index, hitting time statistics and periodicity.* Adv. Math. **231** (2012), 2626–2665.

[39] A.C.M. Freitas, J. M. Freitas and S. Vaienti, *Extreme Value Laws for non stationary processes generated by sequential and random dynamical systems*, Ann. Inst. H. Poincaré Probab. Statist. **53(3)** (2017), 1341–1370.

[40] S. Galatolo, *Dimension and hitting time in rapidly mixing systems*, Math. Res. Lett., **14** (2007), no. 5, 797–805.

[41] S. Galatolo and M. J. Pacifico, *Lorenz like flows: exponential decay of correlations for the Poincaré map, logarithm law, quantitative recurrence*, Erg. Th. Dyn. Sys., **30** (2010), no.6, 1703–1737.

[42] A. Galves, B. Schmitt, *Inequalities for hitting times in mixing dynamical systems*, Random Comput. Dyn. **5** (1997) 337–348.

[43] P. Grassberger, *Generalized dimension of strange attractors*, Phys. Lett. A **97** (1983) 227–230.

[44] P. Grassberger and I. Procaccia, *Measuring the strangeness of strange attractors*, Physica D, **9** (1–2) (1983), 189–208.

[45] P. Grassberger and I. Procaccia, *Characterization of strange attractors*, Phys. Rev. Lett., **50** (1983), 346–349.

[46] M. Guysinsky and S. Yaskolko, *Coincidence of various dimensions associated with metrics and measures on metric spaces*, Discrete Contin. Dyn. Syst., **3** (1997) 591–603.

[47] N. Haydn, *Statistical properties of equilibrium states for rational maps*, Ergodic Theory Dynam. Systems, **20** (2000), 1371–1390.

[48] N T A Haydn: Entry and return times distribution; Dynamical Systems: An International Journal **28(3)** (2013), 333–353.

[49] N. Haydn, Y. Lacroix and S. Vaienti, *Hitting and return times in ergodic dynamical systems*, Ann. Probab., **33** (2005), 2043–2050.

[50] N. Haydn and S. Vaienti, *The Rényi entropy function and the large deviation of short return times*, Ergodic Theory Dynam. Systems, **30** (2010), no. 1, 159–179.

[51] S. C. Hinds, J. L. Fisher, and D. P. D'Amato, *A document skew detection method using run-length encoding and the hough transform*, In 10th International Conference on Pattern Recognition, ICDR 1990, volume 1, 464–468. IEEE, IEEE Comput. Soc. Press, 1990.

[52] M. Hirata, *Poisson law for Axiom A diffeomorphisms*, Ergodic Theory Dynam. Systems, **13** (1993), 533-556.

[53] M. Hirata, B. Saussol and S. Vaienti, *Statistics of return times: a general framework and new applications*, Communications in Mathematical Physics **206** (1999) 33–55.

[54] F. Hofbauer, G. Keller, *Ergodic properties of invariant measures for piecewise monotonic transformations*, Math. Z., **180** (1982), no. 1, 119–140.

[55] Roy Hunter, A. Harry Robinson, *International digital facsimile coding standards*, Proceedings of the IEEE 68 (7) (1980)854–867.

[56] M. Kač, *On the notion of recurrence in discrete stochastic processes*, Bull. A.M.S. 53 (1947), 1002–1010.

[57] S. Karlin, F. Ost, *Counts of long aligned word matches among random letter sequences*, Adv. in Appl. Probab. **19** (1987), no. 2, 293–351.

[58] Y. Kifer and P.-D. Liu, *Random dynamics*, Handbook of Dynamical Systems, Vol. **1B**, Elsevier B. V., Amsterdam, (2006), 379–499.

[59] M. Kotani and T. Sunada, *The pressure and higher correlations for an Anosov diffeomorphism*, Ergodic Theory Dynam. Systems 21 (2001), no. 3, 807–821.

[60] R. Leplaideur and B. Saussol, *Large deviations for return times in non- rectangle sets for Axiom A diffeomorphisms*, Discrete Contin. Dyn. Syst., **22** (2008), 327-344.

[61] T. Luczak and W. Szpankowski, *A suboptimal lossy data compression based on approximate pattern matching*, IEEE Trans. Inform. Theory 43 (1997), 1439–1451.

[62] M. Mansson, *On Compound Poisson Approximation for Sequence Matching,* Combin. Probab. Comput. **9** (2000), 529-548.

[63] C. Neuhauser, *A phase transition for the distribution of matching blocks*, Combin. Probab. Comput., **5** (1996), 139–159.

[64] W. Parry, *On the β-expansions of real numbers*, Acta Math. Acad. Sci. Hungar., **11** (1960) 401–416.

[65] Y. Pesin, *Dimension Theory in Dynamical Systems*, University of Chicago Press, Chicago, 1997.

[66] Y. Pesin and H. Weiss, *A multifractal analysis of equilibrium measures for conformal expanding maps and Moran-like geometric constructions*, J. Stat. Phys., **86** (1997), 233–275.

[67] F. Pène and B. Saussol, *Back to balls in billiards,*Comm. Math. Phys. 293 (2010), 837?866.

[68] W. Philipp, *Some metric theorems in number theory*, Pacific J. Math., **20** (1967) 109–127.

[69] W. Philipp, *Some metrical theorems in number theory II*, Duke Math. J., **37** (1970) 447–458.

[70] B. Pitskel, *Poisson limit law for Markov chains*, Ergodic Theory Dynam. Systems, **11** (1991), 501-513.

[71] A. H. Robinson and C. Cherry, *Results of a prototype television bandwidth compression scheme*, Proceedings of the IEEE, vol. 55 (1967), no. 3, 356–364.

[72] B. Saussol, *Absolutely continuous invariant measures for multidimensional expanding maps,* Israel J. Math., **116** (2000), 223–248.

[73] B. Saussol, *Recurrence rate in rapidly mixing dynamical systems*, Discrete Contin. Dyn. Syst., **15** (2006), 259–267.

[74] B. Saussol, *An introduction to quantitative Poincaré recurrence in dynamical systems,* Rev. Math. Phys., **21** (2009) 949–979.

[75] P. Varandas, *Correlation decay and recurrence asymptotics for some robust nonuniformly hyperbolic maps*, J. Stat. Phys. **133**, 813–839, (2008)

[76] P. Walters, *Ruelle's operator theorem and g-measures*, Trans. Am. Math. Soc., **214**, (1975), 375–387.

[77] M. J. Warrens, *n-way metrics*, J. Classification **27** (2010), no. 2, 173–190.

[78] D.-H. Xu, A. S. Kurani, J. D. Furst, and D. S. Raicu, *Run-length encoding for volumetric texture*, 4th IASTED International Conference on Visualization, Imaging, and Image Processing, VIIP 2004. Acta Press, 2004.

[79] L.-S. Young, *Dimension, entropy and Lyapunov exponents, Ergodic Theory Dynam. Systems* **2** (1982), 109-124.

[80] L-S Young, *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. (2) **147(3)** (1998), 585–650.