GEOMETRIC LAW FOR NUMBERS OF RETURNS UNTIL A HAZARD UNDER $\phi$-MIXING

BY

YURI KIFER

Institute of Mathematics, The Hebrew University of Jerusalem
Givat Ram, Jerusalem 91904, Israel
e-mail: kifer@math.huji.ac.il

AND

FAN YANG

Department of Mathematics, University of Oklahoma
Norman, OK 73012, USA
e-mail: yangfa31@msu.edu

ABSTRACT

We consider a $\phi$-mixing shift $T$ on a sequence space $\Omega$ and study the number of returns $\{T^k \omega \in U\}$ to a union $U$ of cylinders of length $n$ until the first return $\{T^k \omega \in V\}$ to another union $V$ of cylinder sets of length $m$. It turns out that if probabilities of the sets $U$ and $V$ are small and of the same order, then the above number of returns has approximately geometric distribution. Under appropriate conditions we extend this result for some dynamical systems to geometric balls and Young towers with integrable tails. This work is motivated by a number of papers on asymptotical behavior of numbers of returns to shrinking sets, as well as by the papers on open systems studying their behavior until an exit through a “hole”.

Received December 24, 2018 and in revised form July 21, 2020
1. Introduction

Let $\xi_0, \xi_1, \xi_2, \ldots$ be a sequence of independent identically distributed (i.i.d.) random variables and $\Gamma_N, \Delta_N$ be a sequence of sets such that $\Gamma_N \cap \Delta_N = \emptyset$ and

$$\lim_{N \to \infty} \lambda^{-1} N P\{\xi_0 \in \Gamma_N\} = \lim_{N \to \infty} \nu^{-1} N P\{\xi_0 \in \Delta_N\} = 1.$$ 

Then by the classical Poisson limit theorem

$$S^{(\lambda)}_N = \sum_{n=0}^{N-1} I_{\Gamma_N}(\xi_n) \quad \text{and} \quad S^{(\nu)}_N = \sum_{n=0}^{N-1} I_{\Delta_N}(\xi_n),$$

where $I_{\Gamma}$ is the indicator of a set $\Gamma$, converge in distribution as $N \to \infty$ to Poisson random variables with parameters $\lambda$ and $\nu$, respectively. On the other hand, if $\tau_N = \min\{n \geq 0 : \xi_n \in \Gamma_N\}$ then it turns out that the sum

$$S_{\tau_N} = \sum_{n=0}^{\tau_N-1} \sum_{i=1}^{\ell} I_{\Delta_N}(\xi_{q_i(n)})$$

which counts returns to $\Delta_N$ until the first arrival at $\Gamma_N$, converges in distribution as $N \to \infty$ to a geometric random variable $\zeta$ with the parameter $p = \lambda(\lambda + \nu)^{-1}$, i.e., $P\{\zeta = k\} = (1 - p)^k p$.

In [17] a more general setup was considered where $\xi_0, \xi_1, \xi_2, \ldots$ is a $\psi$-mixing stationary process and which included increasing functions $q_i(n)$, $i = 1, \ldots, \ell$, $n \geq 0$ taking on integer values on integers and satisfying $0 \leq q_1(n) < q_2(n) < \cdots < q_\ell(n)$ with all differences $q_i(n) - q_{i-1}(n)$ tending to $\infty$ as $n \to \infty$. There, “nonconventional” sums

$$(1.1) \quad S_{\tau_N} = \sum_{n=0}^{\tau_N-1} \prod_{i=1}^{\ell} I_{\Delta_N}(\xi_{q_i(n)})$$

were considered with

$$(1.2) \quad \tau_N = \min\left\{n \geq 0 : \prod_{i=1}^{\ell} I_{\Gamma_N}(\xi_{q_i(n)}) = 1\right\}$$

and $\tau_N = \infty$ if the set in braces is empty. Now $S_{\tau_N}$ equals the number $N_N$ of multiple returns to $\Delta_N$ until the first multiple return to $\Gamma_N$. It turned out that if

$$(1.3) \quad \lim_{N \to \infty} \lambda^{-1} N (P\{\xi_0 \in \Gamma_N\})^\ell = \lim_{N \to \infty} \nu^{-1} N (P\{\xi_0 \in \Delta_N\})^\ell = 1$$

then, again, $S_{\tau_N}$ converges in distribution to a geometric random variable with the parameter $p = \lambda(\lambda + \nu)^{-1}$. 

We consider in this paper the following setup which comes from dynamical systems but has also a perfect probabilistic sense. Let $\zeta_k, k = 0, 1, 2, \ldots$ be a $\phi$-mixing discrete time process evolving on a finite or countable state space $A$. For each sequence $a = (a_0, a_1, a_2, \ldots) \in A^\mathbb{N}$ of elements from $A$ and any $m \in \mathbb{N}$ denote by $a^{(m)}$ the string $a_0, a_1, \ldots, a_{m-1}$ which determines also an $m$-cylinder set $A^a_m$ in $A^\mathbb{N}$ consisting of sequences whose initial $m$-string coincides with $a_0, a_1, \ldots, a_{m-1}$. Let $\tau^a_m$ be the first $l$ such that starting at the time $l$ the process $\zeta_k = \zeta_k(\omega), k \geq 0$ repeats the string $a^{(m)} = (a_0, \ldots, a_{m-1})$. Let $b = (b_0, b_1, \ldots) \in A^\mathbb{N}$, $b \neq a$. We are interested in the number of $j < \tau^a_m$ such that process $\zeta_k$ repeats the string $b^{(n)} = (b_0, \ldots, b_{n-1})$ starting at the time $j$. Employing the left shift transformation $T$ on the sequence space $A^\mathbb{N}$ we can represent the number in question as a random variable on $\Omega = A^\mathbb{N}$ given by the sum

$$\sum_{j=0}^{\tau^a_m-1} I_{A^b_n}(T^j \omega).$$

We will show that for any $T$-invariant $\phi$-mixing probability measure $P$ on $\Omega$ and $P$-almost all $a, b \in \Omega$ the distribution of random variables $\Sigma^{b,a}_{n,m}$ approaches in the total variation distance as $n \to \infty$ the geometric distribution with the parameter

$$P(A^a_n)(P(A^a_m) + P(A^b_n))^{-1}$$

provided the ratio $P(A^b_n)/P(A^a_m)$ stays bounded away from zero and infinity. In particular, if this ratio tends to $\lambda$ when $m = m(n)$ and $n \to \infty$, then the distribution of $\Sigma^{b,a}_{n,m}$ converges in total variation distance to the geometric distribution with the parameter $(1+\lambda)^{-1}$. In fact, we will prove such results for arbitrary unions of cylinders of the same length. It turns out that under just $\phi$-mixing (and not $\psi$-mixing) our method does not work for stationary processes as considered in [17] and it is not applicable for shifts when the nonconventional sums are considered as in [17].

After proving results for $\phi$-mixing shifts on symbolic spaces we consider dynamical systems which can be modeled by such shifts via corresponding partitions and approximating geometric balls by elements of such partitions we obtain under appropriate conditions geometric limit laws for numbers of returns to a sequence of shrinking balls until first arrival to a ball from another such sequence. We show also that our conditions are satisfied for some Young towers, for instance those which are constructed as discrete time suspensions of Gibbs–Markov maps, obtaining thus a limiting geometric law for dynamical systems which can be modelled by such towers.
Our results are applicable to large classes of dynamical systems. Among such systems are smooth expanding endomorphisms of compact manifolds and Axiom A (in particular, Anosov) diffeomorphisms which have symbolic representations via Markov partitions (see [6]). Then, in place of cylinder sets we can count returns to an element of a Markov partition until first return to another element of this partition. If for such dynamical systems we consider Sinai–Ruelle–Bowen type measures then the results are extended to returns to geometric balls in place of elements of Markov partitions using approximations of the former by unions of the latter (cf. the proof of Theorem 3 in [13]). The results remain true for some systems having symbolic representations with infinite alphabet, for instance, for the Gauss map $T x = \frac{1}{x} \pmod{1}$, $x \in (0, 1]$, $T 0 = 0$ of the unit interval considered with the Gauss measure

$$G(\Gamma) = \frac{1}{\ln 2} \int_{\Gamma} \frac{dx}{1 + x}$$

which is known to be $T$-invariant and $\psi$-mixing with an exponential speed ([15]).

More generally, our weaker $\phi$ mixing assumption enables us to consider additional classes of dynamical systems among them some non uniformly expanding transformations which can be modeled by some Young towers (see [24] and [25]) with sufficiently fast decaying tails, as well as some Markov processes with infinite state space.

The motivation for the present paper is twofold. On one hand, it comes from the series of papers deriving Poisson type asymptotics for distributions of numbers of returns to appropriately shrinking sets (see, for instance, [3], [2], [16] and references there). On the other hand, our motivation was influenced by works on open dynamical systems which study dynamics of such systems until they exit the phase space through a “hole” (see, for instance, [9], [14] and references there). In our setup the number of returns is studied until a “hazard” which is interpreted as a certain visit to a set which can be also viewed as a “hole”.

The structure of this paper is as follows. In the next section we will describe precisely our setups and formulate main results. In Section 3 we derive auxiliary lemmas and the corollary from the main theorem for the symbolic setup proving the latter in Section 4. In Sections 5, 6 and 7 we prove our results for geometric balls. In Section 8 we show that the geometric law is preserved under suspension, proving our results for some Young towers.
2. Preliminaries and main results

2.1. Symbolic setup. Our setup consists of a finite or countable set $\mathcal{A}$ which is not a singleton, the sequence space $\Omega = \mathcal{A}^{\mathbb{N}}$, the $\sigma$-algebra $\mathcal{F}$ on $\Omega$ generated by cylinder sets, the left shift $T : \Omega \to \Omega$, and a $T$-invariant probability measure $P$ on $(\Omega, \mathcal{F})$ which is assumed to be $\phi$-mixing with respect to the $\sigma$-algebras $\mathcal{F}_{mn}$, $n \geq m$ generated by the cylinder sets

$$\{\omega = (\omega_0, \omega_1, \ldots) \in \Omega : \omega_i = a_i \text{ for } m \leq i \leq n\}$$

for some $a_m, a_{m+1}, \ldots, a_n \in \mathcal{A}$. Observe also that $\mathcal{F}_{mn} = T^{-m} \mathcal{F}_{0,n-m}$ for $n \geq m$.

Recall that the $\phi$-dependence (mixing) coefficient between two $\sigma$-algebras $\mathcal{G}$ and $\mathcal{H}$ can be written in the form (see [7])

$$\phi(\mathcal{G}, \mathcal{H}) = \sup_{\Gamma \in \mathcal{G}, \Delta \in \mathcal{H}} \left\{ \left| \frac{P(\Gamma \cap \Delta)}{P(\Gamma)} - P(\Delta) \right|, P(\Gamma) \neq 0 \right\}$$

(2.1)

$$= \frac{1}{2} \sup \{ \| E(g|\mathcal{G}) - E(g) \|_{L^1} : g \text{ is } \mathcal{H}\text{-measurable and } \| g \|_{L^\infty} \leq 1 \}.$$ 

Set also

(2.2)

$$\phi(n) = \sup_{m \geq 0} \phi(\mathcal{F}_{0,m}, \mathcal{F}_{m+n,\infty}).$$

The probability $P$ is called $\phi$-mixing if $\phi(n) \to 0$ as $n \to \infty$.

We will need also the $\alpha$-dependence (mixing) coefficient between two $\sigma$-algebras $\mathcal{G}$ and $\mathcal{H}$ which can be written in the form (see [7]),

$$\alpha(\mathcal{G}, \mathcal{H}) = \sup_{\Gamma \in \mathcal{G}, \Delta \in \mathcal{H}} \left\{ |P(\Gamma \cap \Delta) - P(\Gamma)P(\Delta)| \right\}$$

$$= \frac{1}{4} \sup \{ \| E(g|\mathcal{G}) - E(g) \|_{L^1} : g \text{ is } \mathcal{H}\text{-measurable and } \| g \|_{L^\infty} \leq 1 \}.$$ 

Set also

$$\alpha(n) = \sup_{m \geq 0} \alpha(\mathcal{F}_{0,m}, \mathcal{F}_{m+n,\infty}).$$

For each word $a = (a_0, a_1, \ldots, a_{n-1}) \in \mathcal{A}^{n}$ we will use the notation

$$[a] = \{ \omega = (\omega_0, \omega_1, \ldots) : \omega_i = a_i, i = 0, 1, \ldots, n-1 \}$$

for the corresponding cylinder set. Without loss of generality we assume that the probability of each 1-cylinder set is positive, i.e., $P([a]) > 0$ for every $a \in \mathcal{A}$,
and since $\mathcal{A}$ is not a singleton we have also $\sup_{a \in \mathcal{A}} P([a]) < 1$. Write $\Omega_P$ for the support of $P$, i.e.,

$$\Omega_P = \{\omega \in \Omega : P[\omega_0, \ldots, \omega_n] > 0 \text{ for all } n \geq 0\}.$$ 

For $n \geq 1$ set $\mathcal{C}_n = \{[w] : w \in \mathcal{A}^n\}$. Then $\mathcal{F}_{0,n}$ consists of $\emptyset$ and all unions of disjoint elements from $\mathcal{C}_{n+1}$. Next, for any $U \in \mathcal{F}_{0,n-1}$, $U \neq \emptyset$ and $V \in \mathcal{F}_{0,m-1}$, $V \neq \emptyset$ define

$$\pi(U) = \min\{k \geq 1 : U \cap T^{-k}U \neq \emptyset\}$$

and

$$\pi(U, V) = \min\{k \geq 1 : U \cap T^{-k}V \neq \emptyset \text{ or } V \cap T^{-k}U \neq \emptyset\}.$$ 

It is clear that $\pi(U) \leq n$ and $\pi(U, V) \leq n$, and so

$$\kappa_{U,V} = \min\{\pi(U, V), \pi(U), \pi(V)\} \leq m \wedge n$$

where, as usual, for $n, m \geq 1$ we denote

$$m \vee n = \max\{m, n\} \quad \text{and} \quad m \wedge n = \min\{m, n\}.$$ 

Set

$$\tau_V(\omega) = \min\{k \geq 1 : T^k\omega \in V\}$$

with $\tau_V(\omega) = \infty$ if the event in braces does not occur and define

$$\Sigma_{U,V} = \sum_{k=0}^{\tau_V-1} I_U \circ T^k.$$ 

For any two random variables or random vectors $Y$ and $Z$ of the same dimension denote by $\mathcal{L}(Y)$ and $\mathcal{L}(Z)$ their distribution and by

$$d_{TV}(\mathcal{L}(Y), \mathcal{L}(Z)) = \sup_G |\mathcal{L}(Y)(G) - \mathcal{L}(Z)(G)|$$

the total variation distance between $\mathcal{L}(Y)$ and $\mathcal{L}(Z)$ where the supremum is taken over all Borel sets. We denote also by $\text{Geo}(\rho), \rho \in (0, 1)$ the geometric distribution with the parameter $\rho$, i.e.,

$$\text{Geo}(\rho)\{k\} = \rho(1 - \rho)^k \quad \text{for each } k \in \mathbb{N} = \{0, 1, \ldots\}.$$
2.1. Theorem: For all integers $n, m, M, R \geq 1$ and the sets $U \in \mathcal{F}_{0,n-1}$, $U \neq \emptyset$ and $V \in \mathcal{F}_{0,m-1}$, $V \neq \emptyset$ such that $U \cap V = \emptyset$ we have

\[
\begin{align*}
&d_{TV}\left(\mathcal{L}(\Sigma_{U,V}), \text{Geo}\left(\frac{P(V)}{P(V) + P(U)}\right)\right) \\
\leq & (1 - P(V))^{M+1} + P(U) + 6MR(P(U) + P(V))^2 + 6M\phi(R - n \lor m) \\
&+ 8MR(P(U) + P(V))(P(U_r) + P(V_r) + \phi(\kappa_{U,V} + r - n \lor m))
\end{align*}
\]

(2.3)

where

\[U_r = T^r U, \quad V_r = T^r V\]

and $r$ is an arbitrary integer such that $r \geq n \lor m - \kappa_{U,V}$.

Next, for each $\xi \in \Omega$ and $n \geq 1$ write $A^\xi_n = [\xi_0 \cdots \xi_{n-1}] \in \mathcal{C}_n$. For any $\xi, \eta \in \Omega$ set

\[\tau^\eta_n(\omega) = \tau_{A^\eta_n}(\omega) \quad \text{and} \quad \Sigma^\xi,\eta_{n,m} = \Sigma_{A^\xi_n, A^\eta_m}.\]

The following corollary deals with the limit behavior of $\Sigma^\xi,\eta_{n,m(\cdot)}$ for $P \times P$-typical pairs $(\xi, \eta) \in \Omega \times \Omega$, where $|m(n) - n| = o(n)$ and, as usual, $o(n)$ denotes an unspecified function $f : \mathbb{N} \to \mathbb{N}$ with $\frac{f(n)}{n} \to 0$ as $n \to \infty$.

2.2. Corollary: Let $\{m(n)\}_{n \geq 1} \subset \mathbb{N} \setminus \{0\}$ be a sequence satisfying

\[|m(n) - n| = o(n) \quad \text{as} \quad n \to \infty.\]

Assume that there exists $\beta > 0$ such that $\phi(n) \leq \beta^{-1} e^{-\beta n}$ for all $n \geq 1$. Impose also the finite entropy condition

\[-\sum_{a \in A} P([a]) \ln P([a]) < \infty.\]

Then for $P \times P$-a.e. $(\xi, \eta) \in \Omega \times \Omega$,

\[
\lim_{n \to \infty} d_{TV}\left(\mathcal{L}(\Sigma_{n,m(\cdot)}^{\xi,\eta}), \text{Geo}\left(\frac{P(A^\eta_{m(n)})}{P(A^\eta_{m(n)}) + P(A^\xi_n)}\right)\right) = 0.
\]

(2.4)

In particular, if

\[
\lim_{n \to \infty} \frac{P(A^\xi_n)}{P(A^\eta_{m(n)})} = \lambda
\]

(2.5)

then $\mathcal{L}(\Sigma_{n,m(\cdot)}^{\xi,\eta})$ converges in total variation as $n \to \infty$ to the geometric distribution with the parameter $(1 + \lambda)^{-1}$.
We observe that, in general (in fact, “usually”), the ratio $\frac{P(A^\xi_n)}{P(A^\eta_m)}$ will be unbounded for distinct $\xi, \eta \in \Omega$, and so in order to obtain a nontrivial limiting geometric distribution it is necessary to choose cylinders $A^\xi_n$ and $A^\eta_{m(n)}$ with appropriate relative lengths. In order to have the ratio $\frac{P(A^\xi_n)}{P(A^\eta_{m(n)})}$ bounded away from zero and infinity our condition $|m(n) - n| = o(n)$ is, essentially, necessary (at least, in the finite entropy case) which follows from the Shannon–McMillan–Breiman theorem (see [22]). Corollary 2.2 can be applied to Markov shifts with infinite state space satisfying the Doeblin condition which are known to be $\phi$-mixing with an exponential speed (see [7]).

2.3. Remark: It is not difficult to see that if $\xi, \eta \in \Omega$ are not periodic and not shifts of each other then $\kappa_{A^\xi_n, A^\eta_{m(n)}} \to \infty$ as $m, n \to \infty$. Still, some rate of growth of the latter is needed in order to use the estimate of Theorem 2.1 for an asymptotic result as in Corollary 2.2 if $|m(n) - n|$ is unbounded since to make the right hand side of (2.3) small we need $r = r(n)$ satisfying $n \wedge m(n) - r(n) \to \infty$ and $\kappa_{A^\xi_n, A^\eta_{m(n)}} + r(n) - n \vee m(n) \to \infty$ as $n \to \infty$.

2.4. Remark: Theorem 2.1 and Corollary 2.2 remain true also for the two-sided shift setup. In this case $\Omega = A^\mathbb{Z}$, i.e.,

$$\Omega = \{\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots) : \omega_i \in \mathcal{A}\},$$

and the $\sigma$-algebras $\mathcal{F}_{m,n}$ are defined for $-\infty < m \leq n < \infty$ as unions of cylinder sets $\{\omega = (\ldots, \omega_{-1}, \omega_0, \omega_1, \ldots) : \omega_i = a_i \text{ for } m \leq i \leq n\}$. The $\phi$ and $\alpha$ dependence coefficients are defined by the same formulas as above with

$$\phi(n) = \sup_{-\infty < m < \infty} \phi(\mathcal{F}_{-\infty, m}, \mathcal{F}_{m+n, \infty})$$

and

$$\alpha(n) = \sup_{-\infty < m < \infty} \alpha(\mathcal{F}_{-\infty, m}, \mathcal{F}_{m+n, \infty}).$$

Next, for $n, m \geq 1$ we consider nonempty sets $U \in \mathcal{F}_{-n+1, n-1}$ and $V \in \mathcal{F}_{-m+1, m-1}$ and define $\pi(U)$, $\pi(U, V)$, $\tau_V$, $\kappa_{UV}$ and $\Sigma_{U, V}$ in the same way as above. In the two sided shift case $U_r$ and $V_r$ should be defined differently by considering $U$ and $V$ as unions of cylinders and for each cylinder $A = (a_{-k}, a_{-k+1}, \ldots, a_m)$ considering the cylinder $A = (a_{-k+r}, a_{-k+r+1}, \ldots, a_m)$ and obtaining $U_r$ and $V_r$ as unions of such modified cylinders. The quantities we will have to estimate
in the proof of Theorem 2.1 and Corollary 2.2 have the form
\[ P(U \cap T^{-k}V) = P(T^{-l}U \cap T^{-l-k}V) \]
where \( U \in \mathcal{F}_{-n+1,n-1} \) and \( V \in \mathcal{F}_{-m+1,m-1} \). Choosing \( l \geq \max(m - 1, n - 1) \) we obtain that \( T^{-l}U \in \mathcal{F}_{l-n+1,l+n-1} \) and \( T^{-l-k}V \in \mathcal{F}_{l+k-m+1,l+k+m-1} \) with \( l - n + 1 \geq 0 \) and \( l - m + 1 \geq 0 \) which amounts to the same estimates as in the one-sided shift case.

2.2. Maps with \( \phi \)-mixing partitions. Let \( T \) be a measurable map of a compact metric space \( M \) and \( \mu \) be a \( T \)-invariant probability measure on \( M \). Our setup includes also a countable (one-sided) measurable generating partition \( \mathcal{A} \) of \( M \) with finite entropy and denote by
\[ \mathcal{A}^n = \bigvee_{j=0}^{n-1} T^{-j} \mathcal{A} \]
its \( n \)th join. Recall, that “generating” means that if \( A_n(x) \) is an element of the partition \( \mathcal{A}^n \) which contains a point \( x \), then
\[ \bigcap_{n} A_n(x) = \{x\} \]

Let \( A^j, j = 1, 2, \ldots \) be a numeration of elements of \( \mathcal{A} \); then each \( x \in M \) has a symbolic representation \( \omega(x) = (\omega_0, \omega_1, \ldots) \) so that \( \omega_k = j \) if \( T^k x \in A^j \). Since \( \mathcal{A} \) is generating, then no two different points have the same symbolic representation. Hence, the map \( \omega : M \to \mathcal{A}^\mathbb{N} \) is an injection and it sends the measure \( \mu \) to a probability measure \( \omega(\mu) \) on \( \mathcal{A}^\mathbb{N} \) invariant under the left shift on \( \mathcal{A}^\mathbb{N} \) which provides a symbolic representation of the dynamical system \((M, T, \mu)\). Let \( \mathcal{F}_{0,n-1} \) be the \( \sigma \)-algebra generated by all elements of the partition \( \mathcal{A}^n \). Clearly, \( \mathcal{F}_{0,n-1} \) consists of \( \emptyset \) and all unions of elements of \( \mathcal{A}^n \). We denote also by \( \mathcal{F} \) the minimal \( \sigma \)-algebra which contains all \( \mathcal{F}_{0,n-1}, n \geq 1 \). Recall that the measure \( \mu \) is called (left) \( \phi \)-mixing if
\[ |\mu(\Gamma \cap T^{-n-k}\Delta) - \mu(\Gamma)\mu(\Delta)| \leq \phi(k)\mu(\Gamma) \]
for any \( \Gamma \in \mathcal{F}_{0,n-1} \) and \( \Delta \in \mathcal{F} \), where \( \phi(k) \) is nonincreasing and \( \phi(k) \to 0 \) as \( k \to \infty \). This definition corresponds to the one given above in the symbolic setup if we introduce \( \sigma \)-algebras \( \mathcal{F}_{mn} = T^{-m}\mathcal{F}_{0,n-m+1}, n \geq m \). We will assume the following properties. Denote by \( B_r(x) \) an open ball of radius \( r \) centered at \( x \), though in view of Assumption A4 below our results remain the same whether we consider open or closed balls.
A1. The diameter of $\mathcal{A}^n$. There is a generating partition $\mathcal{A}$ and constants $C, p > 0$, such that for every $n \geq 0$, the diameter of its $n$th join under the map $T$, $\mathcal{A}^n$, satisfies $\text{diam} \mathcal{A}^n \leq C n^{-p}$.

When the diameter of $\mathcal{A}^n$ decays super-polynomially, we say that A1 holds with $p = +\infty$.

A2. Polynomial rate of $\phi$-mixing. The measure $\mu$ is left $\phi$-mixing with respect to the partition $\mathcal{A}$, with $\phi(n) \leq C n^{-\beta}$ for some $\beta > 1$.

A3. Dimension for $\mu$. There is $d > 0$, such that for almost every $x \in \text{supp}(\mu)$, we have

$$\lim_{r \to 0} \frac{\log \mu(B_r(x))}{\log r} = d.$$ 

It is shown in [20] that when this assumption holds, then the Hausdorff dimension $\text{dim}_H \mu$ of $\mu$ equals $d$.

A4. Regularity of $\mu$. There are $C, a > 0$ and $b \in \mathbb{R}$, such that for $\delta \ll r$ and almost every $x$,

$$\mu(B_{r+\delta}(x) \setminus B_{r-\delta}(x)) \leq C \delta^a r^{-b} \mu(B_r(x)).$$

Note that for the Lebesgue measure on $\mathbb{R}^n$, this property is satisfied with $a = b = 1$.

For any $x, y, z \in \mathbf{M}$, write

$$\Sigma^x_y(z) = \sum_{j=0}^{\tau_{B_r(y)}-1} \mathbb{I}_{B_r(x)}(T^j z),$$

which counts the number of arrivals to $B_r(x)$ before hitting $B_r(y)$ for the first time. With these we can state the theorem.

To simplify notation, we will write

$$\rho(x, y) = \frac{\mu(B_r(y))}{\mu(B_r(x)) + \mu(B_r(y))}.$$
2.5. Theorem: Assume that Assumptions A1–A4 are satisfied with
\[(2.6) \quad p > \frac{d + b}{ad}.\]
Then for $\mu \times \mu$-almost every $(x, y) \in \mathcal{M} \times \mathcal{M}$, such that $\rho(x, y, r) \to \rho(x, y) \in (0, 1)$ as $r \to 0$,
\[
\lim_{r \to 0} d_{TV}(\mathcal{L}(\Sigma^{x,y}_r), \text{Geo}(\rho(x, y))) = 0.
\]
If the diameter of $\mathcal{A}^n$ decreases super-polynomially, then the assumption (2.6) is redundant.

2.6. Remark: When the measure $\mu$ is absolutely continuous with respect to the volume with a density $h$ that is bounded from above, we can take $a = b = 1$ in A4. In this case, condition (2.6) reduces to $p > \frac{d+1}{d}$.

2.7. Remark (Radius of the ball): In Theorem 2.5 we take the ball at $x$ and $y$ with the same radius. However, one can easily check that the same results hold as long as
\[
\frac{\mu(B_{r_y}(y))}{\mu(B_{r_x}(x)) + \mu(B_{r_y}(y))}
\]
converges to a limit $\rho \in (0, 1)$ when $r_x$ and $r_y$ tend to 0. For example, when the measure $\mu$ is absolutely continuous with respect to the Lebesgue measure, one can take the balls to be $B_r(x)$ and $B_{cr}(y)$ for any constant $c > 0$.

2.8. Remark (Invertible case): Theorem 2.5 remains true when $T$ is invertible. In this case, assumption A1 should be stated for the two-sided join
\[
\mathcal{A}^n = \bigvee_{j=-(n-1)}^{n-1} T^{-j} \mathcal{A}.
\]
The rest of the proof remains the same, with minor modification described in Remark 2.4. Also note that Theorem 6.1 and Propositions 6.5, 6.6 hold true for both invertible and non-invertible systems.

2.3. Geometric Law under Suspension. Next, we will state a general result which will allow us to generalize the previous theorems to discrete time suspensions over $\phi$-mixing systems. Namely, let $(\hat{\Omega}, \hat{\mu}, \hat{T})$ be a measure preserving dynamical system with $\hat{\mu}$ being a probability measure. Given a measurable function $R: \hat{\Omega} \to \mathbb{Z}^+$ consider the space $\Omega = \hat{\Omega} \times \mathbb{Z}^+ / \sim$ with the equivalence relation $\sim$ given by
\[
(x, R(x)) \sim (\hat{T}(x), 0).
\]
Define the (discrete-time) suspension map over $\hat{\Omega}$ with roof function $R$ as the measurable map $T$ on the space $\Omega$ acting by

$$T(x, j) = \begin{cases} 
(x, j + 1) & \text{if } j < R(x) - 1, \\
(\hat{T}x, 0) & \text{if } j = R(x) - 1.
\end{cases}$$

We will call $\Omega$ a tower over $\hat{\Omega}$ and refer to the set $\Omega_k := \{(x, k) : x \in \hat{\Omega}, k < R(x)\}$ as the $k$th floor, where $\hat{\Omega}$ can be naturally identified with the 0th floor, called the base of the tower.

For $0 \leq k < i$, set $\Omega_{k,i} = \{(x, k) : R(x) = i\}$. The map $\Pi : (x, k) \mapsto x$ is naturally viewed as a projection from the tower $\Omega$ to the base $\hat{\Omega}$ and for any given set $U \subset \Omega$ we will write $\hat{U} = \Pi(U)$.

The measure $\hat{\mu}$ can be lifted to a measure $\mu$ on $\Omega$ by

$$\hat{\mu}(A) = \sum_{i=1}^{\infty} \sum_{k=0}^{i-1} \hat{\mu}(\Pi(A \cap \Omega_{k,i})) = (\hat{\mu} \times \mathcal{N})(A),$$

where $\mathcal{N}$ is the counting measure on $\mathbb{N}$. It is easy to verify that $\hat{\mu}$ is $T$-invariant and if $\hat{\mu}(R) = \int R \, d\hat{\mu} < \infty$ then $\hat{\mu}$ is a finite measure. In this case, the measure

$$\mu = \frac{\hat{\mu}}{\hat{\mu}(R)} = \frac{\hat{\mu}}{\hat{\mu}(R)} \times \mathcal{N}$$

is a $T$-invariant probability measure on $\Omega$. In order to state our main theorem for this section, we will introduce the following class of sets.

2.9. Definition: We say that a positive measure set $U \subset \Omega$ is well-placed (WP) if for $\hat{\mu}$-almost every $x$ the intersection $U \cap \{(x, k) : k < R(x)\}$ contains at most one point.

2.10. Remark: If $U$ is well-placed, then, clearly, $\Pi |_{U}$ is almost injective, i.e., there exists a set $U_0 \subset U$ with $\mu(U \setminus U_0) = 0$, such that $\Pi |_{U_0}$ is injective. In particular, if $U \subset \Omega_0$ then it is well-placed. Also note that if $U$ is well-placed, then so is every positive measure subset of $U$. 
As before, denote by $\tau_U$ the first hitting time of $U \subset \Omega$, i.e.,
$$\tau_U(x) = \min\{l \geq 1 : T^lx \in U\}$$
if the event in braces occurs and $\tau_U(x) = \infty$, if not, and for $U, V \subset \Omega$ we also write
$$\Sigma_{U,V} = \sum_{k=0}^{\tau_V-1} I_U \circ T^k.$$  
Similarly, given $\tilde{U}, \tilde{V} \subset \tilde{\Omega}$, we denote by $\tilde{\tau}_{\tilde{U}}$ the first hitting time of $\tilde{U}$ under iterates of the map $\tilde{T}$ and set
$$\tilde{\Sigma}_{\tilde{U},\tilde{V}} = \sum_{k=0}^{\tilde{\tau}_{\tilde{V}}-1} I_{\tilde{U}} \circ \tilde{T}^k.$$  
In other words, every term that contains a tilde is defined for the base systems $(\tilde{\Omega}, \tilde{\mu}, \tilde{T})$, and every term without a tilde is defined for the suspension $(\Omega, \mu, T)$.

2.11. Theorem: Let $(\Omega, \mu, T)$ be a discrete-time suspension over an ergodic system $(\tilde{\Omega}, \tilde{\mu}, \tilde{T})$ with a roof function $R$ satisfying
$$\int R d \tilde{\mu} < \infty.$$  
Let $\{U_n\}, \{V_n\}$ be two sequences of well-placed subsets of $\Omega$. We consider the following two cases:

(i) If the base system $(\tilde{\Omega}, \tilde{\mu}, \tilde{T})$ and the sets $\tilde{U}_n = A^\omega_{A_n}$, $\tilde{V}_n = A^\eta_{m(n)}$ fall within the symbolic setup of Section 2.1 and satisfy the assumptions of Corollary 2.2, then $\Sigma_{U_n,V_n}$ converges in distribution on $(\Omega, \mu)$ to the geometric distribution $\text{Geo}((1 + \lambda)^{-1})$ with $\lambda$ given by (2.5).

(ii) If the base system $(\tilde{\Omega}, \tilde{\mu}, \tilde{T})$ and the sets $\tilde{U}_n = B_{r_n}(x)$, $\tilde{V}_n = B_{r_n}(y)$ fall within the setup of Section 2.2 and satisfy the assumptions of Theorem 2.5 for some sequence of positive real numbers $\{r_n\}$ with $r_n \to 0$, then $\Sigma_{U_n,V_n}$ converges in distribution on $(\Omega, \mu)$ to the geometric distribution $\text{Geo}(\rho(x,y))$.

3. Some auxiliary lemmas and Corollary 2.2

We start with the following result which appears in [11] and in [1] as Lemma 1 but under the extra condition that $A$ is finite, which is redundant as the following proof shows.
3.1. Lemma: Suppose that $P$ is $\phi$-mixing. Then there exists a constant $\nu > 0$ such that for any $A \in \mathcal{C}_n$, $$P(A) \leq e^{-\nu n}.$$ 

Proof. Since $\gamma = \sup_{a \in A} P([a]) < 1$ and $P$ is $\phi$-mixing, i.e., $\phi(n) \downarrow 0$ as $n \uparrow \infty$, we can set $k = \min \{j : \gamma + \phi(j) < 1\}$. Let $A = [a_0, a_1, \ldots, a_{n-1}]$. Then

$$A \subset \bigcap_{i=0}^{[n/k]} T^{-ik}[a_{ik}].$$

By the definition of the $\phi$-dependence coefficient,

$$P(A) \leq P\left(\bigcap_{i=0}^{[n/k]} T^{-ik}[a_{ik}]\right) \leq (P([a_{k[n/k]}]) + \phi(k))P\left(\bigcap_{i=0}^{[n/k]-1} T^{-ik}[a_{ik}]\right)$$

$$\leq \cdots \leq (\gamma + \phi(k))^{n/k}\gamma \leq e^{-\nu n}$$

where $\nu = -k^{-1} \ln(\gamma + \phi(k))$ and, without causing any confusion, we use $[\cdot]$ both for the integral part of a number and to denote 1-cylinders $[a], a \in A$. 

Next, we prove Corollary 2.2 assuming that Theorem 2.1 is already proved but, first, we will need the following lemma. In what follows, $\{m(n)\}_{n \geq 1}$ is a sequence of positive integers with $m(n) \geq 1$ and $|m(n) - n| = o(n)$ as $n \to \infty$. For $n \geq 1$ we write $\kappa^{\omega, \eta}_{n,m} = \kappa_{A^\omega_n, A^\eta_m}$ and $b(n) = n \wedge m(n)$.

3.2. Lemma: Set $c = 3\nu^{-1}$ and let $\mathcal{E}$ be the set of all $(\omega, \eta) \in \Omega \times \Omega$ for which there exists $N = N(\omega, \eta) \geq 1$ such that $\kappa^{\omega, \eta}_{n,m(n)} \geq b(n) - c \ln b(n)$ for all $n \geq N$. Then $P \times P(\Omega^2 \setminus \mathcal{E}) = 0$.

Proof. For $\omega \in \Omega$ and $n \geq 1$ set

$$B_{\omega,n} = \{\eta \in \Omega : \pi(A^\omega_n, A^\eta_{m(n)}) \leq b(n) - c \ln b(n)\}.$$ 

Assume $b(n) - c \ln b(n) \geq 1$ and set $d = d(n) = [b(n) - c \ln b(n)]$. Then

$$P(B_{\omega,n}) \leq \sum_{r=0}^{d} P\{\eta : T^{-r}A^\omega_n \cap A^\eta_{m(n)} \neq \emptyset\} + \sum_{r=0}^{d} P\{\eta : T^{-r}A^\eta_{m(n)} \cap A^\omega_n \neq \emptyset\}.$$ 

For $0 \leq r \leq d$,

$$\{\eta : T^{-r}A^\omega_n \cap A^\eta_{m(n)} \neq \emptyset\} \subset T^{-r}[\omega_0, \ldots, \omega_{n \wedge (m(n)-r)-1}]$$

and

$$\{\eta : T^{-r}A^\eta_{m(n)} \cap A^\omega_n \neq \emptyset\} \subset [\omega_r, \ldots, \omega_{n \wedge (m(n)+r)-1}].$$
Hence by Lemma 3.1 for all $n$ large enough,

$$P(B_{\omega,n}) \leq \sum_{r=0}^{d} e^{-\nu(n\wedge(m(n)-r))} + \sum_{r=0}^{d} e^{-\nu((n-r)\wedge m(n))} \leq 2 \sum_{r=0}^{d} e^{-\nu(b(n)-r)} \leq 2 \frac{e^{-\nu(b(n)-d)}}{1-e^{-\nu}} \leq \frac{2b(n)^{-3}}{1-e^{-\nu}}.$$ 

From this and since $|b(n) - n| = o(n)$ it follows that $\sum_{n=1}^{\infty} P(B_{\omega,n}) < \infty$, and so by the Borel–Cantelli lemma

$$P\{\eta : \#\{n \geq 1 : \eta \in B_{\omega,n}\} = \infty\} = 0.$$

From Fubini’s theorem we now get

$$P \times P\{(\omega, \eta) : \#\{n \geq 1 : \pi(A_{\omega n}^{\eta}, A_{m(n)}^{\eta}) \leq b(n) - c\ln b(n)\} = \infty\} = \int_{\Omega} P\{\eta : \#\{n \geq 1 : \eta \in B_{\omega,n}\} = \infty\} dP(\omega) = 0.$$

In a similar manner it can be shown that

$$P\{\omega : \#\{n \geq 1 : \pi(A_{\omega n}^{\eta}) \leq b(n) - c\ln b(n)\} = \infty\} = 0$$

and

$$P\{\eta : \#\{n \geq 1 : \pi(A_{m(n)}^{\eta}) \leq b(n) - c\ln b(n)\} = \infty\} = 0.$$ 

This completes the proof of the lemma. \qed

**Proof of Corollary 2.2.** Let $\kappa_{n,m(n)}^\omega \geq b(n) - c\ln b(n)$ for all $n \geq N$ where $N = N(\omega, \eta) \geq 1$.

Let $c$ and $E$ be as in the statement of Lemma 3.2. Denote by $h$ the entropy of the system $(\Omega, P, T)$ which is finite under our assumptions. Let $E_0$ be the set of all $(\omega, \eta) \in E \cap (\Omega_1 \times \Omega_1)$ for which

$$-\lim_{n \to \infty} \frac{\log P(A_{\omega n}^{\eta})}{n} = -\lim_{n \to \infty} \frac{\log P(A_{m(n)}^{\eta})}{n} = h.$$ 

Recall (see, for instance, [7]) that our $\phi$-mixing (and even $\alpha$-mixing) assumption implies that the shift $T$ is mixing in the ergodic theory sense with respect to the invariant measure $P$, and so it is ergodic. Hence, we can apply the Shannon–McMillan–Breiman Theorem (see, for instance, [22]) which implies that the above equalities hold true with probability one and $h \geq \nu > 0$ by Lemma 3.1. This together with Lemma 3.2 yields that

$$P \times P(\Omega^2 \setminus E_0) = 0.$$
Now, let \((\omega, \eta) \in \mathcal{E}_0\). Taking in Theorem 2.1, \(M = M(n) = [e^{(h+\varepsilon)n}]\) with small enough \(\varepsilon > 0\), \(R = R(n) = n^2\) and \(r = [n/2]\) we obtain (2.4) from (2.3) while assuming (2.5) the second assertion follows directly from (2.4). \(\blacksquare\)

For the proof of Theorem 2.1 we will need the following general result.

3.3. Lemma: Let \(\mathcal{F}_1, \mathcal{F}_2\) and \(\mathcal{F}_3\) be sub \(\sigma\)-algebras of \(\mathcal{F}\) such that

\[(3.1) \quad \max(\phi(\mathcal{F}_1, \sigma(\mathcal{F}_2, \mathcal{F}_3)), \phi(\mathcal{F}_2, \mathcal{F}_3)) = \delta\]

where \(\sigma(\mathcal{G}, \tilde{\mathcal{G}})\) denotes the minimal \(\sigma\)-algebra containing \(\mathcal{G}\) and \(\tilde{\mathcal{G}}\). Then

\[(3.2) \quad \alpha(\mathcal{F}_2, \sigma(\mathcal{F}_1, \mathcal{F}_3)) \leq 3\delta.\]

Proof. Let \(U_1 \in \mathcal{F}_1, U_2 \in \mathcal{F}_3\) and \(V \in \mathcal{F}_2\). Then

\[
\begin{align*}
|P(U_1 \cap V \cap U_2) - P(U_1)P(V \cap U_2)| &\leq \delta P(U_1), \\
P(U_1)|P(V \cap U_2) - P(V)P(U_2)| &\leq \delta P(U_1)P(V), \\
\text{and } P(V)|P(U_1)P(U_2) - P(U_1 \cap U_2)| &\leq \delta P(U_1)P(V).
\end{align*}
\]

Hence

\[(3.3) \quad |P(U_1 \cap V \cap U_2) - P(U_1 \cap U_2)P(V)| \leq \delta P(U_1)(1 + 2P(V)) \leq 3\delta P(U_1).\]

Next, consider the collection \(\mathcal{U}\) of all sets of the form \(U = \bigcup_{i=1}^{l}(U_1^{(i)} \cap U_2^{(i)})\) where \(U_1^{(i)} \in \mathcal{F}_1, i = 1, \ldots, l\) are disjoint while \(U_2^{(i)} \in \mathcal{F}_3, i = 1, \ldots, l\) are arbitrary. Clearly, finite unions and intersections of sets from \(\mathcal{U}\) belong to \(\mathcal{U}\), and so \(\mathcal{U}\) is an algebra of sets as \(\Omega\) and \(\emptyset\) belong to \(\mathcal{U}\), as well. Now, if \(U = \bigcup_{i=1}^{l}(U_1^{(i)} \cap U_2^{(i)})\) and \(V\) are as above then, by (3.3),

\[
\begin{align*}
|P(U \cap V) - P(U)P(V)| &\leq \sum_{i=1}^{l} \left|P(U_1^{(i)} \cap V \cap U_2^{(i)}) - P(U_1^{(i)} \cap U_2^{(i)})\right| \\
&\leq \sum_{i=1}^{l} \left|P(U_1^{(i)} \cap V \cap U_2^{(i)}) - P(V)P(U_1^{(i)} \cap U_2^{(i)})\right| \\
&\leq 3\delta P\left(\bigcup_{i=1}^{l} U_1^{(i)}\right) \leq 3\delta.
\end{align*}
\]

The estimate (3.4) being true for all \(U_j \in \mathcal{U}\) remains valid under monotone limits \(U_j \uparrow\) and \(U_j \downarrow\), and so it holds true for any \(U \in \sigma(\mathcal{F}_1, \mathcal{F}_3)\) and \(V \in \mathcal{F}_2\), yielding (3.2). \(\blacksquare\)
We will need the following result which is, essentially, an exercise in elementary probability whose proof can be found in [17].

3.4. LEMMA: Let $Y = \{Y_{k,l} : k \geq 0 \text{ and } l \in \{0,1\}\}$ be independent Bernoulli random variables such that

$$1 > P\{Y_{k,0} = 1\} = p = 1 - P\{Y_{k,0} = 0\} > 0$$
and

$$1 > P\{Y_{k,1} = 1\} = q = 1 - P\{Y_{k,1} = 0\} > 0.$$ 

Set

$$\tau = \min\{l \geq 0 : Y_{l,0} = 1\}.$$ 

Then

$$S = \sum_{l=0}^{\tau-1} Y_{l,1}$$

is a geometric random variable with the parameter $p(p+q-pq)^{-1}$.

4. Proof of Theorem 2.1

Define random variables $X_{k,0}$ and $X_{k,1}$ on $(\Omega, \mathcal{F}, P)$ by

$$X_{k,0} = I_V \circ T^k \quad \text{and} \quad X_{k,1} = I_U \circ T^k,$$

and set

$$S_M = \sum_{k=0}^{M-1} X_{k,1}, \quad \tau = \min\{k \geq 0 : X_{k,0} = 1\} \quad \text{and} \quad \tau_M = \min(\tau, M).$$

Then $S_\tau = \Sigma_{U,V}$ which appears in Theorem 2.1. Let $\{Y_{k,\alpha} : k \geq 0, \alpha = 0,1\}$ be a sequence of independent Bernoulli random variables such that $Y_{k,\alpha}$ has the same distribution as $X_{k,\alpha}$. Set

$$S_M^* = \sum_{k=0}^{M-1} Y_{k,1}, \quad \tau^* = \min\{k \geq 0 : Y_{k,0} = 1\} \quad \text{and} \quad \tau_M^* = \min(\tau^*, M).$$

Denote

$$p_V = P\{X_{k,0} = 1\} = P\{Y_{k,0} = 1\} \quad \text{and} \quad p_U = P\{X_{k,1} = 1\} = P\{Y_{k,1} = 1\},$$

$$k = 0,1,\ldots$$
Observe that $S^*_\tau$ has by Lemma 3.4 the geometric distribution with the parameter
\[ \varrho = \frac{pV}{pV + pU(1-pV)} > \rho = \frac{pV}{pV + pU}. \]
Next, we can write
\[ (4.1) \quad d_{TV}(\mathcal{L}(S_\tau), \text{Geo}(\rho)) \leq A_1 + A_2 + A_3 + A_4 \]
where
\[ A_1 = d_{TV}(\mathcal{L}(S_\tau), \mathcal{L}(S_{\tau_M})), \quad A_2 = d_{TV}(\mathcal{L}(S_{\tau_M}), \mathcal{L}(S^*_\tau)), \]
\[ A_3 = d_{TV}(\mathcal{L}(S^*_\tau), \mathcal{L}(S^*_\tau)) \quad \text{and} \quad A_4 = d_{TV}(\text{Geo}(\varrho), \text{Geo}(\rho)). \]
Introduce random vectors
\[ X_{M,\alpha} = \{X_{k,\alpha}, 0 \leq k \leq M\}, \quad \alpha = 0, 1, \]
\[ X_M = \{X_{M,0}, X_{M,1}\}, \]
\[ Y_{M,\alpha} = \{Y_{k,\alpha}, 0 \leq k \leq M\}, \quad \alpha = 0, 1 \]
and
\[ Y_M = \{Y_{M,0}, Y_{M,1}\}. \]
Observe that the events \{\(S_\tau \neq S_{\tau_M}\) or \(S^*_\tau \neq S^*_\tau_M\)\} can occur only if \{\(\tau > M\)\} or \{\(\tau^* > M\)\}, respectively. Also, we can write
\[ \{\tau > M\} = \{X_{k,0} = 0 \text{ for all } k = 0, 1, \ldots, M\} \]
and
\[ \{\tau^* > M\} = \{Y_{k,0} = 0 \text{ for all } n = 0, 1, \ldots, M\}. \]
Hence,
\[ A_1 \leq P\{\tau > M\} \]
\[ \leq P\{\tau^* > M\} + P\{X_{k,0} = 0 \text{ for } k = 0, 1, \ldots, M\} \]
\[ - P\{Y_{k,0} = 0 \text{ for } k = 0, 1, \ldots, M\} \]
\[ \leq P\{\tau^* > M\} + d_{TV}(\mathcal{L}(X_{M,0}), \mathcal{L}(Y_{M,0})) \]
\[ \leq (1-pV)^{M+1} + d_{TV}(\mathcal{L}(X_M), \mathcal{L}(Y_M)). \]
Also,

\[(4.3)\quad A_3 \leq P\{\tau^* > M\} = (1 - p_V)^{M+1}.\]

The estimate of \(A_4\) is also easy:

\[(4.4)\quad A_4 \leq \sum_{k=0}^{\infty} |\varrho(1 - \varrho)^k - \rho(1 - \rho)^k|\]

\[\leq 2 \sum_{k=1}^{\infty} ((1 - \rho)^k - (1 - \varrho)^k)\]

\[= 2(1 - \rho)\rho^{-1} - 2(1 - \varrho)\varrho^{-1} = \frac{2(\rho - \varrho)}{\rho \varrho} = 2p_U.\]

Next, clearly,

\[(4.5)\quad A_2 \leq d_{TV}(\mathcal{L}(X_M), \mathcal{L}(Y_M))\]

and it remains to estimate the right hand side of (4.5). By Theorem 3 in [4],

\[(4.6)\quad d_{TV}(\mathcal{L}(X_M), \mathcal{L}(Y_M)) \leq 2b_1 + 2b_2 + b_3 + 2 \sum_{0 \leq n \leq M, \alpha = 0, 1} q_{n, \alpha}^2\]

(the additional factor 2 in [4] is due to a different definition of \(d_{TV}\)), where \(q_{n,0} = p_V\) and \(q_{n,1} = p_U\). In order to define \(b_1, b_2\) and \(b_3\) set

\[B_{k,\alpha}^M = \{(l, \alpha), (l, 1 - \alpha) : 0 \leq l \leq M, |l - k| \leq R\}\]

and

\[I_M = \{(k, \alpha) : 0 \leq k \leq M, \alpha = 0, 1\}.\]

Then

\[(4.7)\quad b_1 = \sum_{(k, \alpha) \in I_M} \sum_{(l, \beta) \in B_{k,\alpha}^M} q_{k,\alpha} q_{l,\beta},\]

\[b_2 = \sum_{(k, \alpha) \in I_M} \sum_{(k, \beta) \neq (l, \beta) \in B_{k,\alpha}^M} q_{(k,\alpha), (l,\beta)},\]

where \(q_{(k,\alpha), (l,\beta)} = E(X_{k,\alpha}X_{l,\beta})\), and

\[b_3 = \sum_{(k, \alpha) \in I_M} s_{k,\alpha},\]

where

\[s_{k,\alpha} = E|E(X_{k,\alpha} - q_{k,\alpha})\sigma\{X_{l,\beta} : (l, \beta) \in I_M \setminus B_{k,\alpha}^M\}||.\]
Clearly,
\begin{equation}
(4.8) \quad b_1 \leq 2MR(p_V^2 + 2p_V p_U + p_U^2) = 2MR(p_V + p_U)^2.
\end{equation}
In order to estimate $b_2$ consider two nonempty sets $D \in \mathcal{F}_{0,n-1}$ and $E \in \mathcal{F}_{0,m-1}$. Then both sets are finite or countable unions of corresponding cylinder sets
\[
D = \bigcup_i [d_0^{(i)}, d_1^{(i)}, \ldots, d_{n-1}^{(i)}] \quad \text{and} \quad E = \bigcup_j [e_0^{(j)}, e_1^{(j)}, \ldots, e_{m-1}^{(j)}].
\]
Assume that $D \cap T^{-k}E \neq \emptyset$ and suppose that $k + r \geq n$ for some $r \geq 0$. Set $E_r = T^r E$; then $T^{-r}E_r \supset E$, and so
\[
D \cap T^{-k}E \subset D \cap T^{-(k+r)}E_r.
\]
Since $D \in \mathcal{F}_{0,n-1}$ and $T^{-(k+r)}E_r \in \mathcal{F}_{k+r,\infty}$ we obtain by the definition of the $\phi$-dependence coefficient that
\begin{equation}
(4.9) \quad P(D \cap T^{-k}E) \leq P(D \cap T^{-(k+r)}E_r) \leq P(D)(P(E_r) + \phi(k + r - n + 1)).
\end{equation}
If $k \geq l$ and $k - l < \pi(U)$ then $X_{k,1}X_{l,1} = 0$, if $k \geq l$ and $k - l < \pi(V)$ then $X_{k,0}X_{l,0} = 0$, if $k \geq l$ and $k - l < \pi(U,V)$ then $X_{k,1}X_{l,0} = 0$ and, similarly, if $k \geq l$ and $k - l < \pi(V,U)$ then $X_{k,0}X_{l,1} = 0$. Thus, if $X_{k,\alpha}X_{l,\beta} \neq 0$ and $k \geq l$ then we must have $k - l \geq \kappa_{U,V}$. Hence, by (4.9), if $k - l \geq \pi(U)$ and $\alpha = \beta = 1$ then
\[
EX_{k,\alpha}X_{l,\beta} = P(U \cap T^{-(k-l)}U) \leq p_U(P(U_r) + \phi(\pi(U) + r - n + 1))
\]
where $U_r = T^r U$ and $r \geq n - \pi(U)$. If $k - l \geq \pi(V)$ and $\alpha = \beta = 0$ then
\[
EX_{k,\alpha}X_{l,\beta} \leq p_V(P(V_r) + \phi(\pi(V) + r - n + 1))
\]
where $V_r = T^r V$ and $r \geq m - \pi(V)$. If $k - l \geq \pi(U,V)$ and $\alpha = 0, \beta = 1$ then
\[
EX_{k,\alpha}X_{l,\beta} = P(V \cap T^{-(k-l)}U) \leq p_V(P(U_r) + \phi(\pi(U,V) + r - m + 1))
\]
for any $r \geq n - \pi(U,V)$. Finally, if $k - l \geq \pi(U,V)$ and $\alpha = 1, \beta = 0$ then
\[
EX_{k,\alpha}X_{l,\beta} \leq p_U(P(V_r) + \phi(\pi(U,V) + r - n + 1))
\]
prominence $r \geq n \land m - \pi(U,V)$. It follows that
\begin{equation}
(4.10) \quad b_2 \leq 4MR(p_U + p_V)(P(U_r) + P(V_r) + \phi(\kappa_{U,V} + r - m \lor n))
\end{equation}
for any $r$ satisfying $r \geq m \lor n - \kappa_{U,V}$.

In order to estimate $b_3$ we will use Lemma 3.3 which gives
\[
s_k,\gamma \leq \alpha(\mathcal{F}_{k,k+n}, \sigma(\mathcal{F}_{0,k-R+n\lor m}, \mathcal{F}_{k+R-n\lor m, \infty})) \leq 3\phi(R - n \lor m), \quad \gamma = 0, 1.
\]
It follows that

\[(4.11) \quad b_3 \leq 6M\phi(R - n \lor m).\]

Finally, from (4.1)–(4.8), (4.10) and (4.11) we obtain that

\[
d_{TV} (\mathcal{L}(S_\tau), \text{Geo}(\rho)) \leq 2(1 - p_V)^{M+1} + 2p_U + 8MR(p_V + p_U)^2 \\
+ 16MR(p_U + p_V)\phi(\kappa_{U,V} - |m - n|) \\
+ 12M\phi(R - n \lor m) + 4M(p_U^2 + p_V^2)
\]

and (2.3) follows. \[\square\]

4.1. Remark: Unlike [17] where \(\psi\)-mixing setups were considered, here under only \(\phi\)-mixing we cannot extend the result either to the stationary processes case or to a nonconventional setup. For the former the problem arises in estimates for \(b_2\) (close correlations), since above, close correlations \(EX_{k,\alpha}X_{l,\beta}\) with \(|k - l| < \pi(U,V)\) are zero while in the stationary process case we do not have sufficient control of close correlations under just \(\phi\)-mixing. In a nonconventional setup we would need a result of the type Lemma 3.3 with more than 3 \(\sigma\)-algebras similar to Lemma 3.3 in [16] which was proved under \(\psi\)-mixing. Such a result is not available under \(\phi\)-mixing.

5. Estimates on ball approximations

Throughout this section, \(\mu\) is a \(T\)-invariant ergodic measure, and \(\mathcal{A}\) is a measurable, generating partition of \(M\) satisfying Assumption A1. To simplify the notations, we will also assume that A1 holds true with \(C = 1\). We will demonstrate now how to approximate a geometric ball \(B_r(x)\) by a union of \(n\)-cylinders. For every \(x \in M\), \(r > 0\) and \(k \in \mathbb{N}\) define

\[
U^-(x, r, k) = \bigcup_{A \in \mathcal{A}^k, A \subset B_r(x)} A
\]

which is the union of all \(k\)-cylinders contained in \(B_r(x)\). Since \(\mathcal{A}\) is generating, \(U^-(x, r, k) \neq \emptyset\) for sufficiently large \(k\).

Also define

\[
U^+(x, r, k) = \bigcup_{A \in \mathcal{A}^k, A \cap B_r(x) \neq \emptyset} A.
\]

Then, \(B_r(x) \subset U^+(x, r, k) \subset B_{r+k-r}(x)\) due to Assumption A1.
The difference between $U^-(x, r, k)$ and $U^+(x, r, k)$ is 

$$U^+(x, r, k) \setminus U^-(x, r, k) = \bigcup_{A \in A^k, A \cap B_r(x) \neq \emptyset, A \subsetneq B_r(x)} A.$$  

In order to get a nice approximation of $B_r(x)$ by $U^\pm(x, r, k)$, we need to make the diameter of $k$-cylinders small compared to the size of the ball. For this purpose, we fix some $w > 1$ that will be determined in Section 7 and put

$$n = n(r) = \lfloor r^{-\frac{w}{p}} \rfloor + 1.$$  

This guarantees that diam$A^n < n(r)^{-p} < r^w$. It follows that

$$U^+(x, r, n(r)) \setminus U^-(x, r, n(r)) \subset B_{r+r^w}(x) \setminus B_{r-r^w}(x).$$  

In particular, this shows that

$$B_{r-r^w}(x) \subset U^-(x, r, n(r)) \subset B_r(x),$$  

and the difference between $U^-(x, r, n(r))$ and $B_r(x)$ is small:

$$\mu(B_r(x) \setminus U^-(x, r, n(r))) < \mu(B_r(x) \setminus B_{r-r^w}(x)) \leq C r^{wa-b} \mu(B_r(x)),$$

according to Assumption A4. Similarly we have

$$B_r(x) \subset U^+(x, r, n(r)) \subset B_{r+r^w}(x),$$

with the difference given by

$$\mu(U^+(x, r, n(r)) \setminus B_r(x)) < \mu(B_{r+r^w}(x) \setminus B_r(x)) \leq C r^{wa-b} \mu(B_r(x)).$$

6. Recurrence rate for metric balls and an estimate on the short return.

Define the lower recurrence rate for points $x$ and $y$ by

$$\underline{R}(x, y) = \liminf_{r \to 0} \frac{\log \tau_{B_r(y)}(x)}{-\log r},$$

where

$$\tau_U(x) = \min\{k > 0 : T^k x \in U\},$$

and set $\underline{R}(x) = \underline{R}(x, x)$. The latter was studied in [23, 19, 10]. The following Theorem was first stated in [23] for fast mixing systems (systems that have super-polynomial decay of correlation), and for Young towers with polynomial tails in [19] as Theorem 6.2.
6.1. THEOREM: Assume that Assumptions A1–A4 are satisfied. Then

\[ R(x) = \dim_H \mu, \quad \mu\text{-almost everywhere.} \]

In order to prove Theorem 6.1, we will need the following de-correlation lemma.

6.2. LEMMA: There exists \( C' > 0 \) such that for every \( x \in M, \ r > 0 \) and integers \( 0 \leq k \leq N \),

\[ \mu\{y \in B_r(x) : \tau_{B_r(x)}(T^k(y)) < N\} \leq C' \mu(B_{r+s}(x))k^{-\beta} + 2(N - k)\mu(B_{r+s}(x))^2, \]

where \( s = C(k/2)^{-p} \).

Proof. Recall that \( B_r(x) \subset U^+(x, r, k/2) \subset B_{r+s}(x) \) with \( s = C(k/2)^{-p} \). Also note that

\[ \{y : \tau_{B_r(x)}(T^k(y)) < N\} = T^{-k}\{y : \tau_{B_r(x)}(y) < N - k\} \subset T^{-k}\{y : \tau_{U^+(x, r, k/2)}(y) < N - k\}. \]

By the \( \phi \)-mixing property, we get

\[ \mu\{y \in B_r(x) : \tau_{B_r(x)}(T^k(y)) < N\} \leq \mu(U^+(x, r, k/2) \cap T^{-k}\{y : \tau_{U^+(x, r, k/2)}(y) < N - k\} + \phi(k/2) \]

\[ \leq \mu(B_{r+s}(x))\mu\{y : \tau_{B_{r+s}(x)}(y) < N - k\} + \phi(k/2) \]

\[ \leq C\mu(B_{r+s}(x))(k/2)^{-\beta} + (N - k)\mu(B_{r+s}(x))^2, \]

where in the last inequality we use the coarse estimate \( \mu(\tau_A < n) \leq n\mu(A) \) for every measurable set \( A \).

Now the proof of Theorem 6.1 follows the lines of Theorem 6.2 in [19] and Lemma 6.2 above, which replaces Proposition 5.1 in [19]. Note that although in [19] \( T \) is assumed to be an invertible map, the proof of Theorem 6.2 in [19] does not require that. In fact, the argument in [19] is taken from Lemma 16 in [23] which is stated for any measure preserving system.

Next, define the hitting time exponent for the pair \((x, y) \in M \times M\) to be

\[ R_{x,y} = \min\{R(x), R(y), R(x, y), R(y, x)\}. \]

Combining Theorem 6.1 with Proposition 1 in [10], we immediately obtain the following lower bound on the hitting time exponent:
6.3. Proposition: Assume that Assumptions A1–A4 hold true. Then
\[ R_{x,y} \geq \dim_H \mu \text{ for } \mu \times \mu \text{-a.e. } (x,y) \in M \times M. \]

When trying to apply the arguments from Section 4 to geometric balls, the major difficulty arises in obtaining an appropriate estimate for short returns, namely for \( b_2 \) in (4.7). On one hand, we need large \( n \) for the approximation of metric balls by \( n \)-cylinder sets, as observed in Section 5, and on the other hand, if we put \[ \pi(B_r(x), B_r(y)) = \min\{k \geq 0 : T^{-k}B_r(x) \cap B_r(y) \neq \emptyset \text{ or } T^{-k}B_r(y) \cap B_r(x)\} \]
and define \( \kappa_{B_r(x), B_r(y)} \) as in the case of cylinders, then it is easy to see that \( \kappa_{B_r(x), B_r(y)} \) is of order \( |\log r| \), which is much smaller than \( n(r) = \lceil r^{-\frac{m}{d}} \rceil + 1 \).
In particular, we cannot expect to have sufficient de-correlation before the first return happens. To solve this issue, we make the following crucial observation concerning the \( b_2 \) term in the estimates from [4].

6.4. Remark: The \( b_2 \) term in (4.7) was used in Theorem 3 of [4] to control terms of the form
\[ \sum_{\alpha \in I} E(X_{\alpha}[f_i(W_{\alpha} + e_i) - f_i(V_{\alpha} + e_i)]), \]
where \( W_{\alpha} \) and \( V_{\alpha} \) are certain random \( d \)-vectors which are different only if \( \sum_{\alpha \neq \beta \in B_{\alpha}} X_{\beta} \geq 1 \), \( X_{\alpha} \) with \( \alpha \in I \) are Bernoulli random variables, \( \alpha \in B_{\alpha} \subset I \), \( \forall \alpha \in I \) and \( e_i \) is the unit vector with 1 on the \( i \)th place. Then we can write
\[
\sum_{\alpha \in I} E(X_{\alpha}[f_i(W_{\alpha} + e_i) - f_i(V_{\alpha} + e_i)]) \\
\leq 2\|f_i\|_{\infty} \sum_{\alpha \in I} E(X_{\alpha}I_{W_{\alpha} \neq V_{\alpha}}) \\
\leq 2\|f_i\|_{\infty} \sum_{\alpha \in I} P(\alpha = 1, \sum_{\alpha \neq \beta \in B_{\alpha}} X_{\beta} \geq 1).
\]
The term \( b_2 \) in [4] is obtained by estimating
\[ P\left(X_{\alpha} = 1, \sum_{\alpha \neq \beta \in B_{\alpha}} X_{\beta} \geq 1\right) \leq \sum_{\alpha \neq \beta \in B_{\alpha}} P(X_{\alpha} = 1, X_{\beta} = 1) \]
which is too coarse for our purposes. Thus it is possible to replace \( b_2 \) by
\[
b'_2 = \sum_{\alpha \in I} P\left(X_{\alpha} = 1, \sum_{\alpha \neq \beta \in B_{\alpha}} X_{\beta} \geq 1\right).
\]
and (4.6) will still hold with \(b'_2\) in place of \(b_2\). It turns out, that we will be able to estimate \(b'_2\) in a better way sufficient for our proof.

Let
\[
X_{k,0} = I_{B_r(y)} \circ T^k \quad \text{and} \quad X_{k,1} = I_{B_r(x)} \circ T^k,
\]
and for fixed \(k \in \mathbb{N}, \alpha = 0,1\) and \(M, R > 0\) let
\[
B^{M,R}_{k,\alpha} = \{(l, \alpha), (l, 1 - \alpha) : 0 \leq l \leq M, |l - k| \leq R\}
\]
be the same as in Section 4, with \(U = B_r(x)\) and \(V = B_r(y)\). The following proposition gives an (optimal) estimate on \(b'_2\), which will be used in the next section.

6.5. **Proposition:** Assume that Assumptions A1–A4 hold true. Then for every \(\alpha = 0,1\), positive integers \(M\) and \(k\), positive real number \(\sigma < \dim H \mu\) and \(\mu \times \mu\)-a.e. \((x,y)\) such that the ratio \(\frac{\mu(B_r(x))}{\mu(B_r(y))}\) remains sandwiched between two positive constants as \(r \to 0\),
\[
\mu\left\{ z : X_{k,\alpha}(z) = 1, \sum_{(k,\alpha) \neq (l,\beta) \in B^{M,\sigma,r}_{k,\alpha}} X_{l,\beta} \geq 1 \right\} \leq v_{x,y}(\mu(B_r(x)))
\]
where \(v(u) = v_{x,y}(u) = o(u)\) as \(u \to 0\) and \(v\) does not depend on \(k\) and \(M\). In particular, if \(M = M(r) \to \infty\) is such that
\[
M(r) v_{x,y}(\mu(B_r(x))) \to 0 \quad \text{as} \quad r \to 0,
\]
then
\[
b'_2 = b'_2(x,y) = \sum_{(k,\alpha) \in I_M} \mu\left\{ z : X_{k,\alpha}(z) = 1, \sum_{(k,\alpha) \neq (l,\beta) \in B^{M,\sigma,r}_{k,\alpha}} X_{l,\beta} \geq 1 \right\}
\]
\[
\leq M(r) v_{x,y}(\mu(B_r(x))) \to 0, \quad \text{as} \quad r \to 0.
\]

**Proof.** We will only consider the case \(\alpha = 1\). The case \(\alpha = 0\) is similar by switching \(x\) and \(y\). Observe that since the ratio of \(\mu(B_r(x))\) and \(\mu(B_r(y))\) is sandwiched between two positive constants, we can interchange between \(v(\mu(B_r(x)))\) and \(v(\mu(B_r(y)))\) in the estimates.
Case 1. $\alpha = \beta = 1$.

First note that when $\alpha = \beta = 1$,

$$ \mu \left\{ z : X_{k,1}(z) = 1, \sum_{(k,1) \neq (l,1) \in B_{k,1}^{M,r-\sigma}} X_{l,1} \geq 1 \right\} = \mu \left( (T^{-k}B_r(x)) \cap \bigcup_{(k+r-\sigma) \wedge M \geq l \geq 0 \vee (k-r-\sigma)} T^{-l}B_r(x) \right). $$

We split this into two parts that correspond to $l < k$ and $l > k$, respectively,

$$ b^- = b^-_k(x) = \mu(T^{-k}B_r(x) \cap \bigcup_{l=1}^{k \wedge r-\sigma} T^{-(k-l)}B_r(x)), $$

and

$$ b^+ \leq b^+_k(x) = \mu \left( T^{-k}B_r(x) \cap \bigcup_{l=1}^{r-\sigma} T^{-(k+l)}B_r(x) \right). $$

(i) First, we will prove that $b^- \leq b^+$ and then it will remain only to estimate $b^+$. It suffices to show that

$$ \mu \left( T^{-k}B_r(x) \cap \bigcup_{l=1}^{k \wedge r-\sigma} T^{-(k-l)}B_r(x) \right) = \mu \left( T^{-k}B_r(x) \cap \bigcup_{l=1}^{r-\sigma} T^{-(k+l)}B_r(x) \right). $$

For this purpose, we write $J_l = T^{-k}B_r \cap T^{-(k-l)}B_r$ and $U = \bigcup_{l=1}^{k \wedge r-\sigma} J_l$. Similarly $\tilde{J}_l = T^{-k}B_r(x) \cap T^{-(k+l)}B_r(x)$ and $\tilde{U} = \bigcup_{l=1}^{k \wedge r-\sigma} \tilde{J}_l$. In order to show that $\mu(U) = \mu(\tilde{U})$, we decompose $U$ into a disjoint union

$$ U = \bigcup_{l=1}^{l-1} \tilde{J}'_l \text{ where } \tilde{J}'_l = \tilde{J}_l \setminus \bigcup_{j=1}^{l-1} \tilde{J}_l \cap \tilde{J}_j. $$

Similarly, $\tilde{J}'_l = \tilde{J}_l \setminus \bigcup_{j=1}^{l-1} \tilde{J}_l \cap \tilde{J}_j$ decomposes $\tilde{U}$ into a disjoint union.

Note that $T^{-l}J_l = \tilde{J}_l$ when $l \leq k$, and so

$$ T^{-l}J'_l = T^{-l}J_l \setminus \bigcup_{j=1}^{l-1} T^{-l}(J_l \cap J_k) = \tilde{J}_l \setminus \bigcup_{j=1}^{l-1} (\tilde{J}_l \cap \tilde{J}_k) = \tilde{J}'_l. $$

This shows that $\mu(J'_l) = \mu(\tilde{J}'_l)$ when $l \leq k$. Summing over $l$, we get

$$ \mu(U) = \mu(\tilde{U}). $$
(ii) Next, we estimate \( b^+ \) using the Lebesgue density point technique employed in Proposition 7.1 from [19]. Fix \( \varepsilon_1, \varepsilon_2 > 0 \). Consider the “good” set:

\[
G_{\sigma,r_1} = \{ z \in \mathbb{M} : \forall r < r_1, \tau_{B_{2r}(z)}(z) \geq r^{-\sigma} \}.
\]

By Theorem 6.1, \( G_{\sigma,r_1} \uparrow G_{\sigma} \) as \( r_1 \downarrow 0 \) where \( \mu(\mathbb{M} \setminus G_{\sigma}) = 0 \) and, in particular, we can take \( r_1 \) small enough such that \( \mu(\mathbb{M} \setminus G_{\sigma,r_1}) < \varepsilon_1 \).

Let \( \tilde{G}_{\sigma,r_1} \) be the set of Lebesgue density points of \( G_{\sigma,r_1} \) with respect to the measure \( \mu \), that is, the set of points \( z \in G_{\sigma,r_1} \) such that

\[
\lim_{r \to 0} \frac{\mu(B_r(z) \cap G_{\sigma,r_1})}{\mu(B_r(z))} = 1.
\]

Here \( \tilde{G}_{\sigma,r_1} \) has full measure in \( G_{\sigma,r_1} \). Then we take \( r_2 < r_1 \), such that

\[
\tilde{G}_{\sigma,r_1,r_2} := \left\{ z : \forall r < r_2, \frac{\mu(B_r(z) \cap G_{\sigma,r_1})}{\mu(B_r(z))} > 1 - \varepsilon_2 \right\}
\]

satisfies

\[
\mu(G_{\sigma,r_1} \setminus \tilde{G}_{\sigma,r_1,r_2}) < \varepsilon_1.
\]

Now, for \( x \in \tilde{G}_{\sigma,r_1,r_2} \) and \( r < r_2 \),

\[
b^+ \leq b^+_k(x) = \mu(z \in T^{-k}B_r(x) : \tau_{B_r(x)}(T^k z) < r^{-\sigma})
\]

\[
= \mu(z \in B_r(x) : \tau_{B_r(x)}(z) < r^{-\sigma})
\]

\[
\leq \mu(z \in B_r(x) : \tau_{B_{2r}(x)}(z) < r^{-\sigma})
\]

\[
\leq \mu(B_r(x) \setminus G_{\sigma,r_1}) \leq \varepsilon_2 \mu(B_r(x)),
\]

which fails only on the set \( \mathbb{M} \setminus \tilde{G}_{\sigma,r_1,r_2} \) with measure less than \( 2\varepsilon_1 \) and which decreases as \( r_1, r_2 \downarrow 0 \) to a set of \( \mu \)-measure 0.

**Case 2.** \( \alpha = 1, \beta = 0 \).

First note that when \( l = k \) then \( T^{-k}B_r(x) \cap T^{-l}B_r(y) \neq \emptyset \) if and only if \( B_r(x) \cap B_r(y) \neq \emptyset \), which can be avoided if \( r \) is taken small enough. As in the previous case, we have two sub-cases: \( l < k \) and \( l > k \). Denote

\[
c^-(x,y) = \mu\left( T^{-k}B_r(x) \cap \bigcup_{l=1}^{k \wedge r^{-\sigma}} T^{-(k-l)}B_r(y) \right)
\]

and

\[
c^+(x,y) = \mu\left( T^{-k}B_r(x) \cap \bigcup_{l=1}^{r^{-\sigma}} T^{-(k+l)}B_r(y) \right).
\]
By an argument similar to the $b^-$ case (with one of the $x$ replaced by $y$) we see that $c^-(x,y) \leq c^+(y,x)$, where the latter can be seen as a part of the case $\alpha = 0$, $\beta = 1$. Hence, as before, we only need to estimate $c^+(x,y)$ and the proposition follows.

By Proposition 6.3 and the Fubini theorem, for almost every $y \in M$ we have $R(x,y) \geq \dim_H \mu$ for almost every $x$. For such fixed $y$, consider the “good” set

$$G^y_{\sigma,r_1} = \{z \in M : \forall r < r_1, \tau_{B_r(y)}(z) \geq r^{-\sigma}\}.$$ 

Again, $G^y_{\sigma,r_1} \uparrow G^y_0$ as $r_1 \downarrow 0$, and so we can take $r_1$ small enough so that $\mu(M \setminus G^y_{\sigma,r_1}) < \varepsilon_1$.

The rest of the proof proceeds in the same way as in the case $b^+$. We take $\tilde{G}^y_{\sigma,r_1}$ to be the set of Lebesgue density points of $G^y_{\sigma,r_1}$,

$$\tilde{G}^y_{\sigma,r_1} = \left\{z : \lim_{r \to 0} \frac{\mu(B_r(z) \cap G^y_{\sigma,r_1})}{\mu(B_r(z))} = 1 \right\}.$$ 

Then we take $r_2 < r_1$ so that the set

$$\tilde{G}^y_{\sigma,r_1,r_2} := \left\{z : \forall r < r_2, \frac{\mu(B_r(z) \cap G^y_{\sigma,r_1})}{\mu(B_r(z))} > 1 - \varepsilon_2 \right\}$$

satisfies

$$\mu(G^y_{\sigma,r_1} \setminus \tilde{G}^y_{\sigma,r_1,r_2}) < \varepsilon_1.$$ 

Now, for $x \in \tilde{G}^y_{\sigma,r_1,r_2}$ and $r < r_2$,

$$c^+ = c^+(x,y) = \mu(z \in T^{-n}B_r(x) : \tau_{B_r(y)}(T^n z) < r^{-\sigma})$$

$$= \mu(z \in B_r(x) : \tau_{B_r(y)}(z) < r^{-\sigma})$$

$$\leq \mu(B_r(x) \setminus G^y_{\sigma,r_1}) \leq \varepsilon_2 \mu(B_r(x))$$

concluding the argument as in the previous case and completing the proof of Proposition 6.5. ■

Recall that the sets $U^\pm(x,r,n(r))$ are defined in Section 5 as the approximation of $B_r(x)$ by $n(r)$-cylinders from inside and outside, respectively. Since

$$U^-(x,r,n(r)) \subset B_r(x) \quad \text{and} \quad U^-(y,r,n(r)) \subset B_r(y),$$

it follows that

$$\tilde{X}_{k,0} = \mathbb{I}_{U^-(y,r,n(r))} \circ T^k \leq X_{k,0} \quad \text{and} \quad \tilde{X}_{k,1} = \mathbb{I}_{U^-(x,r,n(r))} \circ T^k \leq X_{k,1}.$$
Hence,
\[
\left\{ \sum_{(l,\beta) \in B_{n,\alpha}^{M,r} - \sigma} \tilde{X}_{l,\beta} \geq 1 \right\} \subset \left\{ \sum_{(l,\beta) \in B_{n,\alpha}^{M,r} - \sigma} X_{l,\beta} \geq 1 \right\}
\]

and so
\[
\tilde{b}_2' = \tilde{b}_2'(x, y) \leq \sum_{(k,\alpha) \in I_M} \mu \left\{ z : \tilde{X}_{k,\alpha}(z) = 1, \sum_{(k,\alpha) \neq (l,\beta) \in B_{n,\alpha}^{M,r} - \sigma} \tilde{X}_{l,\beta}(z) \geq 1 \right\} \leq b_2'.
\]

This together with Proposition 6.5 yields

**6.6. Proposition:** Assume that Assumptions A1–A4 hold true. Then for every \( \alpha = 0, 1 \), a positive integer \( M \), a positive real number \( \sigma < \dim H \mu \) and \( \mu \times \mu \)-a.e. \((x, y)\) for which \( \rho(x, y, r) \to \rho(x, y) \in (0, 1) \) as \( r \to 0 \),
\[
\tilde{b}_2' \leq M \nu(\mu(B_r(x)) \quad \text{where} \quad \nu(u) = \nu_{x,y}(u) = o(u) \text{ as } u \to 0 \text{ and } \nu \text{ does not depend on } M.
\]

**7. Proof of Theorem 2.5**

Recall that \( n(r) = \lfloor r^{-\frac{w}{p}} \rfloor + 1 \) where \( w > 1 \) will be determined later. We will write
\[
\tau = \tau_{B_r}(y), \quad \tau_M = \min(\tau, M), \quad \tilde{\tau} = \tau_{U^-(y, r, n(r))} \quad \text{and} \quad \tilde{\tau}_M = \min(\tilde{\tau}, M).
\]

Then, \( \tilde{\tau} \geq \tau \) and \( \tilde{\tau}_M \geq \tau_M \). Introduce the sums
\[
S_M(z) = \sum_{j=0}^{M} X_{k,1}(z) \quad \text{and} \quad \tilde{S}_M(z) = \sum_{j=0}^{M} \tilde{X}_{k,1}(z)
\]

which count the number of visits to \( B_r(x) \) and to \( U^-(x, r, n(r)) \), respectively.

Set also
\[
\tilde{\rho} = \frac{\mu(U^-(y, r, n(r)))}{\mu(U^-(x, r, n(r))) + \mu(U^-(y, r, n(r)))}.
\]

In other words, every term with a tilde is defined using the approximations \( U^-(x, r, n(r)) \) and \( U^-(y, r, n(r)) \). Similarly to Section 4 we introduce also a sequence of independent Bernoulli random variables \( \{Y_{k,\alpha} : k \geq 0, \alpha = 0, 1\} \) such that \( Y_{k,\alpha} \) has the same distribution as \( \tilde{X}_{k,\alpha} \) and we set
\[
\tilde{S}_M^* = \sum_{k=0}^{M-1} Y_{k,1}, \quad \tilde{\tau}^* = \min\{k \geq 0 : Y_{k,0} = 1\} \quad \text{and} \quad \tilde{\tau}_M^* = \min(\tilde{\tau}^*, M).
\]
As in (4.1) we estimate

\[ d_{TV}(L(\Sigma_r^x y), \text{Geo}(\rho(x, y, r))) \leq B_1 + B_2 + B_3 + B_4 + B_5, \]

where

\[ B_1 = d_{TV}(L(S_r), L(S_{\tilde{\tau}_M})), \quad B_2 = d_{TV}(L(S_{\tilde{\tau}_M}), L(\tilde{S}_{\tilde{\tau}_M})), \]
\[ B_3 = d_{TV}(L(\tilde{S}_{\tilde{\tau}_M}), L(S_{\hat{\tau}_M})), \quad B_4 = d_{TV}(L(S_{\hat{\tau}_M}), L(\tilde{S}_{\hat{\tau}_M})), \]
\[ B_5 = d_{TV}(\text{Geo}(\tilde{\rho}), \text{Geo}(\rho)). \]

Note that in \( B_1 \) we use \( \tilde{\tau}_M \) and not \( \tau_M \) as in (4.1), but this will not make an essential difference.

To obtain an estimate on \( B_1 \), we write

\[ B_1 \leq d_{TV}(L(S_r), L(S_{\tau_M})) + d_{TV}(L(S_{\tau_M}), L(S_{\tilde{\tau}_M})). \]

The first term is similar to \( A_1 \) from (4.1) and is bounded by \( \mu\{z : \tau(z) > M\} \).

For the second term, note that \( \tau_M \leq \tilde{\tau}_M \leq M \). If \( \tau_M \neq \tilde{\tau}_M \), then the first entry to \( B_r(y) \) must be contained in \( B_r(y) \setminus U^-(y, r, n(r)) \). Therefore

\[ d_{TV}(L(S_{\tau_M}), L(S_{\tilde{\tau}_M})) \leq \mu\{z : \tau_B(y) \setminus U^- (y, r, n(r))(z) < M\} \]
\[ \leq \mu\left( \bigcup_{j=0}^{M-1} T^{-j}(B_r(y) \setminus U^-(y, r, n(r))) \right) \]
\[ \leq M \mu(B_r(y) \setminus U^-(y, r, n(r))) \leq M r^{wa-b} \mu(B_r(y)), \]

where the last inequality follows from (5.1).

Next, \( B_5 \) is the same as \( A_4 \) in Section 4, and so

\[ B_5 \leq \frac{2|\hat{\rho} - \rho|}{\hat{\rho} \rho}. \]

To estimate \( B_2 \), note that \( U^-(x, r, n(r)) \subset B_r(x) \) which implies that \( \tilde{S}_M \leq S_M \). Therefore

\[ B_2 \leq \sum_{k=0}^{M} |\mu\{z : S_{\tilde{\tau}_M}(z) = k\} - \mu\{z : \tilde{S}_{\tilde{\tau}_M}(z) = k\}| \]
\[ \leq \sum_{k=0}^{M} \mu\{z : S_{\tilde{\tau}_M}(z) > k, \tilde{S}_{\tilde{\tau}_M}(z) = k\}. \]
Since \( U^{-}(x, r, n(r)) \subset B_{r}(x) \), all the extra visits \( T^{i} z \in B_{r}(x) \), which make \( S_{\tilde{T}, M} \) larger than \( \tilde{S}_{\tilde{T}, M} \), must be contained in \( B_{r}(x) \setminus U^{-}(x, r, n(r)) \). Thus, by Assumption A4,

\[
B_{2} \leq \sum_{k=0}^{M} (M - k) \mu(B_{r}(x) \setminus U^{-}(x, r, n(r))) \leq \frac{1}{2} M^{2} r^{wa-b} \mu(B_{r}(x)).
\]

(7.1)

Observe that \( B_{3} \) and \( B_{4} \) are the same as \( A_{2} \) and \( A_{3} \) in Section 4 but we estimate \( B_{3} \) by (4.5) and (4.6) with \( b_{2} \) replaced by \( \tilde{b}_{2} \) in view of Remark 6.4. Hence,

\[
B_{4} \leq \mu\{z : \tilde{\tau}(z) > M\} = (1 - \mu(U^{-}(y, r, n(r)))) M^{+1},
\]

and setting \( p_{x} = \mu(U^{-}(x, r, n(r))) \) and \( p_{y} = \mu(U^{-}(y, r, n(r))) \) we obtain, using Proposition 6.6, that

\[
B_{3} \leq 8MR(p_{x} + p_{y})^{2} + 4Mv(\mu(B_{r}(x))) + 12M\phi(R - n) + 4M(p_{x}^{2} + p_{y}^{2}).
\]

Combining the above estimates we obtain

\[
d_{TV}(\mathcal{L}(\Sigma_{r}^{x,y}), \text{Geo}(\rho))
\]

\[
\leq \mu\{z : \tau(z) > M\} + Mr^{wa-b} \mu(B_{r}(y))
\]

\[
+ \frac{1}{2} M^{2} r^{wa-b} \mu(B_{r}(x)) + 4MR(p_{x} + p_{y})^{2}
\]

\[
+ 2M v(\mu(B_{r}(x))) + 6M\phi(R - n(r))
\]

\[
+ 2M(p_{x}^{2} + p_{y}^{2}) + (1 - \mu(U^{-}(y, r, n(r)))) M^{+1} + \frac{2|\tilde{\rho} - \rho|}{\tilde{\rho} \rho}.
\]

(7.2)

Next, observe that \( \tilde{\tau} \geq \tau \), and so \( \{z : \tilde{\tau}(z) > M\} \supset \{z : \tau(z) > M\} \). On the other hand,

\[
\{z : \tilde{\tau}(z) > M\} \setminus \{z : \tau(z) > M\} = \{z : \tilde{\tau}(z) > M, \tau(z) \leq M\}
\]

\[
\subset \bigcup_{k=0}^{M} \{z : T^{k} z \in B_{r}(y) \setminus U^{-}(y, r, n(r))\}.
\]

Hence, by Assumption A4,

\[
\mu\{z : \tau(z) > M\} \leq \mu\{z : \tilde{\tau}(z) > M\} + \sum_{k=0}^{M} \mu(B_{r}(y) \setminus U^{-}(y, r, n(r)))
\]

\[
\leq \mu\{z : \tilde{\tau}(z) > M\} + (M + 1) r^{wa+b} \mu(B_{r}(y)).
\]

(7.3)

Similarly to (4.2) we obtain that

\[
\mu\{z : \tilde{\tau}(z) > M\} \leq (1 - p_{y}) M^{+1} + d_{TV}(\mathcal{L}(\tilde{X}_{M}), \mathcal{L}(\tilde{Y}_{M}))
\]

(7.4)
where
\[ \tilde{X}_M = \{ \tilde{X}_{k,\alpha}, 0 \leq k \leq M, \alpha = 0,1 \} \quad \text{and} \quad Y_M = \{ Y_{k,\alpha}, 0 \leq k \leq M, \alpha = 0,1 \}. \]
In the same way as in (4.6) with \( \tilde{b}_2 \) in place of \( b_2 \) we estimate
\[ d_{TV}(\mathcal{L}(\tilde{X}_M), \mathcal{L}(Y_M)) \leq 4M R (p_x + p_y)^2 + 2M v(\mu(B_r(x))) + 6M \phi(R - n(r)) + 2M (p_x^2 + p_y^2). \]

When \( wa - b > 0 \) and \( \rho(x, y, r) \to \rho \in (0,1) \) as \( r \to 0 \), then by (5.1) the pairwise ratios of \( \mu(U^\pm(x, r, n(r))) \), \( \mu(U^\pm(y, r, n(r))) \), \( \mu(B_r(x)) \) and \( \mu(B_r(y)) \) are sandwiched between positive constants. This together with Proposition 6.3 yields that for \( \mu \times \mu \)-almost all \( x, y \) we can choose \( M = M_{x,y}(r) \to \infty \) as \( r \to 0 \) so that
\[
M_{x,y}^2(r) r^{wa-b} \mu(B_r(x)) \to 0, \quad M_{x,y} r \mu(B_r(x)) \to 0, \quad (1 - \mu(U^-(y, r, n(r)))) M_{x,y} \to 0 \quad \text{and} \quad \frac{|\tilde{\rho} - \rho|}{\tilde{\rho} \rho} \to 0 \quad \text{as} \quad r \to 0.
\]
Observe that the assumption (2.6) enables us to choose \( w \) and \( \sigma \) so that \( wa - b > d, d > \sigma, \sigma > \frac{w}{p} \) and \( \sigma \beta > d \). Indeed, these inequalities will hold true if we choose \( \sigma = d - \varepsilon \) and \( w = dp - \delta \) for \( \varepsilon > 0 \) and \( \delta > 0 \) satisfying \( p \varepsilon < \delta < dp - \frac{b+d}{a} \) and \( \varepsilon < d(1 - \beta^{-1}) \), which is possible in view of (2.6) and since \( \beta > 1 \). With such \( w \) and \( \sigma \) take \( R = R(r) = r^{-\sigma} \). Then, as \( r \to 0 \), for \( \mu \times \mu \)-almost all \( (x, y) \) we can choose \( M_{x,y}(r) \to \infty \) so that both (7.6) and
\[ M_{x,y}(r) R(r) (p_x + p_y)^2 \to 0 \quad \text{and} \quad M_{x,y}(r) \phi(R(r) - n(r)) \to 0 \]
hold true. This completes the proof of Theorem 2.5 in view of (7.2)–(7.6).

\section{8. Geometric law limit for suspensions}

In this section we will prove Theorem 2.11. Given a well-placed set \( U \subset \Omega \), the set \( \{(x,0) : (x,k) \in U \text{ for some } k\} \subset \Omega_0 \) can be naturally identified with \( \Pi(U) \) by dropping the second coordinate and it will also be denoted by \( \tilde{U} \).

\textbf{8.1. Lemma:} Let \( (\Omega, \mu, T) \) be a discrete time suspension over a measure preserving dynamical system \( (\tilde{\Omega}, \tilde{\mu}, \tilde{T}) \) and \( U, V \) be a pair of well-placed sets such that their projections to the base \( \tilde{U} = \Pi(U) \) and \( \tilde{V} = \Pi(V) \) are disjoint. Introduce the set \( W = W_U = \{(x,i) \in \Omega : i > 0 \text{ and there exists } j \geq i \text{ such that } (x,j) \in U \} \). Then \( \Sigma_{U,V}(x,i) = \Sigma_{\tilde{U}, \tilde{V}}(x,i) \) for any \( (x,i) \in \Omega \setminus W \) and \( \Sigma_{U,V}(x,i) = \Sigma_{\tilde{U}, \tilde{V}}(x,i) + 1 \) if \( (x,i) \in W \).
Proof. By the definition of a discrete time suspension and taking into account that $\tilde{U} \cap \tilde{V} = \emptyset$, each time the orbit $T^k(x, i), k \geq 0$ visits $\tilde{U}$ it must visit $U$ before it can visit $\tilde{V}$ or $V$. After each visit to $U$ the orbit either visits $\tilde{V}$ and then $V$ which stops the counting or it visits $\tilde{U}$ and then $U$ and the process repeats itself. Since $U$ is well-placed, any orbit must visit once $\tilde{U}$ between two successive visits to $U$. Thus, if an orbit visits $\tilde{U}$ for the first time before it ever visited $U$ then there will be the same number of visits to each of them before the first visit to $V$. On the other hand, if $(x, i) \in W$ then the orbit $T^k(x, i), k \geq 0$ visits first $U$ before it visits $\tilde{U}$, and so in this case there will be one more visit to $U$ than to $\tilde{U}$. □

Now we are ready to prove Theorem 2.11.

Proof of Theorem 2.11. Let $U_n$ and $V_n$, $n \geq 1$ be decreasing sequences of measurable subsets of $\Omega$ such that $\tilde{U}_n \cap \tilde{V}_n = \emptyset$ and $\tilde{\mu}(\tilde{U}_n), \tilde{\mu}(\tilde{V}_n) \to 0$ as $n \to \infty$. Then the set $W_{U_n}$ appearing in Lemma 8.1 satisfies

$$\mu(W_{U_n}) \leq \int_{\tilde{\Omega}_n} R(x) d\tilde{\mu}(x) \to 0.$$ 

This together with Lemma 8.1 yields that Theorem 2.11 will follow if we show that

$$(8.1) \quad \Sigma_{\tilde{U}_n, \tilde{V}_n} \overset{\mu}{\Rightarrow} \text{Geo}(\rho).$$

Next, define the probability measure $\tilde{\mu}$ on $\tilde{\Omega}$ by setting

$$\tilde{\mu}(\Gamma) = \left( \int_{\tilde{\Omega}} R(x) d\tilde{\mu}(x) \right)^{-1} \int_{\Gamma} R(x) d\tilde{\mu}(x)$$

for any measurable set $\Gamma \subset \tilde{\Omega}$. Observe that for each $x \in \tilde{\Omega}$ the number of visits to $\tilde{U}_n$ by the orbit $T^k(x, i), k \geq 1$ until the first visit to $\tilde{V}_n$ is the same for all $i = 0, 1, \ldots, R(x) - 1$. It follows that the asymptotical behavior of $\Sigma_{\tilde{U}_n, \tilde{V}_n}(z)$ is the same for all $z \in \Omega$ having the same projection to $\tilde{\Omega}$, and so the convergence $\Sigma_{\tilde{U}_n, \tilde{V}_n} \overset{\mu}{\Rightarrow} \text{Geo}(\rho)$ is equivalent to the convergence $\Sigma_{\tilde{U}_n, \tilde{V}_n} \overset{\tilde{\mu}}{\Rightarrow} \text{Geo}(\rho)$ where the latter is considered on the probability space $(\tilde{\Omega}, \tilde{\mu})$. By the conditions of Theorem 2.11, Corollary 2.2 and Theorem 2.5 we know that $\Sigma_{\tilde{U}_n, \tilde{V}_n} \overset{\tilde{\mu}}{\Rightarrow} \text{Geo}(\rho)$, and so $\Sigma_{\tilde{U}_n, \tilde{V}_n} \overset{\tilde{\mu}}{\Rightarrow} \text{Geo}(\rho)$ holds true by Theorem 1 from [26], completing the proof of Theorem 2.11. □
9. Applications

9.1. Gibbs–Markov systems. As the first example, we consider a Gibbs–Markov map $T$ on a Lebesgue space $(X, \mu)$. Recall that a map $T$ is called Markov if there is a countable measurable partition $\mathcal{A}$ on $X$ with $\mu(A) > 0$ for all $A \in \mathcal{A}$, such that for all $A \in \mathcal{A}$, $T(A)$ is injective and can be written as a union of elements in $\mathcal{A}$. Write

$$\mathcal{A}^n = \bigvee_{j=0}^{n-1} T^{-j} \mathcal{A}$$

as before; it is also assumed that $\mathcal{A}$ is (one-sided) generating.

Fix any $\gamma \in (0, 1)$ and define the metric $d_\gamma$ on $X$ by $d_\gamma(x, y) = \gamma^{s(x,y)}$, where $s(x, y)$ is the largest positive integer $n$ such that $x, y$ lie in the same $n$-cylinder. Define the Jacobian

$$g = JT^{-1} = \frac{d\mu}{d\mu \circ T}$$

and

$$g_k = g \circ T \circ \cdots \circ T^{k-1}.$$

The map $T$ is called Gibbs–Markov if it preserves the measure $\mu$, and also satisfies the following two assumptions:

(i) The big image property: there exists $C > 0$ such that $\mu(T(A)) > C$ for all $A \in \mathcal{A}$.

(ii) Distortion: $\log g|_A$ is Lipschitz with the same constant for all $A \in \mathcal{A}$.

In view of (i) and (ii), there exists a constant $D > 1$ such that for all $x, y$ in the same $n$-cylinder, we have the following distortion bound:

$$\left| \frac{g_n(x)}{g_n(y)} - 1 \right| \leq Dd_\gamma(T^n x, T^n y),$$

and the Gibbs property:

$$D^{-1} \leq \frac{\mu(A_n(x))}{g_n(x)} \leq D.$$

It is well known (see, for example, Lemma 2.4(b) in [18]) that Gibbs–Markov systems are exponentially $\phi$-mixing. Therefore we have the following corollaries of Theorems 2.1 and 2.5:
9.1. Theorem: Let $T$ be a Gibbs–Markov map on $(X, \mu)$ with finite entropy
\[- \sum_{A \in \mathcal{A}} \mu(A) \log \mu(A) < \infty.\]

Let $\{m(n)\}_{n \geq 1} \subset \mathbb{N} \setminus \{0\}$ be a sequence satisfying $|m(n) - n| = o(n)$ as $n \to \infty$. Then for $\mu \times \mu$-a.e. $(\omega, \eta) \in X \times X$,
\[\lim_{n \to \infty} d_{TV}(L(\Sigma_{n,m(n)}^{\omega,\eta}), \text{Geo}\left(\frac{\mu(A^n_{m(n)})}{\mu(A^n_{m(n)}) + \mu(A^n_{\eta})}\right)) = 0.\]

In particular, if
\[\lim_{n \to \infty} \frac{\mu(A^n_{\eta})}{\mu(A^n_{m(n)})} = \lambda,\]
then $L(\Sigma_{n,m(n)}^{\omega,\eta})$ converges in total variation as $n \to \infty$ to the geometric distribution with the parameter $(1 + \lambda)^{-1}$.

9.2. Theorem: Let $X$ be a compact metric space. Assume that there exists a countable generating partition $\mathcal{A}$, such that the measure preserving system $(T, \mu)$ is Gibbs–Markov. Also assume that Assumptions A1, A3 and A4 are satisfied with $p > d + b$, and $\text{diam} A^n \leq \gamma^n$, and the assumption $p > \frac{d+b}{ad}$ is redundant.

Then for $\mu \times \mu$-almost every $(x, y) \in \mathcal{M} \times \mathcal{M}$, such that $\rho(x, y, r) \to \rho(x, y) \in (0, 1)$ as $r \to 0$,
\[\lim_{r \to 0} d_{TV}(L(\Sigma_{r}^{x,y})^{\rho(x,y)}), \text{Geo}(\rho(x,y))) = 0.\]

9.3. Remark: In many cases (for example, those in [12]), the metric $d_\gamma$ is equivalent to the metric of $X$. Then Assumption A1 is satisfied with $\text{diam} A^n \leq \gamma^n$, and the assumption $p > \frac{d+b}{ad}$ is redundant.

9.2. Conformal repellers. A conformal repeller is a maximal compact set $\Omega \subset M$ so that $T$ acts conformally on $\Omega$ and is expanding, that is there exists a $\beta > 1$ so that $|DT^kv| \geq \beta^k$ for all large enough $k$ and all $v \in T_x M \forall x \in \Omega$.

9.4. Theorem: Let $\Omega \subset M$ be a conformal repeller for the $C^{1+\alpha}$-map $T : M \to M$ and let $\mu$ be an equilibrium state for a Hölder continuous potential $f : \Omega \to \mathbb{R}$. Then the conclusion of Theorem 2.5 is valid.

Proof. We verify Assumptions A1–A4 using the fact that Markov partitions $\mathcal{A}$ of arbitrarily small diameter can be constructed here. Let $\mathcal{A}$ be such a partition. Then:
Assumption A1 holds since the map is uniformly expanding, so diam $A^n \leq \beta^{-n}$.

Assumption A2 follows from the fact that the equilibrium states are $\psi$-mixing with respect to the partition $\mathcal{A}$ at exponential speed, thus left $\phi$-mixing.

Assumption A3 is shown in Theorem 21.3 of [20] that every equilibrium state on a conformal repeller has exact dimension.

Assumption A4 is satisfied for any $w > 1$ as $\mu$ is diametrically regular [21] and thus also has the annular decay property [8]. This yields

$$\frac{\mu(B_{r+r^w}(x) \setminus B_r(x))}{\mu(B_r(x))} = O(r^{(w-1)\delta}) \to 0$$

for some $\delta > 0$.

9.3. Young towers. First introduced by Young [24], [25], Young towers (or Gibbs–Markov–Young structure) is a useful tool to study the statistical property of $C^{1+\alpha}$ systems that are non-uniformly hyperbolic. When the system is non-invertible, a Young tower can be viewed as a discrete time suspension over a Gibbs–Markov system, such that the roof function $R$ (in this case, it is usually call the return time function) is integrable with respect to the measure $\tilde{\mu}$. As an immediate corollary of Theorems 2.11, 9.1 and 9.2, we have:

9.5. THEOREM: Let $(\Omega, \mu, T)$ be a Young tower such that the return time function $R$ is integrable. Assume that $\{U_n\}, \{V_n\}$ are two sequences of well-placed sets.

(i) If $\tilde{U}_n = A_{n}^\omega, \tilde{V}_n = A_{m(n)}^n$ for some $\omega, \eta$ and $m(n)$ satisfy the assumptions of Theorem 9.1 and the base system has finite entropy, then $\Sigma_{U_n, V_n}$ converges in distribution to the geometric distribution $\text{Geo}((1 + \lambda)^{-1})$, where $\lambda$ is given by

$$\lim_{n \to \infty} \frac{\mu(A_n^\omega)}{\mu(A_{m(n)}^n)} = \lambda.$$

(ii) If the base system $(\tilde{\Omega}, \tilde{\mu}, \tilde{T})$ satisfies A1, A3 and A4, and $\tilde{U}_n = B_{r_n}(x), \tilde{V}_n = B_{r_n}(y)$ satisfy the assumptions of Theorem 9.2 for some sequence of positive real numbers $\{r_n\}$ with $r_n \to 0$, then $\Sigma_{U_n, V_n}$ converges in distribution to $\text{Geo}(\rho(x,y))$.

1 When $T$ is a diffeomorphism, in order to obtain a Gibbs–Markov system one needs to take the quotient over the stable leaves.

2 Alternatively one can assume that the roof function is integrable w.r.t. a reference measure $m$, for which the map $\tilde{T}$ has a “good” Jacobian. Under these conditions, it is shown in [25] that there exists an invariant measure $\tilde{\mu}$, absolutely continuous w.r.t. $m$. 
There are plenty dynamical systems which can be modeled by Young towers with the integrable return map being the first return map to the set which serves as a base. Then the dynamical system and its Young tower representation are isomorphic and limit theorems for both systems are equivalent. The corresponding one-dimensional examples include the uniform expanding piecewise $C^2$-map with the Markov property, the Gauss map, Pomeau–Manneville maps (also known as the intermittent map) and certain unimodal maps (see, for instance, [18]). For a more general construction in higher dimensions see [12].

9.4. A PROBABILISTIC APPLICATION. Let $X_n$, $n \geq 0$ be a Markov chain on a countable state space $\mathcal{A}$ with an invariant probability measure $\mu$. Suppose that $X$ is ergodic, aperiodic and satisfies the Doeblin condition: $\exists \Gamma \subset \mathcal{A}$ with $\mu(\Gamma) = 1$, $\exists \varepsilon \in (0,1)$, $\exists n \geq 1$ such that $\forall x \in \Gamma$, $\forall \Delta \subset \mathcal{A}$ with $\mu(\Delta) < \varepsilon$ the $n$-step transition probability satisfies

$$P(n, x, \Delta) = P\{X_n \in \Delta \mid X_0 = x\} \leq 1 - \varepsilon.$$

It is known (see, for instance, [7]) that this property is equivalent to $\phi$-mixing with an exponentially fast decaying $\phi(n)$ coefficient. For given sequences $\xi = (\xi_0, \xi_1, \ldots)$ and $\eta = (\eta_0, \eta_1, \ldots)$ with $\xi_i, \eta_i \in \mathcal{A}$ we set

$$A^\xi_n = \{\omega \in \mathcal{A}^\mathbb{N} : \omega_i = \xi_i, i = 0, 1, \ldots, n - 1\},$$

$$\tau^\eta_n(\omega) = \min\{k \geq 0 : (X_k, X_{k+1}, \ldots, X_{k+n-1}) = (\eta_0, \eta_1, \ldots, \eta_{n-1})\}$$

and

$$\Sigma^\xi,\eta_{n,n} = \sum_{k=0}^{\tau^\eta_n-1} \mathbb{I}_{A^\xi_n}(X_k, X_{k+1}, \ldots, X_{k+n-1}).$$

Then, assuming that $\xi, \eta$ are nonperiodic and not shifts of each other, we obtain from Corollary 2.2 convergence in distribution of $\Sigma^\xi,\eta_{n,n}$ to a geometric random variable with the parameter $(1 + \lambda)^{-1}$ provided

$$\frac{\mu(A^\xi_n)}{\mu(A^\eta_n)} \to \lambda \quad \text{as } n \to \infty.$$
References

[1] M. Abadi, *Exponential approximation for hitting times in mixing processes*, Mathematical Physics Electronic Journal 7 (2001), Article no. 2.

[2] M. Abadi and B. Saussol, *Hitting and returning into rare events for all alpha-mixing processes*, Stochastic Processes and their Applications 121 (2011), 314–323.

[3] M. Abadi and N. Vergne, *Sharp errors for point-wise Poisson approximations in mixing processes*, Nonlinearity 21 (2008), 2871–2885.

[4] R. Arratia, L. Goldstein and L. Gordon, *Two moments suffice for Poisson approximations: the Chen–Stein method*, Annals of Probability 17 (1989), 9–25.

[5] P. Billingsley, *Probability and Measure*, Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, New York, 1995.

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

[7] R. C. Bradley, *Introduction to Strong Mixing Conditions*, Kendrick Press, Heber City, UT, 2007.

[8] S. M. Buckley, *Is the maximal function of a Lipschitz function continuous?* Annales Academiae Scientiarum Fennicae Mathematica 24 (1999), 519–528.

[9] M. Demers, P. Wright and L.-S. Young, *Entropy, Lyapunov exponents and escape rates in open systems*, Ergodic Theory and Dynamical Systems 32 (2012), 1270–1301.

[10] S. Galatolo, J. Rousseau and B. Saussol, *Skew products, quantitative recurrence, shrinking targets and decay of correlations*, Ergodic Theory and Dynamical Systems 35 (2015), 1814–1845.

[11] A. Galves and B. Schmitt, *Inequalities for hitting times in mixing dynamical systems*, Random & Computational Dynamics 5 (1997), 337–348.

[12] S. Gouëzel, *Decay of correlations for nonuniformly expanding systems*, Bulletin de la Société Mathématique de France 134 (2006), 1–31.

[13] N. T. A. Haydn and Y. Psiloyenis, *Return times distribution for Markov towers with decay of correlations*, Nonlinearity 27 (2014), 1323–1349.

[14] N. Haydn and F. Yang, *Local Escape Rates for ϕ-mixing Dynamical Systems*, Ergodic Theory and Dynamical Systems 40 (2020), 2854–2860.

[15] L. Heinrich, *Mixing properties and central limit theorem for a class of non-identical piecewise monotonic C²-transformations*, Mathematische Nachrichten 181 (1996), 185–214.

[16] Yu. Kifer and A. Rapaport, *Poisson and compound Poisson approximations in conventional and nonconventional setups*, Probability Theory and Related Fields 160 (2014), 797–831.

[17] Yu. Kifer and A. Rapaport, *Geometric distribution for multiple returns until a hazard*, Nonlinearity 32 (2019), 1525–1545.

[18] I. Melbourne and M. Nicol, *Almost sure invariance principle for nonuniformly hyperbolic systems*, Communications in Mathematical Physics 260 (2005), 131–146.

[19] F. Pène and B. Saussol, *Poisson law for some non-uniformly hyperbolic dynamical systems with polynomial rate of mixing*, Ergodic Theory and Dynamical Systems 36 (2016), 2602–2626.
[20] Y. Pesin, *Dimension Theory in Dynamical Systems*, Chicago Lectures in Mathematics, University of Chicago Press, Chicago, IL, 1997.

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

[22] K. Petersen, *Ergodic Theory*, Cambridge Studies in Advanced Mathematics, Vol. 2, Cambridge University Press, Cambridge, 1983.

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

[24] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Annals of Mathematics 7 (1998), 585–650.

[25] L.-S. Young, *Recurrence time and rate of mixing*, Israel Journal of Mathematics 110 (1999), 153–188.

[26] R. Zweimüller, *Mixing limit theorems for ergodic transformations*, Journal of Theoretical Probability 20 (2007), 1059–1071.