CLOSEST DISTANCE BETWEEN ITERATES
OF TYPICAL POINTS

BOYUAN ZHAO⊠

University of St Andrews, United Kingdom

(Communicated by Alejandro Maass)

ABSTRACT. The shortest distance between the first n iterates of a typical point can be quantified with a log rule for some dynamical systems admitting Gibbs measures. We show this in two settings. For topologically mixing Markov shifts with at most countably infinite alphabet admitting a Gibbs measure with respect to a locally Hölder potential, we prove the asymptotic length of the longest common substring for a typical point converges and the limit depends on the Rényi entropy. For interval maps with a Gibbs-Markov structure, we prove a similar rule relating the correlation dimension of Gibbs measures with the shortest distance between two iterates in the orbit generated by a typical point.

1. **Introduction.** Consider a topological Markov shift $(\Sigma_A, \sigma, \mathcal{I})$ on a (at most) countably infinite alphabet $\mathcal{I}$ with respect to a transition matrix A equipped with the natural symbolic metric d. The first n iterates of a $\underline{x} \in \Sigma_A$ under σ are the initial n symbols appearing in $\underline{x}$. We are interested in the asymptotic behaviour of the following quantity:

$$M_n(\underline{x}) = \max\{k : \exists\, 0 \le i < j \le n-1\colon x_i, \ldots, x_{i+k-1} = x_j, \ldots, x_{j+k-1}\}, \quad (1.1)$$

which counts the maximum length of self-repetition. Studying quantities of this type is often referred to as the longest common substring matching problem. One motivation comes from the matching of nucleotide sequences in DNA, and early results were established in the 80s by Arratia and Waterman's work [2]. They showed that the length of the longest common substring among two i.i.d stochastic sequences $X_1, X_2, \ldots$ and $Y_1, Y_2, \ldots$ taking letters in a finite alphabet with uniform distribution,

$$M_n(X, Y) := \sup\{k : X_{i+m} = Y_{j+m} \text{ for all } m = 1 \text{ to } k \text{ and } 1 \le i, j \le n-k\}$$

satisfies an Erdős-Rényi law

$$\mathbb{P}\left(\lim_{n \to \infty} \frac{M_n}{\log n} = \frac{2}{\log 1/p}\right) = 1,$$

where $p = \mathbb{P}(X_1 = Y_1)$ the collision probability, and $-\log \mathbb{P}(X_1 = Y_1)$ is often called the collision entropy or Rényi entropy.

The result can easily be translated to topological Markov shifts by replacing the stochastic sequences with two points in the shift spaces Σ_X and Σ_Y with distribution μ_X, μ_Y, and one can verify a similar convergence law holds for $M_n(\underline{x}, \underline{y})$ (see [8] and [7]). In a recent work [3], the authors proved as an improvement of the results in [7], for subshift systems with an invariant probability measure μ admitting good mixing conditions (more precisely, α-mixing with exponential decay or ψ-mixing with polynomial decay), the shortest distance between the n-orbit of a typical *pair* of points $\underline{x}, \underline{y}$,

$$M_n\left(\underline{x}, \underline{y}\right) = \sup\{k : x_{i_1+j} = y_{i_2+j}, \, j = 0, \ldots, k-1, \text{ for some } i_1, i_2 \leq n-k\}$$

converges to a Rényi entropy for $\mu \otimes^2 -$ almost every $(\underline{x}, \underline{y})$ in Σ_A^2. Similar almost sure convergences are proved for $k-$point-orbits in [4] for all $k \geq 2$.

The analogous problem for $M_n(\underline{x})$ is more difficult due to short return phenomenon. For subshifts of finite type, Collet et al in [6] applied first and second-moment analysis to the counting random variable $N(\underline{x}, n, r_n)$, which counts the number of matches of subwords of length r_n among the first n iterates in $\underline{x}$, there exists a constant H_2 which is the Rényi entropy of a Gibbs measure μ, such that

$$\lim_{n \to +\infty} \mu\left(\left|\frac{M_n(\underline{x})}{\log n} - \frac{2}{H_2}\right| > \varepsilon\right) = 0,$$

that is, $\frac{M_n(\underline{x})}{\log n}$ converges to $\frac{2}{H_2}$ in probability for typical $\underline{x}$. Then one may ask if this result can be improved to an almost sure convergence, or if the convergence remains valid when the alphabet is countably infinite. The answer is given by the following theorem.

Theorem 1.1. *For a one-sided subshift system $(\Sigma_A, \sigma, \mu, \mathcal{I})$ admitting a Gibbs measure μ,*

$$\lim_{n \to +\infty} \frac{M_n(\underline{x})}{\log n} = \frac{2}{H_2}$$

for μ-almost every $\underline{x} \in \Sigma_A$, where $M_n(\underline{x})$ is defined in (1.1).

The existence and uniqueness of the Gibbs measure with respect to locally Hölder potentials for finite $\mathcal{I}$ is well-known (see [5] for reference). For $\mathcal{I}$ countably infinite, they are characterised by theorems in [17] and [18]. Detailed discussion is in subsection 2.2 below.

The counterpart of the longest substring matching problem for dynamical systems (T, X, μ) acting on non-symbolic metric spaces (X, d) investigates the shortest distance between the two n-orbits generated by a typical pair of points. To be precise, we care about the following quantity

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

There is a dimension-like object for measures, called *correlation dimension* (denoted as $D_2(\mu)$, Definition 4.1), which plays a similar role to Rényi entropies for symbolic systems. In [3] the authors gave an asymptotic relation between $m_n(x, y)$ and the correlation dimension $D_2(\mu)$ for $\mu \otimes^2$-almost every (x, y), provided good decay of correlations. Later in [4], this rule is generalised for a typical collection of k points, $(x_1, x_2, \ldots, x_k)$, for $k = 2, 3, \ldots$. Again, we extend the investigation to the one-point case. Adopting techniques from [11], we will show that for one-dimensional

Gibbs-Markov interval maps, there is a similar asymptotic relation between $m_n(x)$ and $D_2(\mu)$, where

$$m_n(x) := \min_{0 \leq i < j \leq n-1} d\left(T^i x, T^j x\right). \tag{1.2}$$

Theorem 1.2. *Let X be a closed interval of $\mathbb{R}$, (X, T) a Gibbs-Markov system and μ a Gibbs probability measure admitting exponential decay of correlations for L^1 against $\mathcal{BV}$ observables. Then if its upper correlation dimension $\overline{D}_2(\mu)$ is bounded from 0,*

$$\liminf_{n \to \infty} \frac{\log m_n(x)}{-\log n} \geq \frac{2}{\overline{D}_2}$$

for $\mu-$almost every x in the repeller Λ. If μ is absolutely continuous with respect to the Lebesgue measure, then

$$\lim_{n \to \infty} \frac{\log m_n(x)}{-\log n} = \frac{2}{D_2}$$

for μ-almost every x, and in this case $\underline{D}_2(\mu) = \overline{D}_2(\mu) = 1$.

Remark 1.3. The proof for this setting is more challenging than that for the symbolic setting because the open balls defined by the Euclidean metric and the cylinders generated by the natural partitions disagree. The analysis of short return to balls is crucial for obtaining the upper bound of $\frac{\log m_n(x)}{-\log n}$, which is also generally harder than the recurrence analysis of cylinders. The proof for the lower bound relies on the $4-mixing$ property of Gibbs measures proved in Lemma 4.16.

Remark 1.4. Another difference between the two theorems stated above and theorems proved in [3], [4] is that in the single point case, obtaining asymptotic upper bounds of $M_n(x)$ and $m_n(x)$ requires good mixing properties. This should be expected due to the fundamental difference between one-point orbits and orbits generated by multiple independent points, the strengthening of assumptions is to ensure the iterates decorrelate fast enough to behave like an independent sequence after a relatively small number of iterations.

This theorem is applicable to the a range of systems, for example,

Example 1.5 (k-doubling maps). $f : [0, 1] \to [0, 1]$, $f(x) = kx \pmod 1$ for $k = 2, 3, \ldots$, and $\mu = Leb$.

Example 1.6 (Piecewise affine interval maps). Let $\{a_k\}_k$ be a monotonic sequence with $a_1 = 1$ and $\lim_k a_k = 0$. Then $f : [0, 1] \to [0, 1]$ with

$$f|_{[a_{k+1}, a_k)} = \frac{1}{a_k - a_{k+1}} (x - a_{k+1})$$

satisfies the assumptions above.

Example 1.7 (Gauss Map). Define the Gauss map $G : [0, 1] \to [0, 1]$ by

$$G(x) = \begin{cases} \dfrac{1}{x} \pmod 1 & x \in (0, 1] \\ 0 & x = 0 \end{cases}$$

It is a full-branched map. Let μ_G be the Gauss measure, it is the Gibbs measure for the potential $-\log DF$ with density $\frac{d\mu_G}{dLeb} = \frac{1}{(1+x)\log 2}$, then Theorem 1.2 holds for (G, μ_G).

Example 1.8 (**An induced map**)**.** Let F be the first return function to $[0, \frac{1}{2})$ of a Manneville–Pomeau map $f : [0,1] \to [0,1]$:

$$f(x) = \begin{cases} x(1 + 2^a x^a) & x \in \left[0, \dfrac{1}{2}\right) \\[2ex] 2x - 1 & x \in \left[\dfrac{1}{2}, 1\right] \end{cases}$$

for $a \in (0,1)$. There exists μ_F a Gibbs measure with respect to the potential $-\log DF$ (see [15] or [13, §13.2])and is absolutely continuous with respect to the Lebesgue measure.

2. **Preliminaries for Theorem 1.1.** We consider the one-sided symbolic dynamics. Let $\mathcal{I}$ be an at most countably infinite alphabet, A a $\mathbb{N} \times \mathbb{N}$ transition matrix of $0, 1$ entries and

$$\Sigma_A = \left\{ \underline{x} = (x_0, x_1, x_2, \dots) : x_i \in \mathcal{I}, A_{x_i, x_{i+1}} = 1 \right\}$$

be the symbolic space. It is equipped with a metric $d(\cdot, \cdot) = d_\theta(\cdot, \cdot)$, for some $\theta \in (0,1)$, and

$$d(\underline{x}, \underline{y}) = \theta^{\underline{x} \wedge \underline{y}}, \qquad \underline{x} \wedge \underline{y} := \min \left\{ j \geq 0 : x_j \neq y_j \right\}.$$

Without loss of generality, we may always assume $\theta = e^{-1}$. The shift space in question is denoted by (Σ_A, σ, d), where the left shift map σ

$$\sigma : \Sigma_A \to \Sigma_A, \quad (x_0, x_1, \dots) \mapsto (x_1, x_2, \dots).$$

When $\mathcal{I}$ is finite the space (Σ_A, d) is compact. A measure μ is σ-invariant if for all measurable $E \subseteq \Sigma_A$

$$\mu(E) = \mu(\sigma^{-1} E).$$

In dynamical systems literature, (Σ_A, σ) is often referred to as a *topological Markov chain*.

2.1. **General definitions and lemmas.** First, we need the most basic notion of open sets in the symbolic space.

Definition 2.1 (Cylinders)**.** A *k-cylinder* in Σ_A is a subset set of the form

$$[x_0, x_1, \dots, x_{k-1}] = \{ \underline{y} \in \Sigma_A : y_i = x_i, \forall i = 0, 1, \dots, k-1 \}.$$

The set of all k-cylinders is denoted as $\mathcal{C}_k$. For any $\underline{x} \in \Sigma_A$, let $C_k(\underline{x})$ be the k-cylinder containing $\underline{x}$, i.e. $C_k(\underline{x}) = [x_0, x_1, \dots, x_{k-1}]$. Cylinders are generating in the sense that any subset in Σ_A can be represented by a countable union of cylinders. Also, cylinders are also the open balls defined by the symbolic metric d.

Definition 2.2 (Rényi entropy)**.** For each $n \in \mathbb{N}$, $t > 0$, define the quantities

$$Z_n(t) = \sum_{C \in \mathcal{C}_n} \mu(C)^{1+t}. \tag{2.1}$$

The *Rényi entropy* (with respect to the natural partition given by the alphabet $\mathcal{I}$) of the system is given by

$$H_2(\mu) = \lim_{n \to +\infty} \frac{\log Z_n(1)}{-n}, \tag{2.2}$$

whenever this limit exists, and the generalised Rényi entropy function is

$$\mathcal{R}_\mu(t) = \liminf_{n \to +\infty} \frac{\log Z_n(t)}{-tn}.$$

In the information theory context, this entropy is also called *collision entropy* for Bernoulli systems, as it reflects the probability of two i.i.d. random variables coinciding i.e. $H_2 = -\log \sum_i p_i^2 = -\log \mathbb{P}(X = Y)$.

Rényi entropy does not always exist, especially when the alphabet is not finite. For the finite alphabet case, Haydn and Vaienti proved in [12, Theorem 1] that $\mathcal{R}_\mu(t)$ converges uniformly on compact subsets of $\mathbb{R}^+$ for all weakly ψ-mixing invariant measures, in particular, if μ is a Gibbs measure, $H_2(\mu) = \mathcal{R}_\mu(1) = 2P_{top}(\phi) - P_{top}(2\phi)$ where P_{top} is the topological pressure. For infinite alphabet Markov chains, Rényi entropy is obtained in [9].

Definition 2.3. The system is said to be ψ-*mixing*, if there is some monotone decreasing function $\psi(\cdot) : \mathbb{N} \to [0, \infty)$ such that for all n, k, all $E \in \mathcal{C}_n$, and all $F \in \mathcal{C}^*$, where $\mathcal{C}^* = \bigcup_{j=0}^\infty \mathcal{C}_j$,

$$\left| \frac{\mu(E \cap \sigma^{-n-k}F)}{\mu(E)\mu(F)} - 1 \right| \le \psi(k). \tag{2.3}$$

A measure is called *quasi-Bernoulli* if there is some constant $B > 1$ such that for any finite words $\mathbf{i}, \mathbf{j} \in \mathcal{C}^*$

$$\mu([\mathbf{ij}]) \le B\mu([\mathbf{i}])\mu([\mathbf{j}]).$$

Automatically, ψ-mixing entails the quasi-Bernoulli property.

Lemma 2.4. *If the probability measure μ is ψ-mixing with $\psi(\cdot)$ summable, there exists constant $\rho \in (0, 1)$ such that $\mu(C) \le \rho^n$ for all $C \in \mathcal{C}_n$ and all n.*

The proof is given for finite alphabet case in [10] which remains valid for countable alphabet case.

Notation. For two sequences $\{a_k\}_k, \{b_k\}_k$, the following notation is inherited from [6, Def 2.9].

Say $a_k \approx b_k$ if $\log a_k - \log b_k$ is bounded, or equivalently the ratio $\left| \frac{a_k}{b_k} \right|$ is uniformly bounded away from 0 and $+\infty$.

Say $a_k \preceq b_k$ if there is $\{c_k\}_k$ such that $a_k \le c_k$ for all k, and $b_k \approx c_k$. Both relations are transitive.

Notation. The following notation is also used.
(1) For each $\omega \in \mathbb{N}$ and any finite word $(x_0, x_1, \ldots, x_{k-1})$ of length $k \ge 2$, the notation $(x_0, x_1, \ldots, x_k)^\omega$ means the word is repeated ω-times whenever it is allowed.
(2) For any $\underline{x} \in \Sigma$, denote the k-word starting from position m by

$$x_m^k = (x_m, x_{m+1}, \ldots, x_{m+k-1}).$$

(3) The indicator function of a set E is denoted by $\mathbb{1}_E$.
(4) The expectation and variance of a random variable X are denoted respectively by $\mathbb{E}[X]$ and $Var[X]$.
(5) The cardinality of a set E is denoted by $\#E$.

2.2. **Thermodynamic formalism for Gibbs measures.** In [6] the authors considered the substring matching problem for a single point in a Markov subshift system, the relevant measure μ is Gibbsian with respect to a locally Hölder potential. In this section, we will provide conditions and lemmas which enable us to include certain types of countable Markov subshifts. The majority of references for this section can be found in [17] and [18].

Definition 2.5. A potential $\phi : \Sigma_A \to \mathbb{R}$ is called *locally Hölder* if there exists $M_\phi > 0$, $\zeta \in (0,1)$ such that for all $k \geq 1$, $var_k(\phi) \leq M_\phi \zeta^k$ where

$$var_k(\phi) := \sup \left\{ \left| \phi(\underline{x}) - \phi(\underline{y}) \right| : x_i = y_i, \, \forall i \leq k - 1 \right\}.$$

Definition 2.6. The system (Σ_A, σ) is *topologically mixing* if for all $a, b \in \mathcal{I}$, there is $n_0 \in \mathbb{N}$ such that for all $n > n_0$,

$$[a] \cap \sigma^{-n}[b] \neq \emptyset.$$

The system has *big image and preimage* property, if there is a finite subset $S \subseteq \mathcal{I}$ such that for any $a \in \mathcal{I}$, there are $b_1, b_2 \in \mathcal{S}$ such that $[b_1 \, a \, b_2] \neq \emptyset$. This condition is trivially satisfied for all subshifts of finite type.

Definition 2.7. Given a potential ϕ on Σ_A, a σ-invariant measure μ is said to be *Gibbs* if there are constants $c_\phi > 0$, $P \in \mathbb{R}$ such that for each $m \in \mathbb{N}$,

$$c_\phi^{-1} \leq \frac{\mu(C_m(\underline{x}))}{\exp\left(-mP + S_m\phi(\underline{x})\right)} \leq c_\phi, \tag{2.4}$$

where $S_m\phi(\underline{x}) = \sum_{i=0}^{m-1} \phi(\sigma^i \underline{x})$.

Definition 2.8. Let ϕ be a locally Hölder potential, the *partition functions* with respect to ϕ are defined by

$$P_n(\phi, a) := \sum_{\substack{\sigma^n \underline{x} = \underline{x} \\ x_0 = a}} e^{S_n(\phi)(\underline{x})}, \tag{2.5}$$

for each $a \in \mathcal{I}$, and the *Gurevich pressure* $P_G(\phi)$:

$$P_G(\phi) := \lim_{n \to \infty} \frac{1}{n} \log P_n(\phi, a). \tag{2.6}$$

For topologically mixing countable subshifts, $P_G(\phi)$ exists and is independent of the symbol $a \in \mathcal{I}$ ([17, Theorem 1]).

The existence of Gibbs measure for a countable topological Markov subshift is characterised by the following theorem:

Theorem 2.9. [18, Theorem 1],[17, Theorem 8] *Let ϕ be a locally Hölder potential, $(\Sigma_A, \sigma, \mathcal{I})$ topologically mixing , then ϕ has an invariant Gibbs measure μ if and only if the system satisfies the big image and preimage property and $P_G(\phi) < \infty$. In particular, let $\mathcal{L}_\phi$ be the Ruelle-Perron-Frobenius operator associated with ϕ,*

$$\mathcal{L}_\phi f(\underline{x}) = \sum_{\sigma \underline{y} = \underline{x}} e^{\phi(\underline{y})} f(\underline{y}),$$

then $\lambda = e^{P_G(\phi)}$ is the eigenvalue of $\mathcal{L}_\phi$ and the eigenfunction h is uniformly bounded from 0 and ∞, also $P_G(\phi) = P$ for P in (2.4).

Remark 2.10. It is shown by [17, Theorem 3,Theorem 8] that the Gibbs measure given in Theorem 2.9 is the unique equilibrium state which realises the equality below

$$P_G(\phi) := \sup \left\{ h_\nu + \int \phi \, d\nu : \nu \text{ is } \sigma \text{ invariant and } \int \phi \, d\nu > -\infty \right\}.$$

We will need the following lemmas for the proof of Theorem 1.1.

Lemma 2.11. *For subshifts of finite type or countable shifts satisfying the assumptions of Theorem 2.9, if μ is the (unique) Gibbs measure with respect to a locally Hölder potential ϕ, it has exponential rate ψ-mixing for cylinders.*

Proof. For $\mathcal{I}$ finite, one can verify exponential decay of ψ with [5, Proposition 1.14]. For countable $\mathcal{I}$, because the corresponding Gibbs measures for two comohologous Hölder potentials coincide, without loss of generality, one can assume $\mathcal{L}_\phi 1 = 1$, which implies $P_G(\phi) = 0$ and the conformal measure ν identifies with the Gibbs measure μ on cylinders.

Firstly, by locally Hölder property of ϕ, there is $M_1 > 0$ such that

$$\left| e^{S_n \phi(\underline{x}) - S_n \phi(\underline{y})} - 1 \right| \le M_1 d(\underline{x}, \underline{y}) \tag{2.7}$$

whenever there is $a \in \mathcal{I}$ such that $\underline{x}, \underline{y} \in [a]$

Also, as the invariant density is uniformly bounded from 0 and $+\infty$, there is $M_2 > 0$ such that for each n-cylinder $C \in \mathcal{C}_n$, for all $\underline{x} \in C$,

$$M_2^{-1} e^{S_n \phi(\underline{x})} \le \mu(C) \le M_2 e^{S_n \phi(\underline{x})}. \tag{2.8}$$

Now define the norm for real-valued function f acting on Σ_A,

$$\|f\|_L := \|f\|_\infty + D_\beta f,$$

where β is the $\sigma-$algebra generated by $\{\sigma[a] : a \in \mathcal{I}\}$ and

$$D_\beta := \sup_{b \in \beta} \sup_{\underline{x}, \underline{y} \in b} \frac{|f(\underline{x}) - f(\underline{y})|}{d(\underline{x}, \underline{y})}.$$

The operator $\mathcal{L}_\phi : Lip_{1,\beta} \to L$ where the spaces are defined by $Lip_{1,\beta} := \{f : \Sigma_A \to \mathbb{R} : \|f\|_1, D_\beta f \le \infty\}$ and $L := \{f : \Sigma_A \to \mathbb{R} : \|f\|_L < \infty\}$.

Consider $E = [e_0, e_1, \ldots, e_{n-1}] \in \mathcal{C}_n$ and $F \in \mathcal{C}^*$, as $\mathcal{L}_\phi^* \mu = \mu$,

$$\left| \mu(E \cap \sigma^{-(n+k)} F) - \mu(E)\mu(F) \right| = \left| \int \mathbb{1}_E \cdot \mathbb{1}_F \circ \sigma^{n+k} \, d\mu - \int \mathbb{1}_E \, d\mu \int \mathbb{1}_F \, d\mu \right|$$

$$\le \left| \int \mathbb{1}_F \left((\mathcal{L}_\phi^{n+k} \mathbb{1}_E) - \int \mathbb{1}_E \, d\mu \right) d\mu \right| \le \mu(F) \left\| \mathcal{L}_\phi^{n+k} \mathbb{1}_E - \int \mathbb{1}_F \, d\mu \right\|_\infty$$

$$\le \mu(F) \left\| \mathcal{L}_\phi^k (\mathcal{L}_\phi^n \mathbb{1}_E) - \int \mathcal{L}_\phi^n \mathbb{1}_E \, d\mu \right\|_L$$

It is a standard fact (see for example [1, Theorem 1.6] or [17, Theorem 5]) that there are $K_\phi > 0$ and $\kappa \in (0, 1)$ such that

$$\left\| \mathcal{L}_\phi^k (\mathcal{L}_\phi^n \mathbb{1}_E) - \int \mathcal{L}_\phi^n \mathbb{1}_E \, d\mu \right\|_L \le K_\phi \kappa^k \|\mathcal{L}_\phi^n \mathbb{1}_E\|_L.$$

Claim. $\|\mathcal{L}_\phi^n \mathbb{1}_E\|_L \le M_3 \mu(E)$ *for some* $M_3 > 0$.

Proof of claim. It is easy to see for each $E \in \mathcal{C}_n$ and $\underline{x} \in \Sigma_A$, there is only one $\underline{z} \in E = [e_0, e_1, \ldots, e_{n-1}]$ such that $\sigma^n \underline{z} = \underline{x}$, i.e.

$$\underline{z} = (e_0, \ldots, e_{n-1}, x_0, x_1, \ldots),$$

hence by (2.8), for all $\underline{x}$,

$$\mathcal{L}_\phi^n \mathbb{1}_E(\underline{x}) = \sum_{\sigma^n \underline{y} = \underline{x}} e^{S_n \phi(\underline{y})} \mathbb{1}_E(\underline{y}) = e^{S_n \phi(\underline{z})} \le M_2 \mu(E),$$

and for $\underline{x}, \underline{y} \in [b] \in \beta$, by (2.7)

$$\left| \mathcal{L}_\phi^n \mathbb{1}_E(\underline{x}) - \mathcal{L}_\phi^n \mathbb{1}_E(\underline{y}) \right| \leq \sum_{\substack{\underline{z},\underline{w} \in [e_0,\ldots,e_{n-1},b] \\ \sigma^n \underline{z} = \underline{x},\, \sigma^n \underline{w} = \underline{y}}} e^{S_n\phi(\underline{w})} \left| e^{S_n\phi(\underline{z}) - S_n\phi(\underline{w}) - 1} \right|$$

$$\leq M_2 \mu(E) M_1 d(\underline{x}, \underline{y}),$$

this gives $D_\beta \mathcal{L}_\phi^n \mathbb{1}_E \leq \mu(E) M_1 M_2$, and the claim is proved with $M_3 = M_2(1 + M_1)$. $\qquad \square$

The proof of lemma follows from the claim. $\qquad \square$

Now we will show that for countable topological Markov shifts, Rényi entropy of the Gibbs measure μ exists and is given by a formula involving the pressure function. The analogous statement for subshift of finite types is mentioned in [12] and is easy to verify.

Lemma 2.12. *Under the assumptions of Theorem 2.9, for μ the unique Gibbs measure with respect to ϕ, the Rényi entropy $H_2(\mu)$ exists and is given by*

$$H_2(\mu) = \lim_{n \to +\infty} \frac{\log Z_n(1)}{-n} = 2P_G(\phi) - P_G(2\phi).$$

Proof. Firstly, H_2 is clearly non-negative. By definition (2.1) and (2.4), for $B_1 := \sum_{k \geq 1} var_k(\phi)$,

$$Z_{n+k}(1) = \sum_{C \in \mathcal{C}_{n+k}} \mu(C)^2 \leq c_\phi^2 \sum_{C \in \mathcal{C}_{n+k}} \exp\left(S_{n+k}\phi(\underline{x}) - (n+k)P_G \right)$$

$$\leq c_\phi^4 e^{2B_1} Z_n(1) Z_k(1),$$

so $-\log Z_n(1)$ is almost subadditive, and the limit $\frac{\log Z_n(1)}{-n}$ exists, in particular, every subsequence converges to the same limit. Suppose $\underline{x}$ is a periodic point with period k, then for all n, since $\mu(C_{nk}(\underline{x}))^2 \leq Z_{nk}(1) = \sum_{C \in \mathcal{C}_{nk}} \mu(C)^2$,

$$\liminf_{n \to \infty} \frac{-2\log c_\phi + 2\left(S_{nk}\phi(\underline{x}) - nkP_G \right)}{nk} \leq \liminf_{n \to \infty} \frac{\log Z_n(1)}{n}.$$

Since $\underline{x}$ is periodic, $S_{nk}\phi(\underline{x}) = nS_k\phi(\underline{x})$ and both P_G and $S_k\phi(x)/k$ are finite, we get $\limsup_n \frac{\log Z_n(1)}{-n} < \infty$.

Combining the BIP property and locally Hölder property with [17, Lemma 4] one can show that

$$\limsup_{n \to \infty} \frac{1}{n} \log \sum_{C \in \mathcal{C}_n} \exp\left(\sup_{\underline{x} \in C} 2S_n\phi(\underline{x}) \right) \leq P_G(2\phi),$$

together with the Gibbs property we have,

$$\limsup_{n \to \infty} \frac{1}{n} \log \sum_{C \in \mathcal{C}} \mu(C)^2 \leq P_G(2\phi) - 2P_G(\phi).$$

Also for each $C \in \mathcal{C}_n$, there is at most one $\underline{x} \in C$ such that $\sigma^n \underline{x} = \underline{x}$, thus

$$\sum_{C \in \mathcal{C}_n} \mu(C)^2 \geq c_\phi^{-1} e^{-2nP_G(\phi)} \sum_{C \in \mathcal{C}_n} \exp\left(\sup_{\underline{x} \in C} 2S_n(\phi(\underline{x})) \right)$$

$$\geq c_\phi^{-1} e^{-2nP_G(\phi)} \sum_{\substack{\sigma^n \underline{x} = \underline{x} \in C \\ C \in \mathcal{C}_n,\, C \subseteq [a]}} \exp\left(2S_n\phi(\underline{x}) \right) = c_\phi^{-1} e^{-2nP_G(\phi)} P_n(2\phi, a),$$

which implies

$$\liminf_{n\to\infty} \frac{1}{n} \log \sum_{C\in\mathcal{C}_n} \mu(C)^2 \geq P_G(2\phi) - 2P_G(\phi).$$

Then putting the inequalities for $\limsup$ and $\liminf$ together,

$$H_2 = \lim_{n\to+\infty} \frac{\log \sum_{C\in\mathcal{C}_n} \mu(C)^2}{-n} = 2P_G(\phi) - P_G(2\phi). \qquad \square$$

Remark 2.13. It is also easy to see that $H_2(\mu) \leq 2h_\nu$ for all invariant probability measure ν: for all $\underline{x}$ and all $n \in \mathbb{N}$,

$$\frac{\log Z_n(1)}{-n} \leq \frac{2\log\nu(C_n(\underline{x}))}{-n}.$$

By the Shannon-McMillan-Breiman Theorem, the left hand side converges to $2h_\nu$ for almost every $\underline{x}$, therefore $\limsup_n \frac{\log Z_n(1)}{-n} \leq h_\nu$. So the Rényi entropy is finite whenever the measure-theoretic entropy of ν is finite.

For simplicity, denote $\alpha = \frac{H_2}{2}$. The following lemma is crucial for approximating the values of $Z_n(t)$.

Lemma 2.14. *For countable Markov shifts satisfying the assumptions of Theorem 2.9, for μ the invariant Gibbs measure, we have $\alpha > 0$ and*

$$Z_k(1) = \sum_{C\in\mathcal{C}_k} \mu(C)^2 \approx e^{-2k\alpha},$$

and for each $t > 2$,

$$Z_k(t-1) = \sum_{C\in\mathcal{C}_k} \mu(C)^t \preceq e^{-tk\alpha}.$$

Proof. Let $b_n := \max_{C\in\mathcal{C}_n} \mu(C)$, then by Lemma 2.4 $\sum_{C\in\mathcal{C}_n} \mu(C)^2 \leq b_n \sum \mu(C) \leq \rho^n$, hence

$$\liminf_{n\to\infty} \frac{\log Z_n(1)}{-n} \geq \liminf_{n\to\infty} \frac{-\log b_n}{-n} \geq -\log\rho > 0.$$

The approximation formulae are from [6, Lemma 2.13]. They were originally proved for finite alphabets and the proof remains valid if one combines with [12, Theorem 1 (IV)] which holds whenever the relevant measure admits exponential decay of cylinder measures. $\qquad \square$

3. **Proof of Theorem 1.1.** We will use dynamical Borel-Cantelli argument to show separately:

$$\limsup_{n\to+\infty} \frac{M_n(\underline{x})}{\log n} \leq \frac{2}{H_2} \tag{3.1.1}$$

and

$$\liminf_{n\to+\infty} \frac{M_n(\underline{x})}{\log n} \geq \frac{2}{H_2} \tag{3.1.2}$$

for μ-almost every $\underline{x} \in \Sigma_A$. Together they give

$$\lim_{n\to\infty} \frac{M_n(\underline{x})}{\log n} = \frac{2}{H_2}.$$

3.1. **Upper bound (3.1.1).**

Proof. Set
$$r_n = \frac{1}{\alpha - \varepsilon} \left(\log n + \log \log n \right).$$

As $M_n(\underline{x}) = r_n$ implies the return time of some iterate of $\underline{x}$ under σ to some r_n-cylinder is strictly less than n, we need to approximate the size of short return sets in the system in order to apply Borel-Cantelli Lemmas to obtain almost everywhere statements. Hence, as in [12] and [6], we intend to solve this by considering different cases of overlapping between r_n-substrings in $\underline{x}$.

Overlapping analysis. Let r_n be given; if r_n is not an integer, we simply take the closest integer since when $n \to +\infty$, $r_n \to +\infty$ it will not make any difference in terms of limiting behaviours. Let us define the following auxiliary sets.
$$S_k(r_n) = \left\{ \underline{x} \in \Sigma_A : \sigma^k \underline{x} \in C_{r_n}(\underline{x}) \right\}.$$

In other words, it is the set of points whose return time of $\underline{x}$ to the r_n-cylinder containing itself is k, and
$$\mu \left(\left\{ \underline{x} : \exists i, k \text{ such that } d(\sigma^i \underline{x}, \sigma^{i+k} \underline{x}) \le e^{-r_n} \right\} \right) \le \mu \left(\bigcup_{i=0}^{n-1} \bigcup_{k=1}^{n-i-1} \sigma^{-i} S_k(r_n) \right). \quad (3.2)$$

In order to obtain good estimates of $\mu(S_k(r_n))$, we consider three separate cases according to the range of k.

Let
$$\Sigma_0 = \Sigma_0(n) := \mu \left(\bigcup_{i=0}^{n-1} \bigcup_{k=1}^{\lfloor r_n/2 \rfloor} \sigma^{-i} S_k(r_n) \right).$$

Similarly, set
$$\Sigma_1 := \mu \left(\bigcup_{i=0}^{n-1} \bigcup_{k=\lfloor r_n/2 \rfloor + 1}^{r_n} \sigma^{-i} S_k(r_n) \right),$$

and
$$\Sigma_2 := \mu \left(\bigcup_{i=0}^{n-1} \bigcup_{k=r_n+1}^{n-i-1} \sigma^{-i} S_k(r_n) \right).$$

Moreover,
$$\mu(\{M_n > r_n\}) \le \mu(M_n \ge r_n) \le \Sigma_0 + \Sigma_1 + \Sigma_2. \quad (3.2')$$

Σ_0: *Return time* $1 \le k \le \lfloor r_n/2 \rfloor$. Let $\omega_k = \lfloor \frac{r_n}{k} \rfloor$ and $0 \le \gamma_k < k$ so that $r_n = k\omega_k + \gamma_k$. Then if $\underline{x} \in \sigma^{-i} S_k(r_n)$, $x_j = x_l$ if $j = l \pmod{k}$ for all j, $l \in [i, i+r_n+k-1]$, therefore $\sigma^i \underline{x}$ has the following form:
$$\sigma^i \underline{x} = \left(x_i^k, x_{i+k}^k, \ldots, x_{i+k\omega_k}^k, x_{i+k\omega_k+1}^{\gamma_k}, \ldots \right) = \left((x_i, \ldots, x_{i+k-1})^{\omega_k+1}, x_{i+k\omega_k+1}^{\gamma_k}, \ldots \right),$$

that is, a k-word $(x_i, \ldots, x_{i+k-1})$ will be repeated fully for $\omega_k + 1$ times, followed by a truncated γ_k-word with the same initial symbols. Also, for each $k \le \lfloor r_n/2 \rfloor$, there is at least one $\ell = \ell_k \in [\lceil r_n/4 \rceil, \lfloor r_n/2 \rfloor]$ such that ℓ is a multiple of k, meaning that $\underline{x} \in \sigma^{-i} S_k(r_n) \subseteq \sigma^{-i} S_\ell(r_n)$ where the ℓ word is fully repeated $\omega_{\ell_k} + 1 \le 5$ times, i.e. we may interpret $\sigma^i \underline{x}$ as
$$\sigma^i(\underline{x}) = \left((x_i, x_{i+1} \ldots, x_{i+\ell_k-1})^{\omega_{\ell_k}+1}, x_i^{\gamma_k}, \ldots \right).$$

Therefore for each $k \leq \lfloor r_n/2 \rfloor$, by the quasi-Bernoulli property,

$$\mu \left(\{ \underline{x} : \sigma^{i+k} \underline{x} \in C_{r_n}(\sigma^i \underline{x}) \} \right) = \mu(S_k(r_n)) \leq \mu(S_{\ell_k}(r_n))$$

$$\leq B^6 \sum_{C_{\ell_k} \in \mathcal{C}_\ell} \mu(C_\ell)^{\omega_{\ell_k}+1} \rho^{\gamma_k} \leq B^6 Z_{\ell_k}(\omega_{\ell_k}),$$

where ρ is a given by Lemma 2.4. As $r_n \leq \ell_k \omega_{\ell_k} \leq r_n + 1$, $e^{-\alpha \ell_k \omega_k} \leq e^{-\alpha r_n}$, by definition of r_n,

$$Z_{\ell_k}(\omega_k) \preceq e^{-\alpha(\omega_{\ell_k}+1)\ell_k} \leq e^{-\alpha r_n} \leq \exp\left(-\log n - \log\log n \right) \leq \frac{1}{n \log n},$$

For each $i \leq n-1$ and $1 \leq k \leq \lfloor r_n/2 \rfloor$, we can omit $\sigma^{-i} S_k(r_n)$ if $2k \leq n-i-1$ to avoid overcounting the redundant terms because $\sigma^{-i} S_k(r_n) \subseteq \sigma^{-i} S_{2k}(r_n)$, hence for Σ_0 we only need to consider points $\underline{x}$ such that $\sigma^i \underline{x}$ has short return time and $i \geq n - r_n$. As r_n is in the scale of $\log n$, we may choose n large such that $r_n \leq n^{1/2}$ so that by Lemma 2.14,

$$\Sigma_0 \leq B^6 \sum_{k=1}^{\lfloor r_n/2 \rfloor} k Z_{\ell_k}(\omega_{\ell_k}) \preceq B^6 r_n^2 e^{-\alpha r_n} \leq \frac{B^6 r_n^2}{n \log n} \preceq \frac{1}{\log n}. \tag{3.3}$$

Σ_1: Return time $\lfloor r_n/2 \rfloor + 1 \leq k \leq r_n$. In this case, $\underline{x} \in \sigma^{-i} S_k(r_n)$ implies $x_j = x_l$ if $j = l \pmod{k}$ for all $j, l \in [i, i+r_n+k-1]$, hence $\sigma^i \underline{x}$ has the form

$$\sigma^i \underline{x} = \left(x_i^{r_n-k}, x_{i+r_n-k}^{2k-r_n}, x_{i+k}^{r_n-k}, x_{i+r_n}^{2k-r_n}, x_{i+2k}^{r_n-k} \cdots \right)$$

$$= \left(x_i^{r_n-k}, x_{i+r_n-k}^{2k-r_n}, x_i^{r_n-k}, x_{i+r_n-k}^{2k-r_n}, x_i^{r_n-k} \cdots \right)$$

that is, the (r_n-k) word starting from x_i is repeated three times, separated by two identical $(2k-r_n)$ words. Hence by Lemma 2.14 and the quasi-Bernoulli property, the following upper bound for $\mu(\sigma^{-i} S_k(r_n))$ holds:

$$\mu(\sigma^{-i} S_k(r_n)) = \mu(S_k(r_n)) \leq B^6 \sum_{\substack{C \in \mathcal{C}_{r_n-k} \\ D \in \mathcal{C}_{2k-r_n}}} \mu(C)^3 \mu(D)^2 = B^6 Z_{r_n-k}(2) Z_{2k-r_n}(1).$$

As $\alpha > 0$,

$$\Sigma_1 \leq B^6 \sum_{k=\lfloor r_n/2 \rfloor + 1}^{r_n} (n-k) Z_{r_n-k}(2) Z_{2k-r_n}(1)$$

$$\preceq \sum_{k=\lfloor r_n/2 \rfloor + 1}^{r_n} (n-k) e^{-\alpha(3(r_n-k)+2(2k-r_n))}$$

$$= \sum_{k=\lfloor r_n/2 \rfloor + 1}^{r_n} (n-k) e^{-\alpha(r_n+k)} \leq e^{-\frac{3}{2} r_n \alpha} \sum_{k=\lfloor r_n/2 \rfloor + 1}^{r_n} (n-k) \preceq r_n n e^{-\frac{3}{2} \alpha r_n}$$

$$\leq \frac{n r_n}{(n \log n)^{3/2}} \leq \frac{n^{1/2} n}{(n \log n)^{3/2}} \leq \frac{1}{\log n}, \tag{3.4}$$

Σ_2: Return time $r_n + 1 \leq k \leq n - i - 1$. In this case, $k - r_N \geq 1$ and $\underline{x} \in \sigma^{-i} S_k(r_n)$ implies $x_i^{r_n}$, the r_n-word starting from position i of $\underline{x}$, is repeated from the $i+k$ entry without any overlapping with itself, i.e.

$$\sigma^i \underline{x} = \left(x_i^{r_n}, x_{i+r_n}^{k-r_n}, x_{i+k}^{r_n}, \ldots \right) = \left(x_i^{r_n}, x_{i+r_n}^{k-r_n}, x_i^{r_n}, \ldots \right)$$

then by the ψ-mixing condition,

$$\mu(\sigma^{-i}S_k(r_n)) \leq (1 + \psi(k - r_n)) \sum_{C \in \mathcal{C}_{r_n}} \mu(C)^2 = (1 + \psi(k - r_n))Z_{r_n}(1).$$

therefore

$$\Sigma_2 \leq \sum_{k=r_n+1}^{n-1} (n - k)\mu(S_k(r_n)) \leq \sum_{k=r_n+1}^{n-1} (n - k)(1 + \psi(k - r_n))Z_{r_n}(1).$$

Given that $(1 + \psi(k))$ is monotonically decreasing in k,

$$\Sigma_2 \approx e^{-2\alpha r_n} \sum_{k=r_n+1}^{n-1} (n - k)(1 + \psi(k - r_n)) \leq (1 + \psi(1))e^{-2\alpha r_n} \sum_{k=r_n+1}^{n-1} (n - k)$$

$$\leq (1 + \psi(1))e^{-2\alpha r_n}n^2 \leq \frac{1 + \psi(1)}{(\log n)^2} \leq \frac{1 + \psi(1)}{\log n} \tag{3.5}$$

Then, combining (3.2')-(3.5), there is some constant $K_1 > 0$ independent of n such that

$$\mu(\{M_n > r_n\}) \leq K_1 \frac{1}{\log n}.$$

Using the technique in the proof of [3, Theorem 5], picking a subsequence $n_k = e^{\lceil k^2 \rceil}$, then for all k large enough,

$$\mu(\{M_{n_k} > r_{n_k}\}) \leq K_1 \frac{1}{k^2},$$

then by Borel-Cantelli Lemma, for μ-almost every $\underline{x} \in \Sigma_A$,

$$M_{n_k}(\underline{x}) \leq r_{n_k}$$

which implies for all k large enough,

$$\frac{M_{n_k}(\underline{x})}{\log n_k} \leq \frac{1}{\alpha - \varepsilon}\left(1 + \frac{\log \log n_k}{\log n_k}\right).$$

Now taking the limsup of the inequality above, and since $M_n(\underline{x})$ is non-decreasing in n for all $\underline{x}$, for each n, there is a unique k such that $n_k \leq n < n_{k+1}$ with

$$\frac{\log n_k}{\log n_{k+1}} \cdot \frac{M_{n_k}(\underline{x})}{\log n_k} \leq \frac{M_n(\underline{x})}{\log n} \leq \frac{M_{n_{k+1}}(\underline{x})}{\log n_{k+1}} \cdot \frac{\log n_{k+1}}{\log n_k}, \tag{3.6}$$

the following inequality holds for all $\varepsilon > 0$ small,

$$\limsup_{n \to +\infty} \frac{M_n(\underline{x})}{\log n} = \limsup_{n \to +\infty} \frac{M_{n_k}(\underline{x})}{\log n_k} \leq \frac{1}{\alpha - \varepsilon}$$

since

$$\lim_{k \to +\infty} \frac{\log n_{k+1}}{\log n_k} = 1, \text{ and } \lim_{k \to +\infty} \frac{\log \log n_k}{\log n_k} = 0.$$

Then (3.1.1) is proved by letting $\varepsilon \to 0$ $\square$

3.2. **Lower bound (3.1.2).**

Proof. We apply a similar second-moment analysis as in the proof of [6, Theorem 4.1]. Let

$$r_n = \frac{1}{\alpha + \varepsilon} \left(\log n + \beta \log \log n \right)$$

for some uniform constant $\beta < 0$ to be determined later. Since $\sigma^{i+k}\underline{x} \in C_{r_n}(\sigma^i \underline{x})$ if and only if $\underline{x} \in \sigma^{-i} S_k(r_n)$, then we can define the random variable $\mathcal{S}_n$:

$$\mathcal{S}_n(\underline{x}) := \sum_{i=0}^{n-2r_n-1} \sum_{k=2r_n}^{n-i-1} \mathbb{1}_{C_{r_n}(\sigma^i \underline{x})}(\sigma^{i+k}\underline{x}) = \sum_{i=0}^{n-2r_n-1} \sum_{k=2r_n}^{n-i-1} \mathbb{1}_{\sigma^{-i} S_k(r_n)}(\underline{x}), \quad (3.7)$$

which counts the number of times that $\underline{x}$ is belongs to some $\sigma^{-i} S_k(r_n)$. As $M_n(\underline{x}) < r_n$ implies for all $0 \le i \le n-1$, $1 \le k \le n-i-1$, $\underline{x} \notin \sigma^{-i} S_k(r_n)$, and in particular not in the $\sigma^{-i} S_k(r_n)$ sets with $k \ge 2r_n$, therefore

$$\{ M_n(\underline{x}) < r_n \} \subseteq \{ S_n(\underline{x}) = 0 \}$$

and by Paley-Zygmund's inequality [16],

$$\mu(\{ M_n(\underline{x}) < r_n \}) \le \mu(\{ \mathcal{S}_n(\underline{x}) = 0 \}) \le \frac{Var[\mathcal{S}_n]}{\mathbb{E}[\mathcal{S}_n^2]} \le \frac{Var[\mathcal{S}_n]}{\mathbb{E}[\mathcal{S}_n]^2}. \quad (3.8)$$

By definition of $\sigma^{-i} S_k(r_n)$ with $k \ge 2r_n$, this set corresponds to the set of points in which an r_n-word repeats itself at least once with at least an r_n gap, therefore we have the following lower bound using the ψ-mixing property,

$$\mu\left(\{ C_{r_n}(\sigma^i \underline{x}) = C_{r_n}(\sigma^{i+k}\underline{x}) \} \right) = \mu\left(\sigma^{-i} S_k(r_n) \right) = \sum_{C \in \mathcal{C}_{r_n}} \mu(C \cap \sigma^{-k} C)$$

$$\ge (1 - \psi(k - r_n)) \sum_{C \in \mathcal{C}_{r_n}} \mu(C)^2 \ge (1 - \psi(r_n)) Z_{r_n}(1),$$

therefore

$$\mathbb{E}[\mathcal{S}_n] \ge \frac{1}{2}(1 - \psi(r_n))(n - 2r_n)^2 Z_{r_n}(1). \quad (3.9)$$

Next, we need to consider

$$\mathbb{E}[\mathcal{S}_n^2] = \sum_{i,j=0}^{n-2r_n-1} \sum_{k=2r_n}^{n-i-1} \sum_{l=2r_n}^{n-j-1} \mu(\sigma^{-i} S_k(r_n) \cap \sigma^{-j} S_l(r_n)). \quad (3.10)$$

Define the index set

$$F := \{ (i,j,k,l) : 0 \le i,\, j \le n - 2r_n - 1,\, 2r_n \le k \le n - i - 1,\, 2r_n \le l \le n - j - 1 \},$$

then

$$\mathbb{E}[\mathcal{S}_n^2] = \sum_{(i,j,k,l) \in F} \mu(\sigma^{-i} S_k(r_n) \cap \sigma^{-j} S_l(r_n)), \quad (3.10')$$

and the cardinality of F satisfies

$$\#F = \left(\sum_{i=2r_n}^{n-2r_n-1} n - i \right) \left(\sum_{j=2r_n}^{n-2r_n-1} n - j \right) \le \frac{1}{4}(n - 2r_n)^4.$$

Define the counting function by

$$\theta : F \to \mathbb{N}, \quad \theta(i,j,k,l) = \sum_{\substack{a \in \{i, i+k\} \\ b \in \{j, j+l\}}} \mathbb{1}_{(a - r_n, a + r_n)}(b),$$

i.e. it counts the occurrences that two indices in $\{i, j, i+k, j+l\}$ are r_n-close to each other; $\theta > 0$ translates to overlapping between some r_n words, e.g. $|i - j| < r_n$ implies the r_n word $x_i^{r_n}$ overlaps with the r_n word $x_j^{r_n}$, and both r_n-strings are repeated later.

By our definition of $\mathcal{S}_n$, for each quadruple (i, j, k, l), necessarily $k, l \geq 2r_n$ which implies

$$\theta(i, j, k, l) \leq 2, \qquad \forall (i, j, k, l) \in F,$$

which allows us to split (3.10') again into 3 components,

$$\mathbb{E}[\mathcal{S}_n^2] = \left(\sum_{F_0} + \sum_{F_1} + \sum_{F_2} \right) \mu \left(\sigma^{-i} S_k(r_n) \cap \sigma^{-j} S_l(r_n) \right),$$

where $F_t = \{(i, j, k, l) \in F : \theta(i, j, k, l) = t\}$.

Clearly,

$$\#F_0 \leq \#F \leq \frac{1}{4}(n - 2r_n)^4.$$

For each $(i, j, k, l) \in F_1$, if we fix any three indices, for example, if i, j, k are fixed, $j + l$ can be r_n-close to either i or $i + k$ as it is automatically $2r_n$- apart from j, hence there are at most $4r_n$ choices for the remaining index l. Hence

$$\#F_1 \leq 2r_n(n - 2r_n)^3,$$

and similarly if we fix any two of i, j, k, l in F_2, there are at most $2r_n^2$ choices for the remaining two indices, therefore

$$\#F_2 \leq 2r_n^2(n - 2r_n)^2.$$

Contributions of indices in F_0: We will consider the sum over indices in F_0 first. Since $(i, j, k, l) \in F_0$ implies no overlapping, $\underline{x} \in \sigma^{-i} S_k(r_n) \cap \sigma^{-j} S_l(r_n)$ implies $x_i^{r_n} = x_{i+k}^{r_n}$ and $x_j^{r_n} = x_{j+l}^{r_n}$ while the symbols in these two r_n-strings are independent, e.g. when $i + k < j$, $\underline{x}$ has the following form:

$$\sigma^i \underline{x} = \left(x_i^{r_n}, \ldots, x_{i+k}^{r_n}, x_{i+k+r_n}, x_{i+k+r_n+1}, \ldots, x_j^{r_n}, \ldots, x_{j+l}^{r_n} \cdots \right)$$
$$= \left(x_i^{r_n}, \ldots, x_i^{r_n}, \ldots, x_j^{r_n}, \ldots, x_j^{r_n}, \ldots \right).$$

Hence by ψ-mixing property

$$\mu(\sigma^{-i} S_k(r_n) \cap \sigma^{-j} S_l(r_n)) \leq (1 + \psi(\gamma_{ijkl}))^3 Z_{r_n}(1)^2, \qquad (3.11)$$

where

$$\gamma_{ijkl} = \min \{|a - b| - r_n : a, b \in \{i, j, i+k, j+l\}\}.$$

Let $F_0' \subseteq F_0$ be defined as

$$F_0' := \{(i, j, k, l) \in F_0 : \gamma_{ijkl} \geq r_n\},$$

and $F_0'' := F_0 \setminus F_0'$. Notice also that $\#(F_0'') \leq 2r_n(n - 2r_n)^3$.

Define the notation for any $G \subseteq F$,

$$\mathbb{E}[\mathcal{S}_n^2 | G] := \sum_{i,j,k,l \in G} \mu \left(\sigma^{-i} S_k(r_n) \cap \sigma^{-j} S_l(r_n) \right).$$

Then, using (3.11),

$$\mathbb{E}[\mathcal{S}_n^2 | F_0'] = \sum_{i,j,k,l \in F_0'} \mu \left(\sigma^{-i} S_k(r_n) \cap \sigma^{-j} S_l(r_n) \right) \leq (n - 2r_n)^4 (1 + \psi(g))^3 Z_{r_n}(1)^2,$$

By Lemma 2.11 $\psi(r_n) \leq r_n^{-1}$ for all n large enough, then

$$(1 + \psi(r_n))^3 - (1 - \psi(r_n))^2 \leq (1 + \psi(r_n))^3 - (1 - \psi(r_n))^3$$

$$= 2\,\psi(r_n)\left(2 + 2\psi(r_n)^2 + 1 - \psi(r_n)^2\right) \leq 2r_n^{-1}(3 + r_n^{-2}) \leq 8r_n^{-1} \leq \frac{1}{\log n} \quad (3.12)$$

using (3.9) with (3.12), as $\#(F_0') \leq \frac{1}{4}(n - 2r_n)^4$, for some constant $K_2 > 0$.

$$\frac{\mathbb{E}[\mathcal{S}_n^2 | F_0'] - \mathbb{E}[\mathcal{S}_n]^2}{\mathbb{E}[\mathcal{S}_n]^2} \leq \frac{(n - 2r_n)^4 Z_{r_n}(1)^2 \left((1 + \psi(r_n))^3 - (1 - \psi(r_n))^2\right)}{(n - 2r_n)^4 (1 - \psi(r_n))^2 Z_{r_n}(1)^2}$$

$$\leq \frac{(1 + \psi(r_n))^3 - (1 - \psi(r_n))^3}{(1 - \psi(r_n))^2} \leq K_2 \frac{1}{\log n}, \quad (3.13)$$

And for the sum over F_0'', the term $1 + \psi(\gamma_{ijkl})$ in (3.11) is uniformly bounded above by $1 + \psi(0)$, and $1 - \psi(r_n) \geq \frac{1}{2}$ for all n sufficiently large, therefore

$$\frac{\mathbb{E}[\mathcal{S}_n^2 | F_0'']}{\mathbb{E}[\mathcal{S}_n]^2} \approx \frac{r_n(n - 2r_n)^3 (1 + \psi(0))^3 Z_{r_n}(1)^2}{(1 - \psi(r_n))^2 (n - 2r_n)^4 Z_{r_n}(1)^2} \preceq \frac{r_n}{n - 2r_n}$$

hence for some $K_3 > 0$ and all n sufficiently large,

$$\frac{\mathbb{E}[\mathcal{S}_n^2 | F_0'']}{\mathbb{E}[\mathcal{S}_n]^2} \leq K_3 \frac{1}{\log n} \quad (3.14)$$

Contributions of indices in F_1: Next, for $(i, j, k, l) \in F_1$, without loss of generality, suppose only $|i - j| = r < r_n$, $i < j$ and $i + k < j + l$. The other cases are treated exactly the same since the order of the r_n-strings does not have any effects on estimations of the upper bounds for $\mu(\sigma^{-i} S_k(r_n) \cap \sigma^{-j} S_l(r_n))$.

An $\underline{x} \in \sigma^{-i} S_k(r_n) \cap \sigma^{-j} S_l(r_n)$ means $x_{i+r} = x_j$, $x_{i+r+1} = x_{j+1}$, ..., $x_{i+r_n} = x_{j+r}$, so $\sigma^i \underline{x}$ has the following form:

$$\sigma^i \underline{x} = \left(x_i^{r_n - r}, x_j^{r_n}, \ldots, x_i^{r_n}, \ldots, x_j^{r_n}, \ldots\right)$$

$$= \left(x_i^r, x_j^{r_n - r}, x_{j+r_n}^r, \ldots, x_i^r, x_j^{r_n - r}, \ldots, x_j^{r_n - r}, x_{j+r_n}^r, \ldots\right).$$

then using the quasi-Bernoulli property and Lemma 2.14, for $B > 1$ the relevant quasi-Bernoulli constant,

$$\mu(\sigma^{-i} S_k(r_n) \cap \sigma^{-j} S_l(r_n)) \leq B^{10} \sum_{\substack{A, B \in \mathcal{C}_r \\ C \in \mathcal{C}_{r_n - r}}} \mu(A)^2 \mu(B)^2 \mu(C)^3 = C^{10} Z_r(1)^2 Z_{r_n - r}(2)$$

$$\preceq B^{10} e^{-4\alpha r_n} e^{-3\alpha(r_n - r)} \leq B^{10} e^{-3\alpha r_n}.$$

Recall that $r_n = \frac{1}{\alpha + \varepsilon} (\log n + \beta \log \log n)$ and

$$e^{-\alpha r_n} = \left(n(\log n)^\beta\right)^{-\frac{\alpha}{\alpha + \varepsilon}} = n^{-\frac{\alpha}{\alpha + \varepsilon}} (\log n)^{-\frac{\alpha\beta}{\alpha + \varepsilon}}$$

hence for all n large enough such that

$$\frac{n^{\frac{\alpha}{\alpha + \varepsilon}}}{(n - 2r_n)} \leq 1, \quad (3.15)$$

by (3.9) and (3.15) above, as $\beta < 0$,

$$\frac{\mathbb{E}[\mathcal{S}_n^2 | F_1]}{\mathbb{E}[\mathcal{S}_n]^2} \preceq \frac{2r_n(n - 2r_n)^3 B^{10} e^{-3\alpha r_n}}{(n - 2r_n)^4 (1 - \psi(r_n))^2 Z_{r_n}(1)^2}$$

$$\approx \frac{r_n e^{-3\alpha r_n}}{(n - 2r_n) e^{-4\alpha r_n}} = \frac{1}{\alpha + \varepsilon} \frac{\log n + \beta \log \log n}{(n - 2r_n) e^{-\alpha r_n}}$$

$$\preceq \frac{\log n}{(n - 2r_n)e^{-\alpha r_n}} = \frac{(\log n)^{1+\beta\frac{\alpha}{\alpha+\varepsilon}} n^{\frac{\alpha}{\alpha+\varepsilon}}}{n - 2r_n}$$

$$\leq (\log n)^{\left(1+\beta(1-\frac{\varepsilon}{\alpha+\varepsilon})\right)}.$$

For all $0 < \varepsilon \leq \alpha$ such that $1 - \frac{\varepsilon}{\alpha+\varepsilon} \geq \frac{1}{2}$, one can choose

$$\beta = -4,$$

hence

$$1 + \beta\left(1 - \frac{\varepsilon}{\alpha+\varepsilon}\right) \leq -1$$

which is sufficient to conclude that for some constant K_4,

$$\frac{\mathbb{E}[\mathcal{S}_n^2|F_1]}{\mathbb{E}[\mathcal{S}_n]^2} \leq K_4 \frac{1}{\log n}. \tag{3.16}$$

Contributions of indices in F_2: Finally for indices in F_2, it is enough to know that for any $\underline{x} \in \sigma^{-i}S_k(r_n) \cap \sigma^{-j}S_l(r_n)$ there is some r_n-word repeated twice, so we can bound the measure of $\sigma^{-i}S_k(r_n) \cap \sigma^{-j}S_l(r_n)$ for each $(i,j,k,l) \in F_2$, by Lemma 2.14,

$$B^4 \sum_{C \in \mathcal{C}_{r_n}} \mu(C)^2 \approx B^4 e^{-2\alpha r_n}.$$

Then for $\beta = -4$ and all n verifying (3.15), as $(1 - \psi(r_n))$ is uniformly bounded from below, by (3.9),

$$\frac{\mathbb{E}[\mathcal{S}_n^2|F_2]}{\mathbb{E}[\mathcal{S}_n]^2} \approx \frac{2r_n^2(n - 2r_n)^2 B^4 e^{-2\alpha r_n}}{(n - 2r_n)^4(1 - \psi(r_n))^2 e^{-4\alpha r_n}}$$

$$\approx \frac{r_n^2}{(n - r_n)^2 e^{-2\alpha r_n}} \preceq \frac{n^{\frac{2\alpha}{\alpha+\varepsilon}}}{(n - 2r_n)^2}(\log n)^{2+2\beta\frac{\alpha}{\alpha+\varepsilon}}$$

$$\leq (\log n)^{2(1+\beta(1-\frac{\varepsilon}{\alpha+\varepsilon}))} \leq (\log n)^{-2},$$

and it follows that for some constant $K_5 > 0$,

$$\frac{\mathbb{E}[\mathcal{S}_n^2|F_2]}{\mathbb{E}[\mathcal{S}_n]^2} \leq K_5 \frac{1}{(\log n)^2}. \tag{3.17}$$

Then, combining (3.8) (3.10') (3.13)-(3.17), there is some constant $K_6 > 0$ such that

$$\mu(\{M_n < r_n\}) \leq K_6 \frac{1}{\log n},$$

hence we can repeat the trick of picking a subsequence $n_k = \lceil e^{k^2} \rceil$, and apply Borel-Cantelli Lemma to the sum $\sum_{k=1}^{\infty} \mu(\{M_{n_k} < r_{n_k}\}) < +\infty$, which means for all k large enough,

$$\frac{M_{n_k}(\underline{x})}{\log n_k} \geq \frac{1}{\alpha+\varepsilon}\left(1 - \frac{4\log\log n_k}{\log n_k}\right),$$

then taking the liminf on both sides and apply the arguments which validate (3.6),

$$\liminf_{n \to +\infty} \frac{M_n(\underline{x})}{\log n} = \liminf_{n \to +\infty} \frac{M_{n_k}(\underline{x})}{\log n_k} \geq \frac{1}{\alpha+\varepsilon},$$

then (3.1.2) is verified by letting $\varepsilon \to 0$. $\qquad\square$

4. **Gibbs-Markov interval maps.** This section is inspired by [3], [11] and [13], and we will prove an asymptotic behaviour for $m_n(x)$ for μ-almost every x, where the dynamics are interval maps with a Gibbs-Markov structure. Let us first define the measure theoretic dimension object.

Definition 4.1. Suppose $T : X \to X$ is a measurable map with respect on the probability space (X, μ). Define the *upper* and *lower correlation dimension* of μ by

$$\underline{D}_2(\mu) = \liminf_{r \to 0} \frac{\log \int \mu(B(x, r)) \, d\mu(x)}{\log r},$$

$$\overline{D}_2(\mu) = \limsup_{r \to 0} \frac{\log \int \mu(B(x, r)) \, d\mu(x)}{\log r}$$

respectively, and simply write $D_2(\mu)$ when the two limits coincide.

Clearly, $D_2(m) = 1$ for m the Lebesgue measure.

Definition 4.2. Say the metric space (X, d) satisfies the *bounded local complexity* condition if there exists $C_0 \in \mathbb{N}$ such that for each $r > 0$, there is $k(r) < \infty$, and $\{x_1^r, x_2^r, \ldots, x_{k(r)}^r\} \subseteq X$ such that

$$X \subseteq \bigcup_{p=1}^{k(r)} B(x_p^r, r)$$

and each $x \in X$ belongs to at most C_0 elements of $\{B(x_p^r, 2r)\}_{p=1}^{k(r)}$.

Any compact subset of $\mathbb{R}$ has bounded local complexity: compact implies totally bounded which gives $k(r) < \infty$ for all r and C_0 can be chosen to be 4 because one can choose an r-net such that $d(x_i^r, x_j^r) \geq r$ for $i \neq j \in \{1, \ldots, k(r)\}$.

Definition 4.3. (Piecewise expanding interval map) Let X be a closed interval in $\mathbb{R}$, $T : X \to X$ is a *piecewise expanding interval map* if there is a (at most) countable partition $\mathcal{P} = \{I_1, I_2, \ldots\}$ for T such that T is differentiable on each I_k and there is a uniform constant $\gamma > 0$ such that $|DT_{I_k}| > \gamma$. An n-cylinder $[x_0, x_1, \ldots, x_{n-1}]$ with respect to the partition $\mathcal{P}$ is given by

$$[x_0, x_1, \ldots, x_{n-1}] = \bigcap_{i=0}^{n-1} T^{-i} P_{x_i}.$$

Then a point $x \in X$ has a symbolic representation

$$\underline{x} = (x_0, x_1, \ldots) \text{ if } x \in \bigcap_{i=0}^{\infty} T^{-i} P_{x_i}.$$

Denote $\mathcal{P}_n := \bigvee_{j=0}^{n-1} T^{-j} \mathcal{P}$.

We require the following conditions on the probability preserving system (T, μ).

- Say $T : X \to X$ has exponential decay of correlation for $\mathcal{BV}$ against L^1 observables, where $\mathcal{BV} := \{\varphi \in L^1(m) : \varphi \text{ has bounded variation.}\}$, if there is $\beta : \mathbb{N} \to \mathbb{R}$ with $\beta(n) = C_1 e^{-c_1 n}$ for some $C_1, c_1 > 0$, and for all $f, g : X \to \mathbb{R}$, $f \in \mathcal{BV}$ and $g \in L^1$,

$$\left| \int f \cdot g \circ T^n \, d\mu - \int f \, d\mu \int g \, d\mu \right| \leq \|f\|_{\mathcal{BV}} \|g\|_1 \beta(n)$$

  where the norm $\|f\|_{\mathcal{BV}} := \|f\|_1 + TV(f)$, and $TV(f)$ stands for the total variation of f. In particular, if f is an indicator function of some measurable $A \subseteq X$, $\|f\|_{\mathcal{BV}} = 2$ and $\|f\|_1 \leq 1$.

- T has bounded distortion, that is, there is $C_{bd} > 0$ such that for all $x, y \in I \in \mathcal{P}_n$,

$$\frac{1}{C_{bd}} \leq \frac{DT^n(x)}{DT^n(y)} \leq C_{bd}.$$

Definition 4.4. (Gibbs-Markov maps) The map $T : X \to X$ is said to be *Gibbs-Markov* if

- $T(P)$ is a union of elements in $\mathcal{P}$ and $T|_P$ is injective for all $P \in \mathcal{P}$.
- Let d_s be the symbolic metric as, $d_s(\underline{x}, \underline{y}) = e^{-\underline{x} \wedge \underline{y}}$. The map $\log g|_P$ is Lipschitz with respect to d_s for each $P \in \mathcal{P}$, where $g = \frac{dm}{d(m \circ T)}$.
- When the partition does in fact contains infinitely many continuous components, we assume there is δ_0 such that $|T^n(I)| \geq \delta_0$ for all $I \in \mathcal{P}_n$, for all $n \geq 1$. Otherwise, δ_0 is a positive constant such that $|I| \geq \delta_0$ for all $I \in \mathcal{P}$. This is also known as the big image property.

Remark 4.5. For a Gibbs-Markov map defined above one can check that there is a measure μ verifying the inequality for n-cylinders as (2.4) holds for T, and because the system is conjugate to a topological Markov shift satisfying the conditions of Theorem 2.9, the invariant density $h = d\mu/dm$ is bounded from 0 and $+\infty$.

Definition 4.6. The *repeller* Λ of T is defined as the following:

$$\Lambda := \left\{ x \in I : T^k(x) \in \bigcup_{P \in \mathcal{P}} P \text{ for all } k \geq 0 \right\}.$$

Before the proof our quantitative result for limiting behaviour of m_n, following the convention in [11], define the following quantities,

$$\alpha(n) = (\log n)^2$$

$$m_n^{\leq}(x) := \min_{\substack{0 \leq i < j < n \\ |i-j| \leq \alpha(n)}} d\left(T^i x, T^j x\right),$$

$$m_n^{>}(x) := \min_{\substack{0 \leq i < j < n \\ |i-j| > \alpha(n)}} d\left(T^i x, T^j x\right),$$

$$m_n^{\gg}(x) := \min_{\substack{0 \leq i \leq n/3 \\ 2n/3 \leq j < n}} d\left(T^i x, T^j x\right),$$

Then, Theorem 1.2 can be rephrased as the following.

Theorem 4.7. *Let $T : X \to X$ be a piecewise expanding map as Definition 4.3 with a Gibbs-Markov structure defined in Definition 4.4, and μ its invariant Gibbs measure admitting exponential decay of correlations for $\mathcal{BV}$ against L^1 observables, then one has*

$$\limsup_{n \to \infty} \frac{\log m_n^{>}(x)}{-\log n} \leq \frac{2}{\underline{D_2}}. \tag{4.1.1}$$

If μ is absolutely continuous to Lebesgue measure m, $\underline{D_2} = \overline{D_2} = 1$, and

$$\limsup_{n \to \infty} \frac{\log m_n^{\leq}(x)}{-\log n} \leq \frac{2}{D_2}, \tag{4.1.2}$$

for μ-almost every $x \in \Lambda$. Then as $m_n(x) = \min\{m_n^{\leq}, m_n^{>}(x)\}$,

$$\limsup_{n \to +\infty} \frac{\log m_n(x)}{-\log n} \leq \frac{2}{D_2} = 2.$$

Theorem 4.8. *For all Gibbs measures μ, for μ-almost every $x \in \Lambda$,*

$$\liminf_{n \to \infty} \frac{\log m_n^{\gg}(x)}{-\log n} \geq \frac{2}{\overline{\overline{D}}_2}. \tag{4.2}$$

Since $-\log m_n \geq -\log m_n^{\gg}$,

$$\lim_{n \to \infty} \frac{\log m_n(x)}{-\log n} \geq \frac{2}{\overline{D}_2}$$

for μ-almost every $x \in \Lambda$.

Then these two theorems together imply Theorem 1.2, i.e. if μ is absolutely continuous to the Lebesgue measure m,

$$\lim_{n \to \infty} \frac{\log m_n(x)}{\log n} = 2$$

μ-almost every x.

Remark 4.9. The proof of (4.1.1) uses ideas from the proof of [11, Proposition 1.10], whereas the proof of (4.1.2) requires estimating the measures of sets of short return points. Along the proof, one will see also that (4.1.1) and (4.2) hold for all Gibbs invariant measures with exponential decay of correlations and $D_2(\mu) > 0$.

For Gibbs acip μ, the correlation dimension is well defined in the sense that $\overline{D}_2(\mu) = \underline{D}_2(\mu) = 1$, because the invariant density with respect to Lebesgue measure m is uniformly bounded hence $D_2(\mu) = D_2(m)$.

The following lemmas are analogous to [11, Lemma 3.1, Lemma 3.2], which can be proved by replacing their $\rho_p^{(r)}$ functions with indicator functions.

Lemma 4.10. *For all $x, y \in X$, let $\mathbb{1}_{p,r} := \mathbb{1}_{B(x_p^r, 2r)}$, if X has bounded local complexity defined in Definition 4.2,*

$$\mathbb{1}_{B(x,r)}(y) \leq \sum_{p=1}^{k(r)} \mathbb{1}_{p,r}(x)\mathbb{1}_{p,r}(y) \leq C_0 \mathbb{1}_{B(x,4r)}(y)$$

Lemma 4.11. *The following equations hold.*

$$\begin{cases} \limsup_{r \to 0} \dfrac{\log \sum_{p=1}^{k(r)} \left(\int \mathbb{1}_{p,r} d\mu\right)^2}{\log r} = \overline{D}_2(\mu) \\[4mm] \liminf_{r \to 0} \dfrac{\log \sum_{p=1}^{k(r)} \left(\int \mathbb{1}_{p,r} d\mu\right)^2}{\log r} = \underline{D}_2(\mu), \end{cases} \tag{4.3.1}$$

which means for any $\varepsilon > 0$, there is $r_0 > 0$ such that for all $0 < r < r_0$,

$$r^{\overline{D}_2 + \varepsilon} \leq \sum_{p=1}^{k(r)} \left(\int \mathbb{1}_{p,r} \, d\mu\right)^2 \leq r^{\underline{D}_2 - \varepsilon}. \tag{4.3.2}$$

Also, for simplicity, the following definition is introduced.

Definition 4.12. A term is said to be *admissible*, which is a notion introduced by the authors of [11], if it has the form $r^{-k}g(n)$, for some $k \geq 0$ and a function g which decays in n faster than any polynomial of n, hence for any $k \in \mathbb{N}$, by (4.3.2) and choosing the scale of r as in (4.4) below we can bound any admissible error by $\mathcal{O}(n^{-k})$ for all n large.

Now we can prove (4.1.1).

Upper bound.

Proof of (4.1.1). Let $\varepsilon, r > 0$ be given, in particular, r should be small enough that it verifies (4.3.2). Define the random variable $\mathcal{S}_n^>$,

$$\mathcal{S}_n^>(x) := \sum_{\substack{0 \leq i < j < n \\ |i-j| > \alpha(n)}} \mathbb{1}_{p,r}(T^i x) \mathbb{1}_{p,r}(T^j x).$$

By Lemma 4.10, $\{m_n^>(x) \leq r\} \subseteq \{\mathcal{S}_n^> \geq 1\}$. Therefore, by Markov's inequality and decay of correlation,

$$\mu(x : \mathcal{S}_n^>(x) \geq 1) \leq \mathbb{E}[S_n^>] = \sum_{\substack{0 \leq i < j < n \\ j-i > \alpha(n)}} \sum_{p=1}^{k(r)} \int \mathbb{1}_{p,r}(T^i x) \mathbb{1}_{p,r}(T^j x) \, d\mu(x)$$

$$\leq \sum_{\substack{0 \leq i < j < n \\ j-i > \alpha(n)}} \sum_{p=1}^{k(r)} \left(\left(\int \mathbb{1}_{p,r} \, d\mu \right)^2 + \beta(\alpha(n)) \|\mathbb{1}_{p,r}\|_{\mathcal{BV}} \|\mathbb{1}_{p,r}\|_1 \right)$$

$$\leq \sum_{0 \leq i < j < n} r^{\underline{D}_2 - \varepsilon} + \sum_{0 \leq i < j < n} \sum_{p=1}^{k(r)} \|\mathbb{1}_{p,r}\|_{\mathcal{BV}} \|\mathbb{1}_{p,r}\|_1 \beta(\alpha(n)) \qquad \text{by (4.3.2)}$$

$$\leq n^2 r^{\underline{D}_2 - \varepsilon} + C_1 e^{-c_1 (\log n)^2} n^2 \sum_{p=1}^{k(r)} 2\mu(B(x_p, 2r))$$

$$\leq n^2 r^{\underline{D}_2 - \varepsilon} + 2C_0 C_1 e^{-c_1 (\log n)^2} n^2.$$

The last inequality follows from the definition of bounded local complexity

$$\sum_{p=1}^{k(r)} \mu(B(x_p, 2r)) = \int \sum_{p=1}^{k(r)} \mathbb{1}_{p,r}(y) \, d\mu(y) \leq \int C_0 \, d\mu(y).$$

Then picking

$$r = r_n = \exp\left(-\frac{2 + 2\varepsilon}{\underline{D}_2 - \varepsilon} (\log n + \log\log n) \right) \leq n^{-\frac{2+2\varepsilon}{\underline{D}_2 - \varepsilon}} \tag{4.4}$$

for all n sufficiently large,

$$n^2 r_n^{\underline{D}_2 - \varepsilon} + 2C_0 C_1 n^2 e^{-c_1 (\log n)^2} \leq n^{-2\varepsilon} + 2C_0 C_1 n^2 e^{-c_1 (\log n)^2}.$$

and as $e^{-c_1 (\log n)^2}$ decays faster than any polynomial of n, the second term on the right is admissible by definition, meaning that there is some constant $C_2 > 0$ such that for all n large enough that $n^{-\varepsilon} \leq \frac{1}{\log n}$:

$$\mu(m_n^>(x) \leq r) \leq \mathbb{E}[S_n^>] \leq C_2 n^{-\varepsilon} \leq \frac{C_2}{\log n}.$$

Therefore, eventually for μ-almost every x,

$$\frac{\log m_{n_k}^>(x)}{-\log n_k} \leq \frac{2 + \varepsilon}{\underline{D}_2 - \varepsilon} \left(1 + \frac{\log\log n_k}{\log n_k} \right).$$

Although $m_n^>$ is not monotonically increasing, for each $n \in [n_s, n_{s+1}]$,

$$-\log m'_{n_{k+1}}(x) := -\log \min_{\substack{0 \leq i < j < n_{k+1} \\ j-i > \alpha(n_k)}} \geq -\log m_n^>(x)$$

$$-\log m''_{n_k}(x) := -\log \min_{\substack{0 \le i < j < n_k \\ j-i > \alpha(n_{k+1})}} \le -\log m_n^>(x),$$

and one can show that for μ-almost every x, for all k large,

$$\frac{-\log m'_{n_k}(x)}{\log n_k} \le \frac{2+\varepsilon}{\underline{D}_2 - \varepsilon}\left(1 + \frac{\log\log n_k}{\log n_k}\right),$$

and

$$\frac{-\log m''_{n_k}(x)}{\log n_k} \le \frac{2+\varepsilon}{\underline{D}_2 - \varepsilon}\left(1 + \frac{\log\log n_k}{\log n_k}\right),$$

then the limit can be passed to the whole tail of $m_n^>$, and (4.1.1) is proved. $\qquad\square$

Remark 4.13. Note also that for $|i-j|$ large, the measure of the sets $\{x : d\left(T^i x, T^j x\right) < r\}$ scales like r^{D_2} which is similar to the behaviour of the sequence matching problem in the symbolic setting, and this matches our intuition because $D_2(\mu)$ is analogous to H_2 in many ways.

To prove (4.1.2), we are dealing with iterates of x which return to an r-neighborhood of itself within $\alpha(n)$ units of time, that is, we need to approximate the measure of some short return sets as those $S_m(k)$ sets defined in section 3.1 for the symbolic case. But we cannot expect a similar upper bound for short returns as in (3.3) or (3.4). This is because for symbolic structures, an r_n-cylinder is itself an r_n open ball with respect to the symbolic metric, so analysing the returns is equivalent to analysing the repetition of letters in cylinders and one does not need to consider the case that two iterates $\sigma^i \underline{x}$, $\sigma^j \underline{x}$ are close to the boundaries of two open balls with a common boundary but they belong to different cylinders.

For interval maps, although the Gibbs-Markov structure prescribes a natural partition hence a way to define cylinders, metric balls and symbolic cylinders are different objects so one needs to take more caution and include the case that two points belong to different cylinders $U, V \in \mathcal{P}_n$ but they accumulate on a common boundary of U, V with distance smaller than the contraction scale of n-cylinders.

The measures of the short return sets are approximated by the following lemma for absolutely continuous Gibbs measures.

Lemma 4.14. [13, Lemma 3.4] *Define the sets*

$$\mathcal{E}_n(\epsilon) := \{x \in X : |x - T^n x| \le \epsilon\}.$$

Then for T satisfying Gibbs-Markov property and the invariant Gibbs measure μ absolutely continuous with respect to the Lebesgue measure m with exponential decay of correlation for $\mathcal{BV}$ against L^1 observables, there is some constant C_3 such that for all $n \in \mathbb{N}$ and ϵ small enough,

$$\mu\left(\mathcal{E}_n(\epsilon)\right) \le C_3 \epsilon.$$

Proof. The original lemma states that $m\left(\mathcal{E}_n(\epsilon)\right) = \mathcal{O}(\epsilon)$, and since m is equivalent to μ on Λ with $d\mu/dm$ bounded away from 0 and $+\infty$, there is a uniform constant C_μ such that

$$C_\mu^{-1} m(A) \le \mu(A) \le C_\mu m(A)$$

for each measurable A, therefore there is some $C_3 > 0$ such that

$$\mu\left(\mathcal{E}_n(\epsilon)\right) \le C_3 \epsilon.$$

$\qquad\square$

Proof of (4.1.2). Define the random variable $\mathcal{S}_n^{\leq}$ by

$$\mathcal{S}_n^{\leq}(x) := \sum_{i=0}^{n-1} \sum_{k=1}^{\alpha(n)\wedge(n-i-1)} \mathbb{1}_{B(T^i x, r)}(T^{i+k}x),$$

where $a \wedge b = \min\{a, b\}$.

As $\{x : \mathbb{1}_{B(T^i x, r)}(T^{i+k}x) = 1\} \subseteq \{x : T^i x \in \mathcal{E}_k(r)\} = T^{-i}\mathcal{E}_k(r)$, using Lemma 4.14 we obtain the following bound

$$\mu(x : m_n(x) \leq r) \leq \mu(\mathcal{S}_n^{\leq} \geq 1)$$

$$\leq \mathbb{E}_m[\mathcal{S}_n^{\leq}] \leq \sum_{i=0}^{n-1} \sum_{k=1}^{\alpha(n)\wedge(n-i-1)} \mu(T^{-i}\mathcal{E}_k(r)) \leq n\alpha(n)C_3 r$$

Pick $r = r_n$ as in (4.4), for all n large enough such that

$$\alpha(n) = (\log n)^2 \leq n^{\frac{\varepsilon}{2-\varepsilon}}, \qquad n^{-\frac{\varepsilon}{2-\varepsilon}} \leq \frac{1}{\log n}.$$

As the invariant density $d\mu/dm$ is uniformly bounded, $D_2(\mu) = 1 < 2$, one has

$$\mu\left(x : m_n(x) \leq r_n\right) \leq C_3 n^{1+\frac{\varepsilon}{2-\varepsilon}} r \leq n^{\frac{2}{2-\varepsilon}} C_3 n^{-\frac{2+2\varepsilon}{D_2-\varepsilon}} \leq C_3 n^{-\frac{\varepsilon}{2-2\varepsilon}} \leq \frac{C_3}{\log n}.$$

Therefore, by picking a subsequence $n_k = \lceil e^{k^2} \rceil$, by Borel-Cantelli Lemma, we have $\mu\{x : m_{n_k} \leq r_{n_k}$ for infinitely many $k\} = 0$, for μ-almost every x, for all k large enough,

$$m_{n_k}^{\leq}(x) \geq r_{n_k},$$

and one can apply the arguments used to validate (3.1.1) to obtain (4.1.2) for μ-almost every x. $\qquad\square$

Remark 4.15. The condition that μ is an acip may be not sharp; if $\tilde{\mu}$ is another Gibbs measure with exponential decay of correlation and satisfying $\tilde{\mu}(\mathcal{E}_k(\epsilon)) = \mathcal{O}(\epsilon)$, then Theorem 4.7 remains valid.

As in the symbolic case the proof for the lower bound of $\frac{\log m_n(x)}{-\log n}$ is slightly more complicated and also requires a second-moment computation which exploits the following notion of mixing.

Lemma 4.16. *A Gibbs-Markov interval map (T, μ) has exponential $4 - mixing$, that is, for $a < b \leq c$ in $\mathbb{N}$, there are $C_1', c_1' > 0$ such that for any $f_1, f_2 \in \mathcal{BV}$, $g_1, g_2 \in L^\infty$, such that*

$$\left| \int f_1 \cdot f_2 \circ T^a \cdot g_1 \circ T^b \cdot g_2 \circ T^c d\mu - \int f_1 \cdot g_1 \circ T^a d\mu \int f_2 \cdot g_2 \circ T^{c-b} d\mu \right|$$

$$\leq C_1' e^{-c_1'(b-a)}.$$

The constant C_1' depends on the functions f_1, f_2, g_1, g_2.

In particular, for any given $r > 0$, $0 \leq p, q \leq k(r)$, $f_1 = g_1 = \mathbb{1}_{p,r}$, $f_2 = g_2 = \mathbb{1}_{q,r}$, the constant $C_1' = C_1'(f_1, f_2, g_1, g_2)$ does not depend on r.

Proof. Consider the transfer operator $\mathcal{L}$ associated with the Hölder potential φ, that acts on the space of functions of bounded variation of X, $\mathcal{BV} = \mathcal{BV}(X)$,

$$\mathcal{L} = \mathcal{L}_\varphi : \mathcal{BV} \to \mathcal{BV}, \quad \mathcal{L}_\varphi f(x) = \sum_{Ty=x} e^{\varphi(y)} f(y).$$

Let ν be the conformal measure of $\mathcal{L}$ and h the invariant density, $\frac{d\mu}{d\nu} = h$. By the following well-known fact, (see for example [14, (3)]) for mixing Gibbs-Markov maps, there are $C_{\mathcal{BV}} > 0$, $\kappa \in (0,1)$ such that for any $f \in \mathcal{BV}$,

$$\left\| \mathcal{L}^n f - h \int f \, d\nu \right\|_{\mathcal{BV}} \leq C_{\mathcal{BV}} \cdot \kappa^n \|f\|_{\mathcal{BV}}$$

for all $f \in \mathcal{BV}$, there is

$$\left| \int f_1 \, f_2 \circ T^a \, g_1 \circ T^b \, g_2 \circ T^c \, d\mu - \int f_1 \, f_2 \circ T^a \, d\mu \int g_1 \, g_2 \circ T^{c-b} \, d\mu \right|$$

$$= \left| \int \mathcal{L}^{b-a}(h f_1 \, f_2 \circ T^a) g_1 \circ T^a \, g_2 \circ T^{c-b+a} \, d\nu \right.$$

$$\left. - \int \mathcal{L}^a(h f_1) f_2 \, d\nu \int h \, g_1 \circ T^a \, g_2 \circ T^{c-b+a} \, d\nu \right|$$

$$= \left| \int \left(\mathcal{L}^{b-a}(\mathcal{L}^a(h f_1) f_2) - h \int \mathcal{L}^a(h f_1) f_2 \, d\nu \right) g_1 \circ T^a \, g_2 \circ T^{c-b+a} \, d\nu \right|$$

$$\leq \left\| \mathcal{L}^{b-a}(\mathcal{L}^a(h f_1) \, f_2) - h \int \mathcal{L}^a(h f_1) \, f_2 \, d\nu \right\|_1 \|g_1\|_\infty \|g_2\|_\infty$$

$$\leq C_{\mathcal{BV}} \cdot \kappa^{b-a} \|\mathcal{L}^a(h f_1) \, f_2\|_{\mathcal{BV}} \|g_1\|_\infty \|g_2\|_\infty, \tag{4.5}$$

where the first equality holds by the invariance of μ. As $\mathcal{L}^a h f_1$ is of bounded variation because h is of bounded variation, and the product of functions in $\mathcal{BV}$ has bounded variation, the first part of the lemma is proved.

Now we deal with the case $f_1 = g_1 = \mathbb{1}_{p,r}$, $f_2 = g_2 = \mathbb{1}_{q,r}$ and find a suitable upper bound for

$$\|\mathcal{L}^a h \mathbb{1}_{p,r} \cdot \mathbb{1}_{q,r}\|_{\mathcal{BV}} = \|\mathcal{L}^a h \mathbb{1}_{p,r} \cdot \mathbb{1}_{q,r}\|_1 + TV(\mathcal{L}^a h \mathbb{1}_{p,r} \cdot \mathbb{1}_{q,r}).$$

As $\mathcal{L}$ is a positive operator, by its duality, $\|\mathcal{L}^a h \mathbb{1}_{p,r} \cdot \mathbb{1}_{q,r}\|_1 = \int \mathbb{1}_{p,r} \cdot \mathbb{1}_{q,r} \circ T^a \, d\mu = \int \mathbb{1}_{p,r} \mathbb{1}_{q,r} \circ T^a \, d\mu \leq 1$. Also, for any function $u \in \mathcal{BV}$,

$$\|u\|_\infty \leq \inf_{x \in X} |u(x)| + TV(u) \leq \inf_{x \in X} |u(x)| \mu(X) + \|u\|_{\mathcal{BV}} \leq \|u\|_1 + \|u\|_{\mathcal{BV}},$$

then by Lasota-Yorke inequality for Gibbs Markov systems, there are $\kappa' \in (0,1)$ and $C'_{\mathcal{BV}} > 0$ such that

$$\|\mathcal{L}^a h \mathbb{1}_{p,r}\|_{\mathcal{BV}} \leq \kappa'^a \|h \mathbb{1}_{p,r}\|_{\mathcal{BV}} + C'_{\mathcal{BV}} \|h \mathbb{1}_{p,r}\|_1 \leq \kappa'^a (1 + 2\|h\|_\infty) + C'_{\mathcal{BV}}.$$

therefore, let $C''_{\mathcal{BV}} = (1 + 2\|h\|_\infty) + C'_{\mathcal{BV}}$,

$$TV(\mathcal{L}^a h \mathbb{1}_{p,r} \cdot \mathbb{1}_{q,r}) \leq \|\mathcal{L}^a h \mathbb{1}_{p,r}\|_\infty TV(\mathbb{1}_{q,r}) + \|\mathbb{1}_{q,r}\|_\infty \|\mathcal{L}^a h \mathbb{1}_{p,r}\|_{\mathcal{BV}}$$

$$\leq (\|\mathcal{L}^a h \mathbb{1}_{p,r}\|_{\mathcal{BV}} + 1) \cdot 2 + 1 \cdot \|\mathcal{L}^a h \mathbb{1}_{p,r}\|_{\mathcal{BV}} \leq 3 \cdot C''_{\mathcal{BV}} + 2,$$

which is a uniform constant that only depends on the operator $\mathcal{L}$, and combining this with (4.5), one obtains $C'_1 = C'_1(\mathbb{1}_{p,r}, \mathbb{1}_{q,r}, \mathbb{1}_{p,r}, \mathbb{1}_{q,r}) = \mathcal{O}(1)$. □

Lower bound (4.2).

proof of (4.2). Let $\varepsilon > 0$ small be given. Consider the quantity $m_n^{\gg}$ and the random variable $\mathcal{S}_n^{\gg}$:

$$m_n^{\gg}(x) := \min_{\substack{0 \leq i \leq n/3 \\ 2n/3 \leq j < n}} d\left(T^i x, T^j x\right), \quad \mathcal{S}_n^{\gg}(x) := \sum_{\substack{0 \leq i \leq n/3 \\ 2n/3 \leq j < n}} \sum_{p=1}^{k(r)} \mathbb{1}_{p,r}(T^i x) \mathbb{1}_{p,r}(T^j x).$$

By Lemma 4.10, $m_n^{\gg}(x) > 4r$ implies for all pairs of $0 \le i \le \frac{n}{3}$, $\frac{2n}{3} \le j < n$, if for some p, $d(T^i x, x_p^r) < 2r$, then $d(T^j x, x_p^r) \ge 2r$ hence $\mathcal{S}_n^{\gg}(x) = 0$. By Paley-Zygmund inequality,

$$\mu\left(m_n > 4r\right) \le \mu\left(x : \mathcal{S}_n^{\gg}(x) = 0\right) \le \frac{\mathbb{E}[(\mathcal{S}_n^{\gg})^2] - \mathbb{E}[\mathcal{S}_n^{\gg}]^2}{\mathbb{E}[\mathcal{S}_n^{\gg}]^2}$$

Using decay of correlation and invariance,

$$\mathbb{E}[\mathcal{S}_n^{\gg}(x)] = \sum_{\substack{0 \le i \le n/3 \\ 2n/3 \le j < n}} \sum_p \int \mathbb{1}_{p,r}(T^i x)\,\mathbb{1}_{p,r}(T^j x)\, d\mu(x)$$

$$\le \left(\frac{n}{3}\right)^2 \sum_p \left(\left(\int \mathbb{1}_{p,r}\, d\mu(x)\right)^2 \pm 2\beta\left(n/3\right)\right) \tag{4.6}$$

where $a = b \pm c$ means $a \in [b - c, b + c]$. Consider

$$\left(\mathcal{S}_n^{\gg}(x)\right)^2 = \sum_{i,j} \sum_{s,t} \sum_{p,q} \mathbb{1}_{p,r}(T^i x)\,\mathbb{1}_{p,r}(T^j x)\,\mathbb{1}_{q,r}(T^s x)\,\mathbb{1}_{q,r}(T^t x).$$

As in the proof of symbolic case, we will split this sum in terms of the distance between the indices i, j, s, t. Recall that

$$\alpha(n) = (\log n)^2.$$

Let F be the collection of all possible quadruples of indices (i, j, s, t), and define the counting function

$$\tau : F \to \mathbb{N} \cup \{0\}, \quad \tau(i,j,s,t) = \sum_{\substack{a \in \{i,s\} \\ b \in \{j,t\}}} \mathbb{1}_{[a-\alpha(n), a+\alpha(n)]}(b),$$

then $\tau \le 2$ since i, j and s, t are at both at least $\frac{n}{3}$ iterates apart, this allows us to split F into $F_m := \{(i,j,s,t) \in F : \tau = m\}$ for $m = 0, 1, 2$. Obviously, the following upper bounds hold for the cardinality of each F_m,

$$\#F_m \le (2\alpha(n))^m \left(\frac{n}{3}\right)^{4-m}. \tag{4.7}$$

Recall the notation

$$\mathbb{E}[(\mathcal{S}_n^{\gg})^2 \,|\, F_m] = \sum_{(i,j,s,t) \in F_m} \sum_{p,q} \int \mathbb{1}_{p,r}(T^i x)\,\mathbb{1}_{p,r}(T^j x)\,\mathbb{1}_{q,r}(T^s x)\,\mathbb{1}_{q,r}(T^t x)\, d\mu(x),$$

also for simplicity, let us denote

$$R_p = \int \mathbb{1}_{p,r}\, d\mu = \mu(B(x_p, 2r)).$$

Contribution of indices in F_0: For each $(i, j, s, t) \in F_0$, without loss of generality, suppose $i + \alpha(n) < s$ and $j + \alpha(n) < t$, as the alternative cases can be treated equally by exchanging the roles of i, s or j, t and makes no difference to the calculation. As $\min\{j, t\} - \max\{i, s\} \ge \frac{n}{3}$, by Lemma 4.16 and invariance, one obtains the following upper bound for each such quadruple (i, j, s, t):

$$\sum_{p,q} \int \mathbb{1}_{p,r}(T^i x)\,\mathbb{1}_{p,r}(T^j x)\,\mathbb{1}_{q,r}(T^s x)\,\mathbb{1}_{q,r}(T^t x)\, d\mu(x)$$

$$= \sum_{p,q} \int \mathbb{1}_{p,r}\,\mathbb{1}_{q,r} \circ T^{s-i}\,\mathbb{1}_{p,r} \circ T^{j-i}\,\mathbb{1}_{q,r} \circ T^{t-i}\, d\mu$$

$$\leq C_1' e^{-c_1' \frac{n}{3}} k(r)^2 + \sum_{p,q} \int \mathbb{1}_{p,r} \mathbb{1}_{q,r} \circ T^{s-i} \, d\mu \int \mathbb{1}_{p,r} \mathbb{1}_{q,r} \circ T^{t-j} \, d\mu$$

$$\leq C_1' e^{-c_1' \frac{n}{3}} k(r)^2 + \sum_p \sum_q \left(R_p R_q + 2\beta(\alpha(n)) \right)^2$$

$$\leq C_1' e^{-c_1' \frac{n}{3}} r^{-2C_0'} + 8\beta(\alpha(n)) r^{-2C_0'} + \sum_{p,q} (R_p R_q)^2 .$$

Where the last inequality holds as $R_p, R_q \leq 1$ for any p, q, and by [11, Lemma 3.3] $k(r) \leq r^{-C_0'}$ for some $C_0' = 4 \log C_0$. Any term in the inequality above involving $\beta(\alpha(n))$ or $C_1' e^{-c_1' \frac{n}{3}}$ is admissible, hence for each $k \in \mathbb{R}$ it is bounded by $\mathcal{O}(n^{-k})$ for all n sufficiently large, and now we shall pick

$$r = r_n = n^{-\frac{2-4\varepsilon}{\overline{D}_2 + \varepsilon}},$$

then by (4.3.2)

$$r^{\overline{D}_2 + \varepsilon} \leq \sum_p \left(\int \mathbb{1}_{p,r} \, d\mu \right)^2 = \sum_p R_p^2. \tag{4.8}$$

Therefore, the contribution of indices in F_0 is bounded from above up to an admissible error by

$$\left(\frac{n}{3} \right)^4 \sum_p \sum_q R_p^2 R_q^2 \leq \left(\frac{n}{3} \right)^4 \sum_p R_p^2 \sum_q R_q^2 = \left(\frac{n}{3} \right)^4 \left(\sum_p R_p^2 \right)^2,$$

combining with (4.6), up to an admissible error term,

$$\mathbb{E}[(\mathcal{S}_n^{\gg})^2 | F_0] - \mathbb{E}[\mathcal{S}_n^{\gg}]^2 \leq \mathcal{O}(n^{-\varepsilon})$$

also by (4.6), as $2\beta(\frac{n}{3})$ is admissible we can bound it by n^{-3},

$$\mathbb{E}[\mathcal{S}_n^{\gg}]^2 \geq \left(\frac{n}{3} \right)^4 \left(r_n^{\overline{D}_2 + \varepsilon} - 2\beta(\tfrac{n}{3}) \right) \geq \left(\frac{n}{3} \right)^4 (n^{-2-4\varepsilon} - n^{-3})^2 \approx n^{8\varepsilon}$$

allowing us to conclude that there is some constant $C_4 > 0$:

$$\frac{\mathbb{E}[(\mathcal{S}_n^{\gg})^2 | F_0] - \mathbb{E}[\mathcal{S}_n^{\gg}]^2}{\mathbb{E}[\mathcal{S}_n^{\gg}]^2} \leq \frac{C_4}{n^{\varepsilon}}. \tag{4.9}$$

Contributions of indices in F_1: Now we will deal with the indices in F_1. Without loss of generality, suppose $|i - s| \leq \alpha(n)$, $i < s$ and $j < t$, the other cases can be treated by exchanging the roles of i, s or j, t. By Lemma 4.16,

$$\sum_{p,q} \int \mathbb{1}_{p,r}(T^i x) \mathbb{1}_{p,r}(T^j x) \mathbb{1}_{q,r}(T^s x) \mathbb{1}_{q,r}(T^t x) \, d\mu(x)$$

$$= \sum_{p,q} \int \mathbb{1}_{p,r} \mathbb{1}_{q,r} \circ T^{s-i} \mathbb{1}_{p,r} \circ T^{j-i} \mathbb{1}_{q,r} \circ T^{t-i} \, d\mu \qquad \text{by invariance}$$

$$\leq \sum_{p,q} \left(\int \mathbb{1}_{p,r} \mathbb{1}_{q,r} \circ T^{t-j} \, d\mu \right) \int \mathbb{1}_{p,r}(x) \mathbb{1}_{q,r}(T^{s-i} x) \, d\mu(x) + C_1' e^{-c_1' \frac{n}{3}} k(r)^2$$

$$\leq \sum_{p,q} \left(R_p R_q + 2\beta(\alpha(n)) \right) \int \mathbb{1}_{p,r}(x) \mathbb{1}_{q,r}(T^{s-i} x) \, d\mu(x) + C_1' e^{-c_1' \frac{n}{3}} r^{-2C_0'},$$

and this can be bounded by the following up to an admissible error using Cauchy-Schwarz inequality,

$$\sum_{p,q} \int R_p R_q \mathbb{1}_{p,r}(x) \mathbb{1}_{q,r}(T^{s-i}x) \, d\mu(x)$$

$$= \int \sum_p R_p \mathbb{1}_{p,r}(x) \sum_q R_q \mathbb{1}_{q,r}(T^{s-i}) \, d\mu(x)$$

$$\leq \left(\int \left(\sum_p R_p \mathbb{1}_{p,r}(x) \right)^2 d\mu \right)^{-\frac{1}{2}} \left(\int \left(\sum_q R_q \mathbb{1}_{q,r} \circ T^{s-i} \right)^2 d\mu \right)^{-\frac{1}{2}}$$

$$= \int \left(\sum_p R_p \mathbb{1}_{p,r} \right)^2 d\mu. \qquad\qquad \text{by symmetry and invariance}$$

As there are at most C_0 non-zero terms of $\mathbb{1}_{p,r}(x)$ for any $x \in X$, by the following inequality for $a_1, \ldots, a_m \geq 0$,

$$(a_1 + a_2 + \cdots + a_m)^2 \leq m \left(a_1^2 + a_2^2 + \ldots, a_m^2 \right)$$

together with the fact that $(\mathbb{1}_{p,r})^2 \leq \mathbb{1}_{p,r}$, the sum above can be bounded by

$$\int C_0 \sum_p R_p^2 \mathbb{1}_{p,r} \, d\mu = \sum_p C_0 R_p^2 \int \mathbb{1}_{p,r} \, d\mu = C_0 \sum_p R_p^3.$$

As $(a_1 + \cdots + a_m)^{\frac{2}{3}} \leq \sum_{k=1}^m a_k^{\frac{2}{3}}$, clearly $\sum_k a_k \leq \left(\sum_k a_k^{2/3} \right)^{3/2}$, then we get up to an admissible error and some constant C_5, for all n large such that $\alpha(n) \leq n^\varepsilon$, as $(\sum_p R_p^2)^{-\frac{3}{2}} 4\beta(\alpha(n))$ and $\left(\sum_p R_p^2 \right)^{-\frac{1}{2}} \beta(\alpha(n))$ are both admissible errors,

$$\frac{\mathbb{E}[(\mathcal{S}_n^\gg)^2 | F_1]}{\mathbb{E}[\mathcal{S}_n^\gg]^2}$$

$$\leq \frac{2\alpha(n)(\frac{n}{3})^3 C_0 \left(\sum_p R_p^2 \right)^{\frac{3}{2}}}{(\frac{n}{3})^4 (\sum_p R_p^2 - 2\beta(\alpha(n)))^2}$$

$$= \frac{6\alpha(n)(\frac{n}{3})^3 C_0 \left(\sum_p R_p^2 \right)^{\frac{3}{2}}}{n \left(\left(\sum_p R_p^2 \right)^{1/2} - 4\beta(\alpha(n)) \left(\sum_p R_p^2 \right)^{-1/2} - 4\beta(\alpha(n))^2 \left(\sum_p R_p^2 \right)^{-3/2} \right)}$$

$$\leq \frac{6C_0 \alpha(n)}{n \left((r^{\overline{D}_2 + \varepsilon})^{1/2} - \mathcal{O}(n^{-1}) \right)} = \frac{6C_0 \alpha(n)}{n \left((n^{-1+2\varepsilon} - \mathcal{O}(n^{-1})) \right)} \qquad \text{by (4.8)}$$

$$\leq \frac{C_5 n^\varepsilon}{n \cdot n^{-1+2\varepsilon}} = \frac{C_5}{n^\varepsilon}. \qquad\qquad\qquad\qquad\qquad\qquad (4.10)$$

Contribution of indices in F_2: Finally, let us consider indices (i,j,s,t) such that $|i - s|, |j - t| \leq \alpha(n)$. By Lemma 4.10, $\sum_q \mathbb{1}_{q,r}(T^s x) \mathbb{1}_{q,r}(T^t x) \leq C_0$ for any x, therefore for each i, j, s, t in F_2,

$$\sum_{p,q} \int \mathbb{1}_{p,r}(T^i x) \mathbb{1}_{p,r}(T^j x) \mathbb{1}_{p,r}(T^s x) \mathbb{1}_{p,r}(T^t x) \, d\mu(x)$$

$$\leq C_0 \sum_p \int \mathbb{1}_{p,r}(T^i x) \mathbb{1}_{p,r}(T^j x) \, d\mu(x)$$

$$\leq C_0 \sum_p R_p^2 + C_0 \beta(\frac{n}{3}) k(r),$$

therefore, as $\#F_2 \leq \frac{4}{9}\alpha(n)^2 n^2$, by our choice of r_n in (4.8), up to an admissible error there is some constant C_6 such that,

$$
\begin{aligned}
\frac{\mathbb{E}[(\mathcal{S}_n^{\gg})^2 | F_2]}{\mathbb{E}[S_n^{\gg}]^2} &\leq \frac{4\alpha(n)^2 (\frac{n}{3})^2 C_0 \sum_p R_p^2}{(\frac{n}{3})^4 (\sum_p R_p^2 - 2\beta(\alpha(n)))^2} \\
&= \frac{36 C_0 \alpha(n)^2}{n^2 \left(\sum_p R_p^2 - 4\beta(\alpha(n)) + 4\beta(\alpha(n))^2 \left(\sum_p R_p^2 \right)^{-1} \right)} \\
&\leq \frac{36 C_0 n^{2\varepsilon}}{n^2 \left(r^{\overline{D_2}+\varepsilon} - \mathcal{O}(n^{-2}) \right)} \\
&\leq \frac{C_6 n^{2\varepsilon}}{n^2 n^{-2+4\varepsilon}} = \frac{C_6}{n^{2\varepsilon}}.
\end{aligned}
\tag{4.11}
$$

Hence, putting (4.9), (4.10) and (4.11) together, we can conclude that for all n large enough and $r = r_n = n^{-\frac{2-4\varepsilon}{\overline{D_2}+\varepsilon}}$, there is some constant $C_7 > 0$ such that

$$\mu(m_n^{\gg} > 8 r_n) \leq \frac{Var[\mathcal{S}_n^{\gg}]}{\mathbb{E}[\mathcal{S}_n^{\gg}]^2} \leq \frac{C_7}{n^{\varepsilon}},$$

picking a subsequence $n_k = \lceil k^{2/\varepsilon} \rceil$, the probability is summable along the subsequence which means by the Borel-Cantelli Lemma, for μ-almost every x, for k large

$$-\log m_n^{\gg}(x) \geq -\log 4 r_{n_k}.$$

The proof of (4.2) is yet complete because $m_n^{\gg}$ is not a monotone sequence, but for each $n \in [n_k, n_{k+1}]$,

$$-\log m_n^{\gg}(x) \geq -\log \min_{\substack{0 \leq i \leq n_k \\ 2n_{k+1}/3 \leq j < n_k}} d\left(T^i x, T^j x\right) =: -\log m_{n_k}^*(x),$$

and as for all k, $m_{n_{k+1}}^{\gg}(x) \leq m_{n_k}^*(x)$, repeating the same proof we have done for $m_n^{\gg}$ one can also show that $\liminf_{k \to \infty} \frac{\log m_{n_k}^*}{\log n_k} \geq \frac{2-4\varepsilon}{\overline{D_2}+\varepsilon}$ whence the lower bound will be passed to the entire tail of $-\log m_{n_k}^{\gg}(x)$ and hence $-\log m_n(x)$, which implies

$$\liminf_{n \to \infty} \frac{\log m_n(x)}{-\log n} \geq \frac{2}{\overline{D_2}}$$

for μ-almost every x by sending $\varepsilon \to 0$. $\qquad \square$

Remark 4.17. (4.1.1) and (4.2) still hold if decay of correlations is exponential with respect to other Banach function spaces $\mathcal{B}$, $\mathcal{B}'$, for example, both observables are in $\mathcal{BV}$ or Lip, where $Lip := \{f \in C(X) : f \text{ is Lipschitz}\}$, as long as $\beta(\alpha(n)) \|\mathbb{1}_{p,r}\|_{\mathcal{B}} \|\mathbb{1}_{q,r}\|_{\mathcal{B}'}$ remains an admissible term. For example, if the system has decay of correlations for Lipschitz observables, one can replace $\mathbb{1}_{p,r}$ functions with $\{\rho_p^r\}_{p=1}^{k(r)}$, a set of discretisation functions defined in [11], although it requires

more machinery to adjust the proof for (4.2) and Lemma 4.16. Also, instead of exponential decay of correlations, the proof remains valid under stretched exponential decay by manipulating the scale of k in $\alpha(n) = (\log n)^k$.

REFERENCES

[1] J. Aaronson and M. Denker, Local limit theorems for partial sums of stationary sequences generated by Gibbs-Markov maps, *Stoch. Dyn.*, **1** (2001), 193-237.

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

[3] V. Barros, L. Liao and J. Rousseau, On the shortest distance between orbits and the longest common substring problem, *Adv. Math.*, **344** (2019), 311-339.

[4] V. Barros and J. Rousseau, Shortest distance between multiple orbits and generalized fractal dimensions, *Ann. Henri Poincaré*, **22** (2021), 1853-1885.

[5] R. Bowen, *Equilibrium States and The Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Mathematics, vol. 470. Springer, Berlin, 2008.

[6] P. Collet, C. Giardina and F. Redig, Matching with shift for one-Dimensional Gibbs measures, *Ann. Appl. Probab.*, **19** (2009), 1581-1602.

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

[8] A. Dembo, S. Karlin and O. Zeitouni, Limit distribution of maximal non-aligned two-sequence segmental score, *Ann. Probab.*, **22** (1994), 2022-2039.

[9] C. Gabriela and V. Girardin and L. Loïck, Computation and estimation of generalized entropy rates for denumerable Markov chains, *IEEE Trans. Inform. Theory*, **57** (2011), 4026-4034.

[10] A. Galves and B. Schmitt, Inequalities for hitting times in mixing dynamical systems, *Random Comput. Dynam.*, **5** (1997), 337-347.

[11] S. Gouëzel, J. Rousseau and M. Stadlbauer Minimal distance between random orbits, preprint, 2022, `arXiv:2209.13240`.

[12] N. Haydn and S. Vaienti, The Rényi entropy function and the large deviation of short return times, *Ergodic Theory and Dynam. Systems*, **30** (2010), 159-179.

[13] M. Holland, M. Nicol and A. Török, Extreme value theory for non-uniformly expanding dynamical systems, *Trans. Amer. Math. Soc.*, **364** (2012), 661-688.

[14] G. Keller, On the rate of convergence to equilibrium in one-dimensional systems, *Comm. Math. Phys.*, **96** (1984), 181-193.

[15] C. Liverani, B. Sassoul and S. Vaienti A probabilistic approach to intermittency, *Ergodic Theory and Dynam. Systems*, **19** (1999), 671-685.

[16] R. E. A. C. Paley and A. Zygmund, On some series of functions, (3), *Math. Proc. Cambridge Philos. Soc.*, **28** (1932), 190-205.

[17] O. M. Sarig, Thermodynamic formalism for countable Markov shifts, *Ergodic Theory Dynam. Systems*, **19** (1999), 1565-1593.

[18] O. Sarig, Existence of Gibbs measures for countable Markov shifts, *Proc. Amer. Math. Soc.*, **131** (2003), 1751-1758.