On Coupling Lemma and Stochastic Properties with Unbounded Observables for 1-d Expanding Maps

Jianyu Chen∗1, Hong-Kun Zhang†2, and Yiwei Zhang‡3

1School of Mathematical Sciences & Center for Dynamical Systems and Differential Equations, Soochow University, Suzhou, Jiangsu 215006, P.R.China
2Department of Mathematics and Statistics, University of Massachusetts Amherst, Amherst, MA 01003, USA.
3School of Mathematics and Statistics, Center for Mathematical Sciences, Hubei Key Laboratory of Engineering Modeling and Scientific Computing, Hua-Zhong University of Sciences and Technology, Wuhan 430074, China.

Abstract

In this paper, we establish a coupling lemma for standard families in the setting of piecewise expanding interval maps with countably many branches. Our method merely requires that the expanding map satisfies Chernov’s one-step expansion at q-scale and eventually covers a magnet interval. Therefore, our approach is particularly powerful for maps whose inverse Jacobian has low regularity and those who does not satisfy the big image property. The main ingredients of our coupling method are two crucial lemmas: the growth lemma in terms of the characteristic Z function and the covering ratio lemma over the magnet interval. We first prove the existence of an absolutely continuous invariant measure. What is more important, we further show that the growth lemma enables the liftablity of the Lebesgue measure to the associated Hofbauer tower, and the resulting invariant measure on the tower admits a decomposition of Pesin-Sinai type. Furthermore, we obtain the exponential decay of correlations and the almost sure invariance principle (which is a functional version of the central limit theorem). For the first time, we are able to make a direct relation between the mixing rates and the Z function, see (2.7). The novelty of our results relies on establishing the regularity of invariant density, as well as verifying the stochastic properties for a large class of unbounded observables.

Finally, we verify our assumptions for several well known examples that were previously studied in the literature, and unify results to these examples in our framework.

Keywords: Coupling lemma, Standard families, Chernov’s one-step expansion at q-scale, Characteristic Z function, Growth lemma, Dynamically Hölder series.

Contents

1 Introduction 2
2 Assumptions and Main Results 4
   2.1 Assumptions. . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 4
   2.2 Standard pairs and standard families . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 5
   2.3 Statement of results . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
      2.3.1 Coupling lemma . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 6
      2.3.2 Absolutely continuous invariant measure . . . . . . . . . . . . . . . . . . . . . . . . . 7
      2.3.3 Stochastic properties . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 7

∗E-mail: jychen@suda.edu.cn.
†E-mail: hongkun@math.umass.edu.
‡E-mail: yiweizhang@hust.edu.cn.
1 Introduction

The probabilistic coupling method is a flexible technique to compare random processes with two different initial distributions. This method has been applied in a broad variety of contexts in modern probability theory, e.g., to prove limit theorems, to derive inequalities, or to obtain approximations. For a comprehensive introduction on the developments of this topic, we refer the readers to the books by Lindvall [35] and Thorisson [44]. In the field of dynamical systems, the coupling method is also powerful and has been developed since the celebrated work by Young [46] for Young towers, and later the systematic works by Chernov and Dolgopyat [14, 15, 24] for introducing standard pairs for chaotic billiards and partially hyperbolic systems. These two schemes have been adapted afterwards in various settings, e.g., [2, 3, 18, 36, 45, 47].

In this paper, we are aiming to adapt the works in [18, 45], and to establish a widely applicable version of coupling lemma for standard pairs in the setting of piecewise expanding interval maps with countably many inverse branches. Roughly speaking, our coupling lemma indicates if the dynamical system satisfies Chernov’s one-step expansion condition and eventually covers a magnet interval, then every two proper standard families can be coupled after iterations with an exponential decay for the tail of difference (see Theorem 2). The assumptions of our coupling lemma are purely geometrical and simple to check (see Assumption (H1)-(H3) in Section 2 for the precise statements). Moreover, these assumptions allow the systems under considerations to have lower regularity of the inverse Jacobian (see Assumption (H2)) and to merely satisfy a non-uniform version of “big image property”, which is beyond a large part of the current theory of Markov maps with infinitely many branches (see Assumption (H3)). Based on
this coupling lemma, several statistical properties are further investigated, including the existence of an absolutely continuous invariant probability measure (see Theorem 3), the regularity of the invariant density (see Theorem 1), the exponential decay of correlations (see Theorem 5) and the almost sure invariance principle (see Theorem 6), etc.

As the readers shall see later, our coupling lemma turns out to be completely independent of functional analysis for the transfer operator, and thus is effective on bypassing the difficulties on the construction of suitable Banach spaces. Note that functional analytic method is extremely powerful on proving the spectral properties of the transfer operators, and in particular, on establishing the regularity of the invariant density (See e.g. [1,20,33]). We stress that despite such analytic tool is in absence, we still manage to establish the regularity of the invariant density by a completely new approach. Another novelty that we would like to emphasize is that the observables include a large class of unbounded functions, for which the exponential decay of correlations and the almost sure invariance principle hold.

The main ingredients in our proof are described as follows. In order to obtain delicate estimates in our coupling algorithm, we adopt the notion of characteristic $Z$ function, which was first introduced in [18] (see also §7.4 in [16] for an alternative form of the characteristic $Z$ function), to measure the average length of standard families under the operations of cutting, iterates and splitting over the magnet. In particular, for the second operation, we establish the so-called growth lemma with exponential rate (see Lemma 3.6). This key lemma is due to our assumption (H1), and it guarantees that most of intervals in standard families will grow after sufficient many iterates. The other key lemma is the covering ratio lemma (see Lemma 4.2) over a given magnet by standard families, which results from our assumption (H3). It follows that a fixed portion of standard families is coupled at times with bounded gap.

To the best of our knowledge, it is also the first time to introduce the crucial assumption (H1) - Chernov’s one-step expansion at $q$-scale with the constant $q$ is allowed to be less than 1 in the setting of interval dynamics. The advantages of this assumption are two-fold. On the one hand, the coupling technique still works out, even though the inverse Jacobian of the expanding map may not be summable (see Section 8.1). On the other hand, we are able to introduce the space $\mathcal{H}_{W,\gamma,t}$ of dynamically Hölder series (see Definition 6), which contains a large class of unbounded functions when $t > 0$.

We stress that the Hofbauer tower construction is used in the proof of Theorem 4 which shows that the invariant density is a dynamically Hölder series. A important by-product is that we prove the Lebesgue measure is liftable to the Hofbauer tower, due to the second growth lemma (see Lemma 5.1). Inspired by the work [40] of Pesin and Sinai, we show that the limiting invariant measure on the Hofbauer tower has a decomposition of Pesin-Sinai type, and we further prove that the invariant measure on the unit interval is in fact carried by a standard family.

In the last section of this paper, we apply our results in the following two aspects. Firstly, by revisiting several well known piecewise linear expanding maps in the literature, e.g., [8,42], we provide a unified mechanism on the existence of absolutely continuous invariant probability measure, the regularity of the invariant density and some statistical properties for these examples (see Proposition 8.1). Indeed, compared to Theorem 1 in [8], our Assumption (H1) - Chernov’s one-step expansion at $q$-scale turns out to be rather sharp on guaranteeing the existence of an absolutely continuous invariant probability measure. Secondly, we investigate the function space of dynamically Hölder series for which the almost sure invariant principle (ASIP) holds. In particular, we are able to show the ASIP for the random process generated by certain unbounded observables over the doubling map, which gives a functional improvement of the central limit theorem in the previous studies (See e.g. [19]).

The paper is organized as follows. In Section 2 we introduce the general assumptions (H1) - (H3), as well as the notions of standard pairs and standard families, and then state the results on the coupling lemma and the consequent stochastic properties. In Section 3 we make some preparations on the quantitative behavior of the standard families under the dynamics. We then complete proofs of all the theorems in Section 4-7. Finally, in Section 8 we provide some examples and remarks, for which our assumptions and results apply.
2 Assumptions and Main Results

Let $M = [0, 1]$ be the unit interval endowed with the standard Euclidean metric, and let $m$ be the Lebesgue measure on $M$. Given a sub-interval $W \subset M$, we denote its length by $|W| = m(W)$, and the conditional measure of $m$ on $W$ by $m_W(\cdot) = m(\cdot | W)$.

We consider a one-dimensional map $T : M \to M$ with countably many inverse branches, that is, there is a countable partition $\xi_1$ of $M$ into sub-intervals, on each interior of which $T$ is strictly monotonic and $C^1$-smooth. Note that we do not require $W \in \xi_1$ to be a maximal inverse branch.

2.1 Assumptions.

In the following we list and briefly explain the assumptions.

Set $\xi_n = \xi_1 \vee T^{-1}\xi_1 \vee \cdots \vee T^{-(n-1)}\xi_1$ for any $n \geq 1$. Given an interval $W \subset M$, we let $\{W_\alpha\}_{\alpha \in W/\xi_n}$ be the collection of sub-intervals of $W$ after being cut by $S_n$. In other words, $\{W_\alpha\}_{\alpha \in W/\xi_n}$ is the relative partition of $W$ given by $\xi_n$. For each $\alpha \in W/\xi_n$, we call the interval $T^nW_\alpha$ a component of $T^nW$. We further denote the collection of components of $T^nW$ by $\{T^nW_\alpha\}_{\alpha \in W/\xi_n}$.

Although some intervals in $\xi_n$ may be relatively short, the following expansion condition ensures that a large portion of intervals in $\{T^nW_\alpha\}_{\alpha \in W/\xi_n}$ are relatively long.

(H1) Chernov’s one-step expansion. There exists $q \in (0, 1]$ such that

$$\liminf_{\delta \to 0} \sup_{W : |W| < \delta} \sum_{\alpha \in W/\xi_1} \left( \frac{|W|}{|TW_\alpha|} \right)^q |W_\alpha| = 1,$$

where the supremum is taken over all sub-intervals $W \subset M$.

Remark 1. Assumption (H1) was brought up by Chernov and Zhang in [17] for chaotic billiards with polynomial mixing rates (with $q = 1$), and later in [18][21][22] for two-dimensional general hyperbolic systems with singularities.

To emphasize the choice of $q$, we shall call (2.1) the (Chernov’s) one-step expansion (condition) at $q$-scale. Note that by Jensen’s inequality, the one-step expansion at $q'$-scale implies the one-step expansion at $q$-scale for any $0 < q \leq q' \leq 1$. In particular, the one-step expansion at 1-scale, i.e.,

$$\liminf_{\delta \to 0} \sup_{W : |W| < \delta} \sum_{\alpha \in W/\xi_1} \frac{|W_\alpha|}{|TW_\alpha|} < 1,$$

implies the one-step expansion at $q$-scale for any $q \in (0, 1)$. In Section 8.1, we shall provide a class of piecewise linear maps with infinitely many branches, for which the one-step expansion fails at 1-scale but holds at $q$-scale for some $q < 1$.

Another advantage of Chernov’s one-step expansion at $q$-scale with $q < 1$ is that the observables that we consider can be unbounded (see Definition 6 and Remark 4).

Let $S_n$ be the set of endpoints of intervals in $\xi_n$, and set $S_\infty = \bigcup_{n \geq 1} S_n$. It directly follows from Assumption (H1) that the map $T$ is uniformly expanding on $M \setminus S_1$. Therefore, $\{\xi_1\}$ is a generating partition under iterations of $T^{-n}$, or equivalently, $\xi_\infty := \bigvee_{k=0}^\infty T^{-k}\xi_1$ is the partition into individual points (mod $m$), which makes the separation time given below well-defined on $M \setminus S_\infty$.

Definition 1. Given a pair of points $x$ and $y$ in $M \setminus S_\infty$, the separation time $s(x, y)$ is defined to be the smallest integer $n \geq 1$ such that $x$ and $y$ belong to distinct elements of $\xi_n$.

To make assumptions on the regularity of Jacobian, we first introduce the dynamically Hölder continuous functions.
Definition 2. A function \( f : M \to \mathbb{R} \) is said to be \textit{dynamically Hölder continuous}, supported on an interval \( W \subset M \) with parameter \( \gamma \in (0, 1) \), if \( f|_{M \setminus W} \equiv 0 \) and

\[
|f|_{W,\gamma} := \sup \left\{ \frac{|f(x) - f(y)|}{\gamma^{|x-y|}} : x, y \in W \setminus S_\infty, \text{ and } x \neq y \right\} < \infty.
\]

We denote by \( \mathcal{H}_{W,\gamma} \) the space of such functions. Note that \( \mathcal{H}_{W,\gamma} \subset L^\infty(m) \), and we denote \( \|f\|_{W,\gamma} := \|f\| + |f|_{W,\gamma} \) for any \( f \in \mathcal{H}_{W,\gamma} \).

Denote by \( T' \) the derivative of \( T \), which is well defined on \( M \setminus S_1 \). We assume the following.

(H2) \textbf{Regularity of log Jacobian (with respect to \( \xi_1 \))}. There exist \( C_J > 0 \) and \( \gamma_J \in (0, 1) \) such that for any interval \( W \in \xi_1 \), the function \( 1_W \cdot \log |T'| \) belongs to \( \mathcal{H}_{W,\gamma_J} \) and \( |1_W \cdot \log |T'||_{W,\gamma_J} \leq C_J \).

Finally, since we do not have an invariant measure to begin with, we impose the following topological condition in order to establish the coupling lemma.

(H3) \textbf{Eventual covering}. There exists an interval \( U \), which is called a \textit{magnet}, such that any interval \( W \subset M \) will eventually cover \( U \) in the following sense: there is an integer \( n_W \geq 1 \) such that for any \( n \geq n_W \), at least one component of \( T^nW \) contains \( U \).

Remark 2. Our magnet interval is a topological analogy of the magnet rectangle in two-dimensional hyperbolic systems, see e.g. §7.12 in \cite{16}.

Assumption (H3) is easy to check when the map \( T \) admits a Markov partition, of which \( U \) is an element. In general, this assumption may be verified by studying the combinatorial structure of one-dimensional maps (see Section 5.1).

2.2 Standard pairs and standard families

To establish the coupling lemma for the one-dimensional maps, we introduce the concepts of standard pairs and standard families.

Let \( C_J > 0 \) and \( \gamma_J \in (0, 1) \) be constants given in Assumption (H2). Fix

\[
\gamma \in (\gamma_J, 1), \quad \text{and} \quad C_\gamma \geq \max\{1, 2C_J/(\gamma^{-1} - 1)\}.
\]

Definition 3 (Pair and standard pair). \((W, \nu)\) is called a \textit{pair} if \( W \) is an interval in \( M \) and \( \nu \) is an absolutely continuous probability measure supported on \( W \).

A pair \((W, \nu)\) is called a \textit{standard pair} if the density \( \rho := d\nu/dm \) is \textit{regular} on \( W \) in the sense that \( \log \rho \in \mathcal{H}_{W,\gamma} \) with the semi-norm \( |\log \rho|_{W,\gamma} \leq C_\gamma \).

In the coupling process, forward iterates of standard pairs require the definition of standard families, which can be viewed as a convex sum of standard pairs.

Definition 4 (Family and standard family). Let \( \mathcal{G} = \{(W_\alpha, \nu_\alpha), \alpha \in A, \lambda_\alpha\} \) be a countable family of pairs, endowed with non-negative weights \( \lambda_\alpha \) on the index set \( A \).

The total measure of a family \( \mathcal{G} \) is given by

\[
\nu_\mathcal{G}(A) = \sum_{\alpha \in A} \lambda_\alpha \nu_\alpha(A),
\]

for any Borel set \( A \subset M \). For simplicity, we also denote

\[
\mathcal{G} = \sum_{\alpha \in A} \lambda_\alpha (W_\alpha, \nu_\alpha) \quad \text{and} \quad \nu_\mathcal{G} = \sum_{\alpha \in A} \lambda_\alpha \nu_\alpha.
\]

A family \( \mathcal{G} \) is called a \textit{standard family} if each \((W_\alpha, \nu_\alpha)\) is a standard pair and \( \sum_{\alpha \in A} \lambda_\alpha = 1 \).
We denote $\mathcal{Q}$ the supremum of scales for which Assumption ($H_1$) holds, i.e.,

$$\mathcal{Q} := \sup \{ q \in (0, 1] : \text{the one-step expansion (2.1) holds at } q\text{-scale} \}. \quad (2.3)$$

From now on, we fix a scale $q_0 \in (0, \mathcal{Q})$. Then there exists $\delta_0 > 0$ such that

$$\theta_0 := \sup_{W : |W| < \delta_0} \sum_{a \in W/\xi} \left( \frac{|W|}{|TW_a|} \right)^{q_0} \frac{|W_a|}{|W|} < 1, \quad (2.4)$$

where the supremum is taken over all sub-intervals $W \subset M$.

The average length of intervals in a family $G = \sum_{a \in A} \lambda_a(W_a, \nu_a)$ is measured by the following characteristic $Z$ function

$$Z(G) := \sum_{a \in A} \lambda_a |W_a|^{-q_0}. \quad (2.5)$$

Note that $Z(G) \geq 1$ for any standard family $G$. Let $\mathcal{G}$ be the collection of all families $G$ with $Z(G) < \infty$.

We fix constants

$$c_0 := \max \left\{ 1, \frac{2 \delta_0 \delta_0^{q_0}}{1 - \theta_0} \right\}, \quad \text{and } C_p \geq 10 c_0 e^{7 C_r}. \quad (2.6)$$

**Definition 5.** A family $G$ is called proper if $Z(G) \leq C_p$. We say that two families $G_1$ and $G_2$ are equivalent if $\nu_{G_1} = \nu_{G_2}$, denoted by $G_1 \equiv G_2$. Further, we denote $G \equiv \sum_{n=1}^{\infty} G_n$ if $\nu_G = \sum_{n=1}^{\infty} \nu_{G_n}$.

### 2.3 Statement of results

In this paper, we always assume that the map $T : M \ni \cdot$ satisfies Assumptions ($H_1$)-($H_3$) given in Section 2.1.

**2.3.1 Coupling lemma**

With the preparations in Section 2.2, we are now ready to state our first main result - the coupling lemma over magnets.

**Theorem 1.** Given a magnet $U$, there exist $N_c \geq 1$ and $\Theta_c \in (0, 1)$ such that the total measure of any proper standard family $G$ can be decomposed as

$$\nu_G = \sum_{n=1}^{\infty} \nu_n,$$

where each $\nu_n$ is a non-negative finite measure on $M$. Moreover,

1. **Coupling:** If $n$ is an integer multiple of $N_c$, then $T_n^* \nu_n = \Theta_c m_U$; otherwise, $\nu_n$ is null.
2. **Exponential tail:** $\sum_{k \geq n} \nu_k(M) \leq (1 - \Theta_c)^{n/N_c}$.

**Remark 3.** Note that the choices of $\{\nu_n\}_{n \geq 1}$ are not unique in the coupling lemma. As we do not pursue the optimal values for the constants $N_c$ and $\Theta_c$, we shall construct a slow coupling process in the proof of Theorem 1.
2.3.2 Absolutely continuous invariant measure

The equidistribution property immediately follows from Theorem 1.

**Theorem 2.** For any two proper standard families $\mathcal{G}^1$ and $\mathcal{G}^2$ and any $n \geq 0$,

$$\|
T^n \nu_{\mathcal{G}^1} - T^n \nu_{\mathcal{G}^2}\|_{TV} \leq 2(1 - \Theta_c)n/N_c,$$

where $\| \cdot \|_{TV}$ denotes the total variation norm, and $\Theta_c, N_c$ are given by Theorem 1.

The existence of an absolutely continuous invariant measure is a direct consequence of Theorem 2. Furthermore, iterates of any standard family converge exponentially to such measure.

**Theorem 3.** There exists an absolutely continuous $T$-invariant probability measure $\mu$ on $M$. Moreover, there exist constants $C_\gamma > 0$ and $\vartheta_\gamma \in (0, 1)$, such that for any standard family $\mathcal{G} \in \mathfrak{G}$ and any $n \geq 0$,

$$\|T^n \nu_{\mathcal{G}} - \mu\|_{TV} \leq C_\gamma \vartheta_\gamma^n Z_n(\mathcal{G}).$$

(2.7)

The above Theorem establish a new relationship between the $Z$ function and the rates of mixing for initial measures associated to standard families. Equation (2.7) makes it a much clearer picture to understand that $Z$ function is the only factor that dominates the mixing rates for expanding maps.

In general, the invariant density $h = d\mu/dm \in L^1(m)$ could be unbounded when $T$ has infinitely many inverse branches. To describe such function, we introduce the space of dynamically Hölder series.

**Definition 6.** Let $\mathcal{W} := \{W_\alpha : \alpha \in \Lambda\}$ be a collection of countably many intervals in $M$. Choose $\gamma \in (0, 1)$ and $t \in [0, 1]$. A function $f : M \to \mathbb{R}$ is called a dynamically Hölder series supported on $\mathcal{W}$ with parameter $\gamma$ and power $t$, if $f = \sum_{\alpha \in \Lambda} f_\alpha$ such that each $f_\alpha \in \mathcal{H}_{W_\alpha, \gamma}$ and

$$\|f\|_{\mathcal{W}, \gamma, t} := \sum_{\alpha \in \Lambda} |W_\alpha|^t \|f_\alpha\|_{W_\alpha, \gamma} < \infty.$$

We denote by $\mathcal{H}_{\mathcal{W}, \gamma, t}$ the space of such functions.

**Remark 4.** It is easy to see that $\mathcal{H}_{\mathcal{W}, \gamma, t} \subset \mathcal{H}_{\mathcal{W}, \gamma, t'} \subset L^1(m)$ for any $0 \leq t \leq t' \leq 1$, and $\mathcal{H}_{\mathcal{W}, \gamma, 0} \subset L^\infty(m)$. In particular, if the collection $\mathcal{W} = \{W\}$, then the space $\mathcal{H}_{\mathcal{W}, \gamma, 0}$ coincides with the space $\mathcal{H}_{\gamma, \gamma}$, which consists of dynamically Hölder continuous functions supported on $W$ with parameter $\gamma$. Also, the space $\mathcal{H}_{\mathcal{W}, \gamma, t}$ contains unbounded functions if $t > 0$ and the collection $\mathcal{W}$ has intervals of arbitrary short length.

Let $\gamma$ be the constant given in $[2.2]$, and let $\bar{\gamma}$ be given in $[2.3].$

**Theorem 4.** There exists a collection $\mathcal{W}_h$ of countably many intervals such that the invariant density $h = d\mu/dm \in \mathcal{H}_{\mathcal{W}_h, \gamma, s}$ for any $s \in (1 - \bar{\gamma}, 1]$.

2.3.3 Stochastic properties

In the rest of this subsection, we let $\gamma$ and $\bar{\gamma}$ be given by $[2.2]$ and $[2.3]$ respectively. Also, let $\mu$ be the absolutely continuous invariant measure obtained in Theorem 4. We first show the system $(T, \mu)$ enjoys exponential decay of correlations for dynamically Hölder series against bounded observables.

**Theorem 5.** For any $t \in [0, \bar{\gamma})$, there are constants $C_t > 0$ and $\vartheta_t \in (0, 1)$ such that for any $f \in \mathcal{H}_{\mathcal{W}, \gamma, t}$ on some collection $\mathcal{W}$ of countably many intervals and for any $g \in L^\infty(m)$, we have

$$\left| \int fg \circ T^n \, d\mu - \int f \, d\mu \int g \, d\mu \right| \leq C_t \vartheta_t^n \|f\|_{\mathcal{W}, \gamma, t} \|g\|_{\infty}. $$

(2.8)
Note that (2.8) is automatic for any bounded dynamically Hölder continuous function \( f \in \mathcal{H}_{M,\gamma} \). In fact, for such bounded observables, we can show the exponential multiple decay of correlations, and thus prove the central limit theorem (CLT) by the “big small block technique” (see §7.6-7.8 in [16] for more details). Moreover, we can further establish a functional generalization of the CLT - the almost sure invariance principle (ASIP), which asserts that the stationary random process \( \{f \circ T^n\}_{n \geq 0} \) can be well approximated by a Brownian motion with an almost sure error. We refer the readers to the papers [14,20,27,34,37,38,43] for the ASIP of stationary process generated by bounded observables in various smooth dynamics.

However, when \( f \) is an unbounded observable, the CLT and ASIP may fail for some obvious reasons, for instance, \( f \not\in L^2(\mu) \) and thus the corresponding process \( \{f \circ T^n\}_{n \geq 0} \) has no finite variance. In order to establish the limiting theorems for such process, we need some to add some extra conditions, such as moment controls in [11,12]. In this paper, we impose the following conditions on the dynamically Hölder series \( f \in \mathcal{H}_{W,\gamma,t} \).

**Definition 7.** Recall that \( S_n \) is the set of endpoints of intervals in the partition \( \xi_n \). A collection \( W = \{W_\alpha : \alpha \in A\} \) is adapted if for any \( \alpha \in A \), there exists \( n(\alpha) \in \mathbb{N} \) such that the two endpoints of \( W_\alpha \) belong to \( S_{n(\alpha)} \).

A function \( f \in \mathcal{H}_{W,\gamma,t} \) is adapted if the collection \( W \) is adapted, and

\[
\|f\|_{W,\gamma,t}^\alpha := \sum_{\alpha \in A} \mu(W_\alpha)^t \|f_\alpha\|_{W,\gamma} < \infty. \tag{2.9}
\]

We denote by \( \mathcal{H}_{W,\gamma,t}^\alpha \) the space of functions satisfying (2.9).

Assume that \( t \in [0,\frac{1}{2}) \). We further say that \( f \in \mathcal{H}_{W,\gamma,t}^\alpha \) has fast tail if there is \( a > \max\left\{\frac{11}{2}, \frac{24+3t}{1-2t}\right\} \) such that

\[
\sum_{\alpha \in A : n(\alpha) \geq n} \|f_\alpha\|_{L^1(\mu)} = O\left(n^{-a}\right). \tag{2.10}
\]

**Remark 5.** Note that \( \mathcal{H}_{W,\gamma,t}^\alpha \in L^{1/t}(\mu) \subset L^2(\mu) \) for \( t \in [0,\frac{1}{2}) \). Also, it is automatic that a dynamically Hölder function \( f \in \mathcal{H}_{W,\gamma} \) is adapted and has fast tail if \( W = M \) or \( W \in \xi_n \) for some \( n \geq 1 \). As we shall see in the proof of Theorem 6 below, an adapted function \( f \in \mathcal{H}_{W,\gamma,t}^\alpha \) with fast tail can be well approximated by its conditional expectations with respect to the partition \( \xi_n \).

We denote \( \mathbb{E}(f) = \int fd\mu \) for any \( f \in L^1(\mu) \), and denote the covariance for \( f, g \in L^2(\mu) \) by \( \text{Cov}(f,g) := \mathbb{E}(fg) - \mathbb{E}(f)\mathbb{E}(g) \). Then the variance of \( f \in L^2(\mu) \) is given by \( \text{Var}(f) = \text{Cov}(f,f) \).

We now state the ASIP (and thus CLT) for the stationary process generated by an adapted observable.

**Theorem 6.** Fix any \( t \in [0,\frac{1}{2}) \). Let \( f \in \mathcal{H}_{W,\gamma,t}^\alpha \) be of fast tail, such that its auto-correlations satisfy that

\[
|\text{Cov}(f, f \circ T^n)| = O\left(n^{-\frac{11}{2}}\right). \tag{2.11}
\]

Then the stationary process \( \{f \circ T^n\}_{n \geq 0} \) satisfies the ASIP, that is, there exist a constant \( \lambda \in (0,\frac{1}{2}) \) and a Wiener process \( W(\cdot) \) such that

\[
\left|\sum_{k=0}^{n-1} f \circ T^k - n \mathbb{E}(f) - W(n\sigma_f^2)\right| = O(n^\lambda), \quad \text{a.s.}
\]

where \( \sigma_f^2 \) is given by the Green-Kubo formula, i.e.,

\[
\sigma_f^2 := \text{Var}(f) + 2 \sum_{n=1}^\infty \text{Cov}(f, f \circ T^n) \in [0,\infty). \tag{2.12}
\]
Remark 6. Condition (2.11) implies that the ASIP might hold for unbounded functions with fairly slow decay rates (in fact, polynomial decay) of auto-correlations. We remark that the exponent \(-\frac{16}{15}\) in (2.11) is due to a classical result on invariance principle by Philipp and Stout in [41] (see Proposition 7.1 in Section 7). Of course, we may improve this exponent by using some recent results on ASIP in probability theory, but we shall not pursue it in this paper. We shall provide an example in Section 8.2 on how to check Condition (2.11).

By Theorem 5, Condition (2.11) is automatic for any function \(f \in H^{ad}_{W,\gamma,0} \subset L^\infty (m)\).

3 Quantitative Estimates on Standard Families

In this section, we establish quantitative estimates on the density function and the average length (in terms of growth lemmas for the Characteristic \(Z\) functions) for a standard family under iterates. These estimates will be the basis for understanding our coupling algorithm afterwards.

3.1 Estimates for the density function on standard pairs

We first provide the bounds for the density function of a standard pair, that is:

**Lemma 3.1.** If \((W, \nu)\) is a standard pair with the density function \(\rho\), then

\[e^{-Cr} \leq \frac{\rho(x)}{|W|^{-1}} \leq e^{Cr}, \quad \text{for any } x \in W.\]

Moreover, for any \(x, y \in W\),

\[|\rho(x)^{\pm 1} - \rho(y)^{\pm 1}| \leq \frac{C_r e^{Cr}}{|W|} \gamma^{s(x,y)}.\]

**Proof.** By Definition 3 of a standard pair \((W, \nu)\), we have that for any \(x, y \in W\),

\[\rho(x) \leq e^{Cr} \gamma^{s(x,y)} \leq \rho(y) e^{Cr}.\]

Taking integral over \(W\) with respect to \(dm(y)\) on both sides, we obtain that \(|W| \rho(x) \leq e^{Cr}\). The proof for the other direction is similar.

Regarding the second assertion, for any \(z, w \in \mathbb{R}\) with \(|z|, |w| \leq C_r - \log |W|\),

\[|e^z - e^w| \leq |z - w| \sup_{|u| \leq C_r - \log |W|} |e^u| \leq e^{Cr} |z - w|,\]

and hence

\[|\rho(x)^{\pm 1} - \rho(y)^{\pm 1}| = |e^{\pm \log \rho(x)} - e^{\pm \log \rho(y)}| \leq \frac{e^{Cr}}{|W|} |\log \rho(x) - \log \rho(y)| \leq \frac{C_r e^{Cr}}{|W|} \gamma^{s(x,y)}.\]

This completes the proof of the lemma. 

The next lemma concerns the mergence of standard pairs over the same interval.

**Lemma 3.2.** Let \(\{(W, \nu_\alpha)\}_{\alpha \in A}\) be a countable collection of standard pairs. For any non-negative weights \(\lambda_\alpha\) on the index set \(A\) such that \(\sum_{\alpha \in A} \lambda_\alpha = 1\), then the mergence pair \((W, \nu)\) is also a standard pair, where \(\nu = \sum_{\alpha \in A} \lambda_\alpha \nu_\alpha\).
Proof. Let \( \rho_\alpha \) be the density of \( \nu_\alpha \), then the density of mergence pair is given by \( \rho = \sum_{\alpha \in A} \lambda_\alpha \rho_\alpha \). By Definition of standard pairs, for any \( x, y \in W \), we have
\[
e^{-C_r \gamma^{(x,y)}} \leq \frac{\rho_\alpha(x)}{\rho_\alpha(y)} \leq e^{C_r \gamma^{(x,y)}}
\] and thus
\[
e^{-C_r \gamma^{(x,y)}} \leq \frac{\rho(x)}{\rho(y)} = \sum_{\alpha \in A} \lambda_\alpha \rho_\alpha(x) \leq \sum_{\alpha \in A} \lambda_\alpha \rho_\alpha(y) \leq e^{C_r \gamma^{(x,y)}},
\] which immediately implies that \( |\log \rho|_{W, \gamma} \leq C_r \). So the mergence pair \((W, \nu)\) is a standard pair. \( \square \)

3.2 Iterates of standard families

Definition 8 (Iterates of families). For any integer \( n \geq 0 \) and any pair \((W, \nu)\), let \( \{W_\alpha\}_{\alpha \in W/\xi_n} \) be the relative partition of \( W \) given by \( \xi_n \), and set \( \nu_\alpha(\cdot) := \nu(T^{-n}(\cdot)|W_\alpha) \). We define
\[
T^n(W, \nu) = \sum_{\alpha \in W/\xi_n} \nu(W_\alpha) \cdot (T^nW_\alpha, \nu_\alpha).
\]
In general, for a family \( \mathcal{G} = \sum_{\beta \in A} \lambda_\beta(W_\beta, \nu_\beta) \), we define
\[
T^n \mathcal{G} = \sum_{\beta \in A} \lambda_\beta \cdot T^n(W_\beta, \nu_\beta).
\]

Lemma 3.3. If \( \mathcal{G} \) is a standard family, then \( T^n \mathcal{G} \) is also a standard family for any \( n \geq 1 \).

Proof. It suffices to show that for any standard pair \( \mathcal{G} = (W, \nu) \) with density \( \rho = \frac{d\nu}{dm} \), the first iterate
\[
T \mathcal{G} = \sum_{\alpha \in W/\xi_1} \nu(W_\alpha) \cdot (TW_\alpha, \nu_\alpha)
\]
is a standard family, where \( \{W_\alpha\}_{\alpha \in W/\xi_1} \) is the relative partition of \( W \) given by \( \xi_1 \), and \( \nu_\alpha(\cdot) = \nu(T^{-1}(\cdot)|W_\alpha) \).

It is clear that \( \sum_{\alpha \in W/\xi_1} \nu(W_\alpha) = \nu(W) = 1 \), and it remains to show that each \((TW_\alpha, \nu_\alpha)\) is a standard pair. Indeed, for any Borel subset \( A \subset TW_\alpha \),
\[
\nu_\alpha(A) = \frac{\nu(T^{-1}A|W_\alpha)}{m(A)} = \frac{1}{\nu(W_\alpha)} \frac{\nu(T^{-1}A \cap W_\alpha)}{m(T^{-1}A \cap W_\alpha)} \frac{m(T^{-1}A \cap W_\alpha)}{m(A)}.
\]
Since \( T|_{W_\alpha} : W_\alpha \to TW_\alpha \) is invertible, we denote \( x_\alpha = (T|_{W_\alpha})^{-1}(x) \) for any \( x \in TW_\alpha \). Then the density function \( \rho_\alpha := \frac{d\nu_\alpha}{dm} \) is given by
\[
\rho_\alpha(x) = \frac{1}{\nu(W_\alpha)} \frac{\rho(x_\alpha)}{|T'(x_\alpha)|}.
\]

For any \( x, y \in TW_\alpha \), by Assumption (H2) and the choice of \( \gamma \) and \( C_r \) given by (2.2),
\[
\left| \log \rho_\alpha(x) - \log \rho_\alpha(y) \right| \\
\leq \left| \log \rho(x_\alpha) - \log \rho(y_\alpha) \right| + \left| \log |T'(x_\alpha)| - \log |T'(y_\alpha)| \right| \\
\leq C_r \gamma^{s(x, y_\alpha)} + C_J \gamma^{s(x, y_\alpha)} \leq (C_r + C_J) \gamma^{s(x, y) + 1} \leq C_r \gamma^{s(x, y)}.
\]

Hence the density \( \rho_\alpha \) is regular on \( TW_\alpha \). This completes the proof of the lemma. \( \square \)

Remark 7. Along the same lines in the proof of Lemma 3.3, we can show that \( T \mathcal{G} \) is a standard family if the family \( \mathcal{G} = \sum_{\beta \in B} \lambda_\beta(W_\beta, \nu_\beta) \) is a convex sum of pairs with densities \( \rho_\beta = \frac{d\nu_\beta}{dm} \) satisfies that
\[
|\log \rho_\beta|_{W_\beta, \gamma} \leq \frac{1 + \gamma}{2\gamma} C_r.
\]
3.3 Cuttings of standard families

Definition 9 (Cut family). Let \((W, \nu)\) be a pair, and \(W\) is cut into countable sub-intervals \(\{W_i\}_{i \geq 1}\). The cut family of \((W, \nu)\) is defined as

\[
(W, \nu)' = \sum_{i = 1}^{\infty} \nu(W_i) \cdot (W_i, \nu(\cdot|W_i)).
\]

In general, let \(\mathcal{G} = \sum_{\alpha \in A} \lambda_{\alpha}(W_\alpha, \nu_\alpha)\) be a family. Given an index subset \(A' \subset A\) and a set \(C\) of countable points in \(M\), we define the cut family \(\mathcal{G}'\) from \(\mathcal{G}\) with only pairs in \(A'\) being cut by points in \(C\), that is,

\[
\mathcal{G}' = \sum_{\alpha \in A} \lambda_{\alpha}(W_\alpha, \nu_\alpha)'.
\]

We shall simply say \(\mathcal{G}'\) is a cut family from \(\mathcal{G}\) if there is no need to mention \(A'\) and \(C\).

It is easy to see that if \(\mathcal{G}\) is a standard family, then any cut family \(\mathcal{G}'\) from \(\mathcal{G}\) is also a standard family.

Lemma 3.4. Let \(\mathcal{G}'\) be a cut family from a standard family \(\mathcal{G}\) by \(k\) points,

\[
Z(\mathcal{G}) \leq Z(\mathcal{G}') \leq (k + 1)e^{Cr}Z(\mathcal{G}).
\]

Proof. It suffices to show for a standard pair \(\mathcal{G} = (W, \nu)\), which is cut into \((k+1)\) sub-intervals \(W_1, W_2, \ldots, W_{k+1}\). Then

\[
Z(\mathcal{G}') = \sum_{1 \leq i \leq k+1} \frac{\nu(W_i)}{|W_i|^{q_0}} \geq \frac{\sum_{1 \leq i \leq k+1} \nu(W_i)}{|W|^{q_0}} = \frac{1}{|W|^{q_0}} = Z(\mathcal{G}).
\]

On the other hand, by Lemma 3.1,

\[
\nu(W_i) = \int_{W_i} \rho(x)dm(x) \leq e^{Cr} \frac{|W_i|}{|W|},
\]

and thus,

\[
Z(\mathcal{G}') = \sum_{1 \leq i \leq k+1} \frac{\nu(W_i)}{|W_i|^{q_0}} \leq \sum_{1 \leq i \leq k+1} e^{Cr} \frac{|W_i|}{|W|^{q_0}} = \frac{e^{Cr}}{|W|^{q_0}} \sum_{1 \leq i \leq k+1} \left(\frac{|W_i|}{|W|}\right)^{1-q_0} \leq (k + 1)e^{Cr}Z(\mathcal{G}).
\]

This completes the proof of this lemma.

Remark 8. It is not hard to check that if a family \(\mathcal{G}\) is a convex sum of countably many families, say, \(\mathcal{G} = \sum_i \lambda_i \mathcal{G}_i\), then

\[
Z(T^n \mathcal{G}) = \sum_i \lambda_i Z(T^n \mathcal{G}_i), \quad \text{for any } n \geq 0.
\]

This together with Lemma 3.4 implies that if \(\mathcal{G}'\) is a cut family from a standard family \(\mathcal{G}\), then

\[
Z(T^n \mathcal{G}) \leq Z(T^n \mathcal{G}'), \quad \text{for any } n \geq 0.
\]
3.4 Growth lemmas

We establish the growth lemma in this section. Roughly speaking, it means the value of \( \mathcal{Z}(T^n S) \) decreases exponentially in \( n \) until it becomes small enough, providing that the initial standard family \( S \) belongs to \( \mathcal{G} \), i.e., \( \mathcal{Z}(S) < \infty \). This fundamental property was first introduced and proved by Chernov for dispersing billiards in [13], and later generalized by Chernov and Zhang in [18].

To begin with, we first state the growth lemma for the Lebesgue standard pairs.

**Lemma 3.5.** Let \( \theta_0, \delta_0 \) and \( q_0 \) be the constants given in (2.4). For any Lebesgue standard pair \((W, m_W)\) and any \( n \geq 1 \), we have

\[
\mathcal{Z}(T^n(W, m_W)) \leq \theta_0^n \mathcal{Z}((W, m_W)) + 2\delta_0^{-q_0}(\theta_0 + \cdots + \theta_0^{n-1}).
\]  

(3.5)

**Proof.** For any \( n \geq 0 \), we denote \( \{W_\alpha\}_{\alpha \in W/\xi_n} \) the relative partition of \( W \) given by \( \xi_n \), then

\[
T^n(W, m_W) = \sum_{\alpha \in W/\xi_n} m_W(W_\alpha) \cdot (T^n W_\alpha, T^n m_{W_\alpha}),
\]

and thus

\[
\mathcal{Z}(T^n(W, m_W)) = \sum_{\alpha \in W/\xi_n} m_W(W_\alpha) |T^n W_\alpha|^{-q_0} = \sum_{\alpha \in W/\xi_n} |W_\alpha| \frac{1}{|W| |T^n W_\alpha|^{q_0}}.
\]  

(3.6)

We now prove (3.5) by making induction on \( n \). When \( n = 1 \), if \(|W| < \delta_0 \), then by (2.4),

\[
\mathcal{Z}(T(W, m_W)) = \sum_{\alpha \in W/\xi_1} |W_\alpha| \frac{1}{|W| |T W_\alpha|^{q_0}} \leq \theta_0 |W|^{-q_0} = \theta_0 \mathcal{Z}((W, m_W)).
\]  

(3.7)

Otherwise, if \(|W| \geq \delta_0 \), we divide \((W, m_W)\) into \( k = \lfloor |W|/\delta_0 \rfloor + 1 \) pieces \( \{(W_1, m_{W_1}), \ldots, (W_k, m_{W_k})\} \) of equal length which belongs to \( \{\delta_0/2, \delta_0\} \). In other words, \((W, m_W)\) is cut into a sum of standard pairs \( \{(W_i, m_{W_i})\}_{1 \leq i \leq k} \) with equal weights \( 1/k \). By (3.4) and (3.7), we have

\[
\mathcal{Z}(T(W, m_W)) \leq \frac{1}{k} \sum_{i=1}^k \mathcal{Z}(T(W_i, m_{W_i})) \leq \frac{1}{k} \sum_{i=1}^k \theta_0 \mathcal{Z}((W_i, m_{W_i})) \leq \theta_0 \left( \frac{\delta_0}{2} \right)^{-q_0} \leq 2\theta_0 \delta_0^{-q_0}.
\]

In either case, we obtain (3.5) for \( n = 1 \).

Suppose now (3.5) holds for some \( n \). By (3.6),

\[
\mathcal{Z}(T^{n+1}(W, m_W)) = \sum_{\alpha \in W/\xi_{n+1}} \sum_{\beta \in W_{\alpha}/\xi_{n+1}} \frac{|W_{\alpha\beta}|}{|W|} \frac{1}{|T^{n+1} W_{\alpha\beta}|^{q_0}} = \sum_{\alpha \in W/\xi_1} \frac{|W_\alpha|}{|W|} \sum_{\beta \in W_\alpha/\xi_n} \frac{|W_{\alpha\beta}|}{|W_\alpha|} \frac{1}{|T^n (T W_{\alpha\beta})|^{q_0}}
\]

\[
= \sum_{\alpha \in W/\xi_n} \frac{|W_\alpha|}{|W|} \mathcal{Z}(T^n(W_\alpha, m_{W_\alpha}))
\]

12
Lemma 3.6. Let $c_0$ be given in (2.6). For any standard family $\mathcal{G} \in \mathcal{F}$ and any $n \geq 0$, 
\[ Z(T^n \mathcal{G}) \leq e^{2Cr} (Z(\mathcal{G}) \theta_0^n + c_0). \] (3.8)

Proof. By (3.3), it is enough to prove (3.8) for standard pairs. Let $(W, \nu)$ be a standard pair with the density $\rho$. For any $n \geq 0$, we denote $(W_\alpha)_{\alpha \in W/\xi_n}$ the relative partition of $W$ given by $\xi_n$, then the standard family $T^n(W, \nu)$ has weights $\nu(W_\alpha)$. We consider the corresponding Lebesgue standard pair $(W, m_W)$, then $T^n(W, m_W)$ has weights $m_W(W_\alpha)$. By Lemma 3.6,
\[ e^{-C_r} \leq \frac{\nu(W_\alpha)}{m_W(W_\alpha)} = \frac{\int_{W_\alpha} \rho \, dm}{\int_{W_\alpha} |W|^{-1} \, dm} \leq e^{C_r}, \] (3.9)
which implies that
\[ e^{-C_r} \leq \frac{Z(T^n(W, \nu))}{Z(T^n(W, m_W))} \leq e^{C_r}. \] (3.10)

By Lemma 3.6 and the definition of $c_0$ in (2.6), we have for any $n \geq 1$,
\[ Z(T^n(W, m_W)) \leq \theta_0^n Z((W, m_W)) + c_0, \] By (3.10), (3.8) holds for $\mathcal{G} = (W, \nu)$.

Remark 9. From the proofs of Lemma 3.1 and Lemma 3.6 we have that
\[ Z(T^n \mathcal{G}) \leq e^{4Cr} (Z(\mathcal{G}) \theta_0^n + c_0), \] if the family $\mathcal{G} = \sum_{\alpha \in A} \lambda_\alpha(W_\alpha, \nu_\alpha) \in \mathcal{F}$ is a convex sum of pairs with density $\rho_\alpha = d\nu_\alpha / dm$ satisfies that
\[ |\log \rho_\alpha|_{W, \mathcal{G}} \leq 2C_r. \] (3.11)

Lemma 3.7. For any standard family $\mathcal{G} \in \mathcal{F}$, $T^n \mathcal{G}$ is proper for any $n \geq n_p(\mathcal{G})$, where
\[ n_p(\mathcal{G}) := \left\lfloor \frac{-\log Z(\mathcal{G})}{\log \theta_0} \right\rfloor + 1. \]

Proof. By Lemma 3.6 and the definition of $C_p$ in (2.6), for any $n \geq n_p(\mathcal{G})$, 
\[ Z(T^n \mathcal{G}) \leq e^{2Cr} (Z(\mathcal{G}) \theta_0^n + c_0) \leq e^{2Cr}(1 + c_0) < C_p, \] and thus $T^n \mathcal{G}$ is proper for any $n \geq n_p(\mathcal{G})$.

We set
\[ n_p := \left\lfloor \frac{-\log C_p}{\log \theta_0} \right\rfloor + 1. \] (3.12)

If $\mathcal{G}$ is a proper standard family, then $n_p(\mathcal{G}) \leq n_p$, and hence $T^n \mathcal{G}$ is proper for all $n \geq n_p$. 

Therefore, (3.5) also holds for $(n+1)$. Thus, we complete the proof of this lemma by induction. \[\square\]
4 Proof of Theorem [1 - 3]

4.1 Proof of Theorem 1

Throughout the section, let us fix a magnet $U$ given by Assumption (H3). Theorem 1 will be proven by a coupling algorithm over $U$. Before we describe the algorithm, let us first introduce two crucial lemmas (Lemma 4.1 and 4.2) whose proofs are postponed to Appendix A.

4.1.1 Lemmas for standard families over the magnet

We first apply a special splitting of a standard family into two parts, one of which is Lebesgue over the magnet $U$. To be more precise, let $G = \sum_{\alpha \in A} \lambda_{\alpha}(W_{\alpha}, \nu_{\alpha})$ be a standard family. The split family from $G$ over the magnet $U$ with Lebesgue ratio $\rho \in (0, e^{-C_r})$ is defined as

$$\delta \cdot G + (\rho \delta) \cdot \hat{G}$$

where $\delta = \sum_{\alpha \in A} \lambda_{\alpha}$, the Lebesgue part $\hat{G}$ and the split part $\hat{G}$ are families given by

$$\hat{G} := \sum_{\alpha \in A, \lambda_{\alpha} \delta - 1} (W_{\alpha}, \nu_{\alpha} \cdot m_{W_{\alpha}}) = \sum_{\alpha \in A} \lambda_{\alpha}(U, m_U) \equiv (U, m_U)$$

and

$$\tilde{G} := \sum_{\alpha \in A} \frac{(1-\rho)\lambda_{\alpha}}{1-\rho \delta} (W_{\alpha}, \frac{\nu_{\alpha} - \rho m_{W_{\alpha}}}{1-\rho}) + \sum_{\alpha \in A \setminus \lambda_{\alpha}} \frac{\lambda_{\alpha}}{1-\rho \delta} (W_{\alpha}, \nu_{\alpha}).$$

(4.1)

With this convention, we have

**Lemma 4.1.** There is $\rho_c = \rho_c(U) \in (0, e^{-C_r})$ such that for any $\rho \in (0, \rho_c)$ and standard family $\mathcal{G}$ over the magnet $U$ with Lebesgue ratio $\mathcal{G}$, we denote by $\hat{G}$ the split part of $\mathcal{G}$ over the magnet $U$ with Lebesgue ratio $\rho$, then $T \hat{G}$ is a standard family, and

$$Z(T \hat{G}) \leq e^{4C_r} (Z(\mathcal{G}) + c_0).$$

Next, We define the covering ratio of a family $\mathcal{G} = \sum_{\alpha \in A} \lambda_{\alpha}(W_{\alpha}, \nu_{\alpha})$ over the magnet $U$ by

$$\delta(\mathcal{G}) = \sum_{\alpha \in A(U)} \lambda_{\alpha},$$

where $A(U) := \{ \alpha \in A : W_{\alpha} \text{ contains } U \}$. Note that following properties of $\delta(\cdot)$ are straightforward from the definition.

1. If a family $\mathcal{G}$ is a sum of countably many families, say, $\mathcal{G} = \sum_i \lambda_i \mathcal{G}_i$, then for any $n \geq 0$,

$$\delta(T^n \mathcal{G}) = \sum_i \lambda_i \delta(T^n \mathcal{G}_i).$$

(4.2)

2. By (3.9), for any standard pair $(W, \nu)$ and any $n \geq 0$,

$$\delta(T^n(W, \nu)) \geq e^{-C_r} \delta(T^n(W, m_W)).$$

(4.3)

It is clear the splitting operation preserves the total measure, and the Lebesgue part $\hat{G}$ is a standard family and $\hat{G} \equiv (U, m_U)$. Although the split part $\hat{G}$ is a convex sum of pairs, it might not be a standard family, since the pairs in the first summation of (4.1) may not have regular densities. Also, the average length of $\hat{G}$ could become shorter than that of $\mathcal{G}$. 

14
(3) If $\mathcal{G}'$ is a cut family from a family $\mathcal{G}$, then for any $n \geq 0$,

$$\delta(T^n \mathcal{G}) \geq \delta(T^n \mathcal{G}').$$

(4) If $\mathcal{G}$ is a standard family, and $\mathcal{G}'$ is the cut family from $\mathcal{G}$ with pairs in $\mathcal{A}(U)$ being cut by the two endpoints of $U$, then by Lemma 3.1

$$\delta(\mathcal{G}') \geq e^{-Cr}|U|\delta(\mathcal{G}).$$

Based on these properties, we have the following quantitative estimation on $\delta(\cdot)$.

**Lemma 4.2.** There are $n_c = n_c(U) \geq n_p$ and $d_c = d_c(U) \in (0, 1)$ such that for any proper standard family $\mathcal{G}$, we have $\delta(T^{n_c} \mathcal{G}) \geq d_c$.

### 4.1.2 The coupling algorithm for Theorem 1

We are now ready to describe our coupling algorithm. Fix a magnet $U$ given by Assumption (H3). Let $\overline{\rho}_c \in (0, e^{-Cr})$ be given by Lemma 4.2, and let $n_c \geq n_p$, $d_c \in (0, 1)$ be given by Lemma 4.2. Set

$$\Theta_c := e^{-Cr}|U|d_c \overline{\rho}_c.$$

Given a proper standard family $\mathcal{G}$, we set $\widehat{\mathcal{G}}_0 = \mathcal{G}$ and $\widehat{\mathcal{G}}_0 = T^{1+n_p} \widehat{\mathcal{G}}_0$. By (3.12), $\widehat{\mathcal{G}}_0$ is still a proper standard family. Starting from $\widehat{\mathcal{G}}_0$, we apply the following inductive procedure. Assume that a proper standard family $\mathcal{G}_k$ is defined, we shall obtain $\mathcal{G}_{k+1}$ and $\mathcal{G}_{k+1}$ as follows:

1. **Iteration:** By Lemma 4.2, $\delta(T^{n_c} \mathcal{G}_k) \geq d_c$. Also, by (3.12), $T^{n_c} \mathcal{G}_k$ is a proper standard family.

2. **Cutting:** Let $\mathcal{G}'_{k+1}$ be the cut family from $T^{n_c} \mathcal{G}_k$ with pairs that contains $U$ being cut by the two endpoints of $U$. By Inequality (4.3), $\delta(\mathcal{G}'_{k+1}) \geq e^{-Cr} |U|d_c = \Theta_c / \overline{\rho}_c$. By Lemma 3.4, $\mathcal{Z}(\mathcal{G}'_{k+1}) \leq 3e^{Cr} \mathcal{Z}(T^{n_c} \mathcal{G}_k) \leq 3e^{Cr} C_p$.

3. **Splitting:** Set $\overline{\mathcal{G}}_{k+1} := \Theta_c / \delta(\mathcal{G}'_{k+1})$. We split $\mathcal{G}'_{k+1}$ over the magnet $U$ with Lebesgue ratio $\overline{\mathcal{G}}_{k+1}$, and obtain

$$\mathcal{G}'_{k+1} = \Theta_c \overline{\mathcal{G}}_{k+1} + (1 - \Theta_c) \widehat{\mathcal{G}}_{k+1},$$

where $\overline{\mathcal{G}}_{k+1}$ is the Lebesgue part and $\widehat{\mathcal{G}}_{k+1}$ is the split part. By Lemma 4.1, $T^{n_c} \mathcal{G}_{k+1}$ is a standard family, and

$$\mathcal{Z}(T^{n_c} \mathcal{G}_{k+1}) \leq e^{4Cr} (\mathcal{Z}(\mathcal{G}'_{k+1}) + c_0) \leq 3e^{5Cr} (C_p + c_0).$$

By Lemma 3.6 Equations (2.8) and (3.12),

$$\mathcal{Z}(T^{1+n_p} \mathcal{G}_{k+1}) \leq e^{2Cr} (\mathcal{Z}(T^{n_c} \mathcal{G}_{k+1}) + c_0) \leq e^{2Cr} (3e^{5Cr} (C_p + c_0) + c_0) \leq e^{2Cr} (3e^{5Cr} (1 + c_0) + c_0) \leq 7c_0 e^{7Cr} < C_p.$$

Therefore, $\mathcal{G}_{k+1} := T^{1+n_p} \mathcal{G}_{k+1}$ is a proper standard family.

Set $N_c := (1 + n_p + n_c)$. At the $k$-th step of the above coupling construction, the Lebesgue part $\overline{\mathcal{G}}_k = \sum_{\alpha \in \mathcal{A}_k} \lambda_\alpha(U, m_U)$ has the following property: the index set $\mathcal{A}_k \subset M/\xi_k N_c$, and there is an interval $W_\alpha$ inside some element of $\xi_k N_c$ such that $T^{kN_c} W_\alpha = U$. In particular, $T^{kN_c}$ is invertible on $W_\alpha$. Then we can define the family

$$T^{-kN_c} \mathcal{G}_k := \sum_\alpha T^{-kN_c} \lambda_\alpha(W_\alpha, [(T^{kN_c} W_\alpha)^{-1}], m_U).$$
Let \( \nu_n := \nu_{G_n} \). By (4.6), we have for any \( k \geq 1 \),

\[
T^{kN} G = T^{kN} \tilde{G}_0 = T^{(k-1)N} \left( \Theta_{e} \tilde{G}_1 + (1 - \Theta_{e}) \tilde{G}_1 \right) = \Theta_{e} T^{(k-1)N} \tilde{G}_1 + (1 - \Theta_{e}) T^{(k-1)N} \tilde{G}_1 = T^{kN} \tilde{G}_N + \Theta_{e} (1 - \Theta_{e}) T^{(k-2)N} \tilde{G}_2 + (1 - \Theta_{e}) T^{(k-2)N} \tilde{G}_2 = \ldots = T^{kN} \left( \sum_{i=1}^{k} G_{iN} \right) + (1 - \Theta_{e})^k \tilde{G}_k.
\]

It is obvious that \( \nu_0 - \sum_{i=1}^{k} \nu_{G_{iN}} \) is a non-negative measure, and thus,

\[
\left\| \nu_0 - \sum_{i=1}^{k} \nu_{G_{iN}} \right\|_{TV} = \left( \nu_0 - \sum_{i=1}^{k} \nu_{G_{iN}} \right)(M) = T^{kN} \left( \nu_0 - \sum_{i=1}^{k} \nu_{G_{iN}} \right) (T^{kN} M) \leq (1 - \Theta_{e})^k,
\]

which implies that

\[
\nu_0 = \sum_{k=1}^{\infty} \nu_{G_{kN}} = \sum_{n=1}^{\infty} \nu_n.
\]

This provides the decomposition of \( \nu_0 \) in Theorem 1. Moreover, the exponential tail bound in Statement (2) directly follows from (4.7). Therefore, the proof of Theorem 1 is complete.

### 4.2 Proof of Theorem 2

Let \( \mathcal{G}^1 \) and \( \mathcal{G}^2 \) be two proper standard families. By Theorem 1, we decompose their total measures as \( \nu_{G^1} = \sum_{k=1}^{\infty} \nu_{G^1_k} \) and \( \nu_{G^2} = \sum_{k=1}^{\infty} \nu_{G^2_k} \). Therefore,

\[
\left\| T^n \nu_{G^1} - T^n \nu_{G^2} \right\|_{TV} \leq \left\| T^{n-k} \left( T^k \nu_{G^1_k} - T^k \nu_{G^2_k} \right) \right\|_{TV} + \left\| T^n \sum_{k>n} \nu_{G^1_k} \right\|_{TV} + \left\| T^n \sum_{k>n} \nu_{G^2_k} \right\|_{TV} 
\leq \left\| \sum_{k>n} \nu_{G^1_k} \right\|_{TV} + \left\| \sum_{k>n} \nu_{G^2_k} \right\|_{TV} \leq 2(1 - \Theta_{e})^{n/N}.
\]

This completes the proof of Theorem 2.
4.3 Proof of Theorem 3
Let $\theta_0$ be given in (2.4), and set
\[
\vartheta_c := \max \{\theta_0, (1 - \Theta_\varepsilon)^{1/N_\varepsilon}\}, \quad \text{and} \quad C_\varepsilon = \frac{2}{\theta_0(1 - \vartheta_c)}.
\]
We first show that there is a probability measure $\mu$ on $M$ such that $T^n_\varepsilon \nu_\varepsilon = \nu_{T^n_\varepsilon}$ converges to $\mu$ in the total variation norm for any standard family $\mathcal{G} \in \mathfrak{G}$. By Lemma 3.7, $T^n_\varepsilon \mathcal{G}$ is a proper standard family for any $n \geq n_\varepsilon(\mathcal{G})$, and note that $\theta_0 \leq \varepsilon(\mathcal{G}) \theta_0^{\nu^0(\mathcal{G})} \leq 1$. Apply Theorem 2 to the proper standard families $\mathcal{G}^1 = T^{\nu^0(\mathcal{G})+1} \mathcal{G}$ and $\mathcal{G}^2 = T^{\nu^0(\mathcal{G})+1} \mathcal{G}$, we get
\[
\|T^{n+1}_\varepsilon \nu_\varepsilon - T^n_\varepsilon \nu_\varepsilon\|_{TV} \leq \begin{cases} 2, & n < n_\varepsilon(\mathcal{G}), \\ 2\vartheta_c^{n-n_\varepsilon(\mathcal{G})}, & n \geq n_\varepsilon(\mathcal{G}) \end{cases}
\leq 2\vartheta_c^{n-n_\varepsilon(\mathcal{G})} \leq 2\vartheta_c^{n-n_\varepsilon(\mathcal{G})} \leq 2\vartheta_c^{n-n_\varepsilon(\mathcal{G})} \leq 2\vartheta_c^{n-n_\varepsilon(\mathcal{G})}.
\]
It follows that $T^n_\varepsilon \nu_\varepsilon$ is a Cauchy sequence in the total variation norm, and hence it converges to some probability measure $\mu$, such that
\[
\|T^n_\varepsilon \nu_\varepsilon - \mu\|_{TV} \leq \sum_{k=n}^{\infty} 2\vartheta_c^{k} \varepsilon(\mathcal{G}) = C_\varepsilon \varepsilon(\mathcal{G}).
\]
Given another standard family $\mathcal{G}' \in \mathfrak{G}$, and applying Theorem 2 to $\mathcal{G}^1 = T^{\nu^0(\mathcal{G})} \mathcal{G}$ and $\mathcal{G}^2 = T^{\nu^0(\mathcal{G})} \mathcal{G}'$, we get
\[
\|T^n_\varepsilon \nu_\varepsilon - T^n_\varepsilon \nu_{\mathcal{G}'}\|_{TV} \leq 2\vartheta_c^{n-n_\varepsilon(\mathcal{G})} \leq 2\vartheta_c^{n-n_\varepsilon(\mathcal{G})} \leq 2\vartheta_c^{n-n_\varepsilon(\mathcal{G})} \leq 2\vartheta_c^{n-n_\varepsilon(\mathcal{G})}.
\]
It is obvious that $\mu$ is $T$-invariant. It remains to show that $\mu$ is absolutely continuous, that is, $m(A) > 0$ for any Borel subset $A \subset M$ with $\mu(A) > 0$. To see this, we consider the Lebesgue standard pair $\mathcal{G}_0 = (M, m)$, then there is a large $n \geq 1$ such that $\|T^n_\varepsilon m - \mu\|_{TV} \leq 0.5 \mu(A)$, and thus $m(T^{-n} A) \geq 0.5 \mu(A) > 0$. Since $T$ is non-singular with respect to $m$, we must have $m(A) > 0$.

5 Proof of Theorem 4
To prove Theorem 4 we need the following preparations.

5.1 Second growth lemma
We recall an alternative definition of the characteristic $\varepsilon$ function (see Section 5 in [18] or §7.4 in [16] with $q_0 = 1$). Given an interval $W \subset M$ and a point $x \in W$, we denote $r_W(x) := \text{dist}(x, \partial W)$, that is, the Euclidean distance from $x$ to the closest endpoint of $W$. Further, given a family $\mathcal{G} = \sum_{\alpha \in A} \lambda_\alpha(W_\alpha, \nu_\alpha)$ and a point $x \in W_\alpha$, we shall denote $r_\alpha(x) = r_{W_\alpha}(x)$ if the choice of $\alpha$ is clear. We then denote
\[
\tilde{\varepsilon}(\mathcal{G}) := \sup_{\varepsilon > 0} \nu_\varepsilon \left( r_\varepsilon < \varepsilon \right) \quad \varepsilon > 0 \sum_{\alpha \in A} \lambda_\alpha \nu_\alpha \left\{ x \in W_\alpha : r_{W_\alpha}(x) < \varepsilon \right\}.
\]
Using the fact that $m(r_\varepsilon < \varepsilon) = \min\{2\varepsilon, |W|\}$ and Lemma 5.1, it is easy to show that $\tilde{\varepsilon}(\mathcal{G}) \leq 2e^{2C_\varepsilon} \varepsilon(\mathcal{G})$ for any standard family $\mathcal{G}$.

The growth lemma that we establish in Lemma 3.6 is usually called the first growth lemma, which immediately implies the following second growth lemma.
Lemma 5.1. For any \( \varepsilon > 0 \) and any standard pair \( \mathcal{G} = (W, \nu) \), we have

\[
\nu(r_{W,n}(x) < \varepsilon) := \nu\{x \in W : r_{T^n}(T^n x) < \varepsilon\} < C_p \varepsilon^q_0
\]

for all \( n > q_0 \log \theta_0 |W| \), where \( q_0, \theta_0 \) are given in (2.6) and \( C_p \) is given in (2.6).

Proof. By Lemma 3.6 and the choice of constants in (2.6), for any \( m \) and any \( n > q_0 \log \theta_0 |W| \), we have

\[
\tilde{Z}(T^n \mathcal{G}) \leq 2e^{2C r} \tilde{Z}(T^n \mathcal{G}) \leq 2e^{4C r} (\theta_0^n / |W|^{q_0} + c_0) \leq 2e^{4C r}(1 + c_0) < C_p.
\]

In other words, (5.1) holds for any \( \varepsilon > 0 \).

Lemma 5.1 is a slight generalization of the second growth lemma in §5.9 of [18], in which \( q_0 = 1 \) and \( \mathcal{G} \) is restricted to a normalized Lebesgue standard pair. To avoid confusion, we point out that we use the notation \( m_W(\cdot) \) to represent the normalized Lebesgue measure on \( W \) in this paper, while \( m_W(\cdot) \) is the unnormalized one in [18].

5.2 Hofbauer tower and liftability

In order to show that the invariant density \( h = d\mu / dm \) is a dynamically Hölder series, we first need to construct the corresponding collection \( \mathcal{W}_h \) of supporting intervals. To this end, we introduce a Markov extension over the system \( (M, T, \xi_1) \) which is nowadays called Hofbauer tower. For references on this subject, see [4, 6, 7, 8, 9, 10, 23, 29, 32, 39], etc.

For our purpose, we construct the Hofbauer tower as follows: we set \( D_0 := \{M\} \) and for \( n \geq 1 \),

\[
D_n := \{T(W \cap V) : W \in \xi_1 \text{ and } V \in D_{n-1}\}.
\]

It is not hard to see that \( D_n = \{T^n W_\alpha : \alpha \in M/\xi_n\} \), that is, \( D_n \) is the collection of components of \( T^n M \). We further set \( D = \cup_{n \geq 0} D_n \), which is a collection of countably many intervals. The Hofbauer tower extension over \( (M, T, \xi_1) \) is the triple \( (\hat{M}, \hat{T}, \hat{\xi}) \) where

1. the tower is given by \( \hat{M} := \{(x, D) : x \in M \times D : x \in \overline{D}\} \);
2. the map \( \hat{T} : \hat{M} \setminus \pi^{-1}(S_1) \rightarrow \hat{M} \) is given by \( \hat{T}(x, D) = (T(x), T(D \cap W(x))) \), where \( W(x) \) is the interval in \( \xi_1 \) containing \( x \) and \( \pi : \hat{M} \rightarrow M \) is the canonical projection, i.e., \( \pi(x, D) = x \);
3. the partition of \( \hat{M} \) is given by \( \hat{\xi} := \{\hat{D} : D \in D\} \), where for any interval \( D \in D \), we set \( \hat{D} := \{(x, D) : x \in \overline{D}\} \), which is an identical copy of \( D \).

It is easy to see that \( \hat{\xi} \) is a Markov partition for \( \hat{T} \). Also, \( \hat{T} \) is an extension of \( T \) via the projection \( \pi \), i.e., \( \pi \circ \hat{T} = T \circ \pi \). By extending the Euclidean metric of the unit interval \( M \) to the tower \( \hat{M} \) in a natural way, we have that \( \hat{M} \) is a complete separable metric space, which is not necessarily to be compact unless the map \( T \) is already Markov. For any \( D \in D \), we define the level of \( D \) as

\[
\ell(D) := \min\{n \geq 0 : D \in D_n\}.
\]

Further, for any \( \hat{x} = (x, D) \in \hat{M} \), we define the level of \( \hat{x} \) as \( \ell(\hat{x}) = \ell(D) \). Then we set the \( n \)-level set of \( \hat{M} \) to be \( \hat{M}_n := \{\hat{x} \in \hat{M} : \ell(\hat{x}) = n\} \). In particular, we call \( \hat{M}_0 \) the base of the tower \( \hat{M} \), which is an identical copy of \( M \).

We now discuss the liftability property of the Lebesgue measure. Let \( \mathcal{B} \) be the Borel \( \sigma \)-algebra of \( M \), then by extension, \( \mathcal{B} := \xi \vee \pi^{-1} \mathcal{B} \) is the Borel \( \sigma \)-algebra of \( \hat{M} \). We then extend the normalized Lebesgue
measure \( m \) on \( M \) to a (possibly infinite) measure \( \overline{m} \) on \( \hat{M} \) by setting \( \overline{m}(A) = \sum_{D \in \mathcal{D}} m \left( \pi \left( A \cap \hat{D} \right) \right) \) for any \( A \in \hat{\mathcal{B}} \). Define a sequence of measures on \( \hat{M} \) by
\[
\overline{m}_n(A) = m \left( \hat{T}^{-n} A \cap \hat{M}_0 \right), \quad \text{for any} \ n \geq 0.
\]
(5.2)

Note that \( \overline{m}_n \) are all probability measures and \( \pi_* \overline{m}_n = T^n m \), that is, \( \overline{m}_n \) projects to \( T^n m \), or equivalently, we say that \( T^n m \) is lifted to \( \overline{m}_n \). Similarly, we denote the Cesaro means of \( \overline{m}_n \) by \( \hat{m}_n \), that is,
\[
\hat{m}_n := \frac{1}{n} \sum_{k=0}^{n-1} \overline{m}_k, \quad \text{for any} \ n \geq 1.
\]

Note that \( \hat{m}_n \) projects to \( \frac{1}{n} \sum_{k=0}^{n-1} T^k m \). We say that \( \{ \hat{m}_n \}_{n \geq 0} \) is liftable if \( \hat{m}_n \) has a subsequence which converges weak star to a non-vanishing, in fact, probability measure on \( \hat{M} \). To show the liftability, we will prove that

**Lemma 5.2.** The sequence of measures \( \hat{m}_n \) is tight, i.e., for any \( \delta > 0 \), there exists a compact subset \( F \subset \hat{M} \) such that \( \hat{m}_n(\hat{M} \setminus F) < \delta \) for all \( n \).

**Proof.** It suffices to show that \( \overline{m}_n \) is tight. Choose \( \varepsilon_0 > 0 \) such that \( C_p e_{\varepsilon_0}^q < \delta/2 \), where \( q_0 \) and \( C_p \) are given by (2.4) and (2.6) respectively. Since \( \xi_1 \) is a generating partition, we can choose \( L \in \mathbb{N} \) such that \( \xi_L = \bigvee_{k=0}^{L-1} T^{-k} \xi_1 \) has diameter smaller than \( \varepsilon_0 \). Furthermore, we may assume \( C_c \theta_{L} < \delta/4 \), where \( C_c \) and \( \theta_{L} \) are the constants given by Theorem 3. We then set
\[
E := \left\{ \hat{x} \in \hat{M} : \ell(\hat{x}) \leq L \right\}.
\]

By the definition of \( \overline{m}_n \) in (5.2), it is easy to see that when \( n \leq L \), the measure \( \overline{m}_n \) is supported on \( E \) and thus \( \overline{m}_n(\hat{M} \setminus E) = 0 \). When \( n > L \), we consider the Lebesgue standard pair \( g_0 = (M, m) \), and we denote \( r_{M,k}(x) := r_{T^k g_0}(T^k x) \) for any \( k \geq 0 \). For any \( \hat{x} = (x, M) \in \hat{M}_0 \), if \( r_{M,n-L}(x) \geq \varepsilon_0 \), i.e., \( \text{dist}(T^{n-L} x, \partial D_{n-L}) \geq \varepsilon_0 \), where we denote \( \hat{T}^{n-L}(\hat{x}) = (T^{n-L} x, D_{n-L}) \), then there is \( \alpha \in \xi_L \) such that \( T^{n-L} x \in W_{\alpha} \) and \( D_{n-L} \) fully contains \( W_{\alpha} \). It follows that \( \hat{T}^{n-L}(\hat{x}) \in E \). By Lemma 5.1 and note that \( \log_\theta_{\varepsilon_0} |M| = 0 \), we have
\[
\overline{m}_n(\hat{M} \setminus E) = m \left( M \setminus \pi \left( \hat{T}^{-n} E \cap \hat{M}_0 \right) \right) \leq m \left( r_{M,n-L}(x) < \varepsilon_0 \right) \leq C_p e_{\varepsilon_0}^q < \delta/2.
\]

Now we construct a compact subset \( F \) of \( E \) as follows. Note that \( E \) can be rewritten as the following disjoint union \( E = \bigcup_{k=0}^{L} E_k \), where each
\[
E_k := \left\{ \hat{x} \in \hat{M} : \ell(\hat{x}) = k \right\}
\]
consists of countably many intervals. For each \( k \in [0, L] \), we can pick a subset \( F_k \subset E_k \) such that \( F_k \) is a union of finitely many intervals and
\[
\sigma \left( \pi (E_k \setminus F_k) \right) < \frac{\delta}{8L}, \quad \text{for measures} \ \sigma = \mu, m, T_m, \ldots, T_{L} m.
\]
Here \( \mu \) is the invariant measure that we obtain in Theorem 3. It is clear that \( F = \bigcup_{k=0}^{L} F_k \) is a compact
subset of $E$. Moreover, by Theorem 3,

$$m_n(E \setminus F) = T_n^\ast (\pi(E \setminus F)) \leq \sum_{k=0}^L T_k^\ast m((E_k \setminus F_k)) \leq \left\{ \begin{array}{ll}
(L + 1) \cdot \frac{\delta}{2L}, & \text{if } 0 \leq n \leq L,
(L + 1) \cdot \frac{\delta}{2L} + C_L \theta_L, & \text{if } n > L.
\end{array} \right.$$ 

Therefore, we have $m_n(\hat{M} \setminus F) \leq m_n(\hat{M} \setminus E) + m_n(E \setminus F) < \delta$. Hence $m_n$ is tight, so is $\hat{m}_n$. \hfill \Box

Recall that $\mu$ is the invariant measure that we obtain in Theorem 3. The following is a direct consequence of Lemma 5.2.

**Lemma 5.3.** $\hat{m}_n$ has a subsequence converging weak star to a probability measure $\hat{\mu}$ on $\hat{M}$ such that $\pi \ast \hat{\mu} = \mu$.

**Proof.** By Helly-Prohorov theorem, Lemma 5.2 implies that there is an increasing sequence of natural numbers $\{n_j\}_{j \geq 1}$ such that $\hat{m}_{n_j}$ converges weak star to a probability measure $\hat{\mu}$ on $\hat{M}$. Applying Theorem 3 to the Lebesgue standard pair $\hat{S}_0 = (\hat{M}, m)$, we have that $T_n^\ast m$ converges to $\mu$ in total variation, and hence in the weak star topology as well. Since $\pi \ast \hat{m}_n = \frac{1}{n} \sum_{k=0}^{n-1} T_k^\ast m$, we get

$$\pi \ast \hat{\mu} = \lim_{j \to \infty} \pi \ast \hat{m}_{n_j} = \lim_{j \to \infty} \frac{1}{n_j} \sum_{k=0}^{n_j-1} T_k^\ast m = \mu.$$ 

Here the above limits are taken in the weak star topology. \hfill \Box

### 5.3 Pesin-Sinai decomposition

In this section, we would like to show that the invariant measure $\mu$ on $\hat{M}$ is the total measure of a standard family. By Lemma 5.3, we shall instead show that the lifted measure $\hat{\mu}$ on $\hat{M}$ has the following structure.

**Definition 10.** A probability measure $\hat{\nu}$ on $\hat{M}$ is said to have **Pesin-Sinai decomposition** if the conditional decomposition of $\hat{\nu}$ with respect to the countable partition $\hat{\xi} = \{\hat{D}\}_{\hat{D} \in \hat{D}}$ has the following form:

$$\hat{\nu}(A) = \sum_{\hat{D} \in \hat{D}} \lambda(\hat{D}) \cdot \hat{\nu}_D(A)$$

for any $A \in \hat{\mathcal{B}}$, where

1. $\{\lambda(\hat{D})\}_{\hat{D} \in \hat{D}}$ is a probability vector on $\hat{D}$, that is, $0 \leq \lambda(\hat{D}) \leq 1$ for any $\hat{D} \in \hat{D}$ and $\sum_{\hat{D} \in \hat{D}} \lambda(\hat{D}) = 1$;

2. $\hat{\nu}_D$ is a probability measure on $\hat{D}$ such that its projection $(\hat{D}, \pi \ast \hat{\nu}_D)$ is a standard pair.

**Remark 10.** Definition 10 is motivated by the work [40], in which Pesin and Sinai used a crucial lemma (Lemma 13 therein) to construct the $\mu$-Gibbs measure of partially hyperbolic attractors. We adapt their notions in our setting.

If $\hat{\mu}$ has Pesin-Sinai decomposition, then by Lemma 5.3 it is easy to see that $\mu$ is carried by a standard family. To this end, we need the following lemma, which may be regarded as a variant of Lemma 13 of [40].

**Lemma 5.4.** Let $\hat{\nu}_n$ be a sequence of probability measures on $\hat{M}$ with the following properties:
(1) each $\tilde{v}_n$ has Pesin-Sinai decomposition $\tilde{v}_n = \sum_{D \in \mathcal{D}} \lambda_n(D) \cdot \tilde{v}_{n,D}$;

(2) let $\rho_{n,D}$ be the density of the standard pair $(D, \pi_n \tilde{v}_{n,D})$, and assume that $\rho_{n,D}$ converges uniformly in $D$ to a continuous function $\rho_D$ as $n \to \infty$;

(3) the sequence of measures $\frac{1}{n_j} \sum_{k=0}^{n_j} \tilde{v}_{n_k}$ converges weakly to a measure $\tilde{v}$ on $\hat{M}$, where $n_j$ is a subsequence of natural numbers.

Then the measure $\tilde{v}$ has Pesin-Sinai decomposition $\tilde{v} = \sum_{D \in \mathcal{D}} \lambda(D) \cdot \tilde{v}_D$, such that the density of $(D, \pi_n \tilde{v}_D)$ is exactly given by $\rho_D$.

The proof of Lemma 5.4 is almost the same as that of Lemma 13 in [40], by noticing that the uniform limit of regular density is still regular, as well as that the space of probability vectors on $\mathcal{D}$ is weakly compact. Hence we omit the proof here. In the rest of this subsection, we prove

Lemma 5.5. $\hat{\mu}$ has Pesin-Sinai decomposition.

Proof. It suffices to show that $\{\overline{m}_n\}_{n \geq 0}$ satisfies the first two conditions of Lemma 5.4 since the third condition is already shown by Lemma 5.3.

Recall that $\pi_n \overline{m}_n = T^n m$ and for any $D \in \mathcal{D}$, the interval $\hat{D}$ is an identical copy of $D$ via the projection $\pi$. Consider the Lebesgue standard pair $S_0 = (M, m)$, then $T^n m$ is exactly carried by the standard family

$$T^n S_0 = \sum_{\alpha \in M/\xi_n} m(W_\alpha) \cdot (T^n W_\alpha, \nu_\alpha), \quad (5.3)$$

where we denote $\xi_n = \{W_\alpha\}_{\alpha \in M/\xi_n}$ and set $\nu_\alpha(\cdot) := m(T^{-n}(\cdot)|W_\alpha)$. By the construction in Section 5.2, it is easy to see that each $T^n W_\alpha \in \mathcal{D}_n \subset \mathcal{D}$. Note that it is possible that $T^n W_\alpha = T^n W_{\alpha'}$ for distinct index $\alpha$ and $\alpha'$. We would like show that $T^n S_0$ is equivalent to a standard family of the form

$$T^n S_0 \equiv \sum_{D \in \mathcal{D}} \lambda_n(D) \cdot (D, \nu_{n,D}). \quad (5.4)$$

To this end, we need to combine standard pairs of $(5.3)$ over the same interval $D$ as follows. For simplicity, write $A_n := M/\xi_n$ and for any $D \in \mathcal{D}$, set

$$A_n(D) := \{\alpha \in A_n : T^n W_\alpha = D\}.$$

For any $D \in \mathcal{D}$ and any $n \in \mathbb{N}$, if $A_n(D) \neq \emptyset$, we define

$$\lambda_n(D) = \sum_{\alpha \in A_n(D)} m(W_\alpha) \quad \text{and} \quad \nu_{n,D} = \frac{\sum_{\alpha \in A_n(D)} m(W_\alpha) \cdot \nu_\alpha}{\sum_{\alpha \in A_n(D)} m(W_\alpha)}; \quad (5.5)$$

otherwise, we let $\lambda_n(D) = 0$ and $\nu_{n,D} = m_D$. By Lemma 5.2, the pair $(D, \nu_{n,D})$ is a standard pair. In this way, we obtain the equivalent standard family given by the RHS of $(5.4)$, whose total measure is $T^n m$. By lifting $T^n m$ to $\overline{m}_n$ and noting that $\pi^{-1}_D : D \to \hat{D}$ is trivial, we set $\tilde{v}_{n,D} = (\pi^{-1}_D)_{\nu_{n,D}}$, then $\overline{m}_n$ has Pesin-Sinai decomposition $\overline{m}_n = \sum_{D \in \mathcal{D}} \lambda_n(D) \cdot \tilde{v}_{n,D}$, that is, $\{\overline{m}_n\}_{n \geq 0}$ satisfies Condition (1) of Lemma 5.4.

Before we verify Condition (2) of Lemma 5.4 we introduce the following notations. For any $\alpha \in A_n$, we can associate a symbolic representation $\alpha = (\alpha_{-1}, \alpha_{-2}, \ldots, \alpha_{-n}) \in (M/\xi_1)^n$ such that $W_\alpha := \bigcap_{k=1}^n T^k W_{\alpha_{-k}}$. We further denote $A_\infty$ the inverse limit space of $\{A_n\}_{n \geq 1}$, that is, if $\alpha = (\alpha_{-1}, \alpha_{-2}, \ldots) \in A_\infty$, then $W_\alpha \neq \emptyset$ for all $n$, where $|\alpha|_n := (\alpha_{-1}, \alpha_{-2}, \ldots, \alpha_{-n})$ is the truncation of first $n$-words of $\alpha$. For any $D \in \mathcal{D}$, we set

$$A_\infty(D) := \{\alpha \in A_\infty : |\alpha|_n \in A_n(D) \text{ for any } n\}.$$
Note that \( A_\infty(D) \neq \emptyset \) for some \( D \in \mathcal{D} \), due to the Hofbauer tower construction and Assumption (H3). We also denote the subset \( A_n'(D) \) of \( A_n(D) \) such that any \( \alpha \in A_n'(D) \) cannot be extended to an element in \( A_\infty(D) \).

Now we are ready to check Condition (2) of Lemma 5.4. We shall only show the case given by (5.5), since the other case is trivial. Suppose that \( A_n(D) \neq \emptyset \). For any \( \alpha = (\alpha_1, \ldots, \alpha_n) \in A_n(D) \), let \( \nu_\alpha \) be the probability measure given by (5.3). Similar to (3.1), the density of \( \nu_\alpha \) is given by

\[
\rho_\alpha(x) = \frac{1}{m(W_\alpha)} \cdot \frac{1}{|T^n(x)|},
\]

where \( x_\alpha := (T^n|_{W_\alpha})^{-1}(x) \), i.e., \( x_\alpha \) is the \( n \)-th preimage of \( x \) in \( W_\alpha \). Alternatively, we define

\[
p_{n,\alpha}(x, y) := \frac{|(T^n)(y)|}{|(T^n)(x)|} = \prod_{k=1}^n \frac{|T'(y_{\alpha|k})|}{|T'(x_{\alpha|k})|}
\]

for all \((x, y) \in D \times D\), and we notice that

\[
\frac{1}{|(T^n)(x)|} = \int_D p_{n,\alpha}(x, y) dm(y), \quad m(W_\alpha) = \int_D p_{n,\alpha}(x, y) dm(x) dm(y).
\]

Therefore, the measures \( \nu_{n,D} \) given by (5.5) has density

\[
\rho_{n,D}(x) = \frac{\sum_{\alpha \in A_n(D)} m(W_\alpha) \cdot \rho_\alpha}{\sum_{\alpha \in A_n(D)} m(W_\alpha)} = \frac{\sum_{\alpha \in A_n(D)} \int_D p_{n,\alpha|n}(x, y) dm(y) + \sum_{\alpha \in A_n'(D)} m(W_\alpha) \cdot \rho_\alpha}{\sum_{\alpha \in A_n(D)} \int_D p_{n,\alpha|n}(x, y) dm(y) + \sum_{\alpha \in A_n'(D)} m(W_\alpha)}
\]

for any \( x \in D \). On the one hand, for any \( \alpha \in A_\infty(D) \), by Assumption (H2) and the formula (5.6), it is not hard to see that \( p_{n,\alpha|n} \) uniformly converges to \( p_\alpha \) on \( D \times D \), where

\[
p_\alpha(x, y) := \prod_{k=1}^\infty \frac{|T'(y_{\alpha|k})|}{|T'(x_{\alpha|k})|}, \quad \text{for any } (x, y) \in D \times D.
\]

On the other hand, we claim that \( \sum_{\alpha \in A_n'(D)} m(W_\alpha) \to 0 \) as \( n \to \infty \). Indeed, \( A_n'(D) = \bigcup_{k>n} A_{n,k}'(D) \), where \( A_{n,k}'(D) \) consists of all \( \alpha \in A_{n,k}'(D) \) which cannot be extended to an element in \( A_k(D) \). Note that the sets \( A_{n,k}'(D) \) is increasing in \( k \). Since \( G_0 = (M, m) \) and all its iterates \( T^k G_0 \) are proper standard families, we apply Theorem 2 to \( G_0 \) and \( T^k G_0 \) and get

\[
\sum_{\alpha \in A_{n,k}'(D)} m(W_\alpha) \leq |T^n m(D) - T^k m(D)| \leq 2(1 - \Theta_\epsilon)^{n/N}\epsilon
\]

Since \( k \) is arbitrary, we have \( \sum_{\alpha \in A_n'(D)} m(W_\alpha) \to 0 \) as \( n \to \infty \). By the above two observations, we conclude that \( \rho_{n,D} \) uniformly converges to

\[
\rho_D(x) := \frac{\sum_{\alpha \in A_\infty(D)} \int_D p_{n,\alpha|n}(x, y) dm(y)}{\sum_{\alpha \in A_n(D)} \int_D p_{n,\alpha|n}(x, y) dm(x) dm(y)}.
\]

This completes the verification of Condition (2) of Lemma 5.4 and hence \( \mu \) has Pesin-Sinai decomposition. \( \square \)
5.4 Proof of Theorem 4

In the previous subsections, we have shown that the measure $T_n^*m$ is lifted to the measure $\bar{m}_n$ given by (5.2), which has Pesin-Sinai decomposition. So its Cesaro mean $\bar{m}_n$ also has Pesin-Sinai decomposition, say,

$$\bar{m}_n = \sum_{D \in D} \eta_n(D) \cdot \bar{\omega}_{n, D}.$$ 

For any $\varepsilon > 0$, we set $D_\varepsilon := \{ D \in D : |D| < \varepsilon \}$. Consider the Lebesgue standard pair $S_0 = (M, m)$, and denote $r_{M,k}(x) := r_{T^kG_0}(T^kx)$ for any $k \geq 0$. By Lemma 3.1, we have

$$\sum_{D \in D_\varepsilon} \eta_n(D) \leq \frac{1}{n} \sum_{k=0}^{n-1} m(r_{M,k}(x) < \varepsilon) < C \varepsilon^{q_0}.$$ 

By Lemma 5.3 and Lemma 5.5, there is a subsequence $n_j$ such that $\bar{m}_{n_j} \rightarrow \bar{\mu}$ in the weak star topology. Moreover, $\bar{\mu}$ has Pesin-Sinai decomposition, say,

$$\bar{\mu} = \sum_{D \in D} \eta(D) \cdot \mu_D.$$ (5.7)

Moreover, the sequence of probability vectors $\{\eta_n(D)\}_{D \in D}$ converges to the probability vector $\{\eta(D)\}_{D \in D}$ in the weak star topology as $j \rightarrow \infty$. Therefore,

$$\sum_{D \in D_\varepsilon} \eta(D) = \lim_{j \rightarrow \infty} \sum_{D \in D_\varepsilon} \eta_{n_j}(D) < C \varepsilon^{q_0}. \quad (5.8)$$

Now we proceed the proof of Theorem 4. Since $\pi_* \bar{\mu} = \mu$, by (5.7), the density $h = \frac{d\pi_* \bar{\mu}}{dm}$ is given by

$$h = \sum_{D \in D} \eta(D) \cdot \frac{d(\pi_* \bar{\mu})}{dm} =: \sum_{D \in \mathcal{W}_h} h_D,$$

where we set $\mathcal{W}_h := \{ D \in D : \eta(D) > 0 \}$ and $h_D = \eta(D) \cdot \frac{d(\pi_* \bar{\mu})}{dm}$. Since $(D, \pi_* \bar{\mu})$ is a standard pair, by Lemma 3.1, we have $\frac{d(\pi_* \bar{\mu})}{dm}$ has $L_\infty$-norm bounded by $c^{C_r} |D|^{-1}$ and dynamically Hölder semi-norm bounded by $C_r e^{C_r} |D|^{-1}$. Hence for any $D \in \mathcal{W}_h$, we have

$$\|h_D\|_{D, \gamma} \leq (1 + C_r) e^{C_r} \eta(D) |D|^{-1}.$$ 

For any $s \in (1 - \gamma, 1]$, as the choice of $q_0$ is flexible and can be arbitrarily close to $\gamma$, it is not harm to assume that $s > 1 - q_0$. Then we have

$$\|h\|_{\mathcal{W}_h, \gamma, s} = \sum_{D \in \mathcal{W}_h} |D|^s \|h_D\|_{D, \gamma} \leq \sum_{D \in \mathcal{D}} \sum_{n=0}^{\infty} \eta(D) |D|^{s-1} < \infty.$$ 

The above convergence is shown as follows: we set $\Gamma_n = D_{2^n} \setminus D_{2^{n-1}}$, by (5.8), we get

$$\sum_{D \in \mathcal{D}} \sum_{n=0}^{\infty} \eta(D) |D|^{s-1} \leq \sum_{n=0}^{\infty} 2^{(1-s)(n+1)} \sum_{D \in D_{2^n}} \eta(D) \leq \sum_{n=0}^{\infty} 2^{(1-s)(n+1)} C_p (2^{-n})^{q_0} = C_p 2^{1-s} \sum_{n=0}^{\infty} 2^{n(1-q_0-s)} < \infty.$$ 

This completes the proof of Theorem 4.
6 Proof of Theorem 5

We first show that the system is exponential mixing with respect to the Lebesgue measure, that is,

Lemma 6.1. For any \( t \in [0, 1) \), we choose a scale \( q_0 \leq \min\{7, 1 - t\} \) satisfying \( 2.4 \). Then for any \( f \in \mathcal{H}_{W, \gamma, t} \) on some collection \( \mathcal{W} \) of countably many intervals and for any \( g \in L^\infty(m) \), we have

\[
\left| \int fg \circ T^n dm - \int f dm \int g dm \right| \leq 6C_\varepsilon\vartheta_\varepsilon \|f\|_{\mathcal{H}_{W, \gamma, t}} \|g\|_\infty.
\]

Here constants \( C_\varepsilon \) and \( \vartheta_\varepsilon \) are given by Theorem 3.

Remark 11. Note that the choice of \( q_0 \) in \( 2.4 \) is quite flexible. It is not hard to see from the proof of Theorem 3 that the constants \( C_\varepsilon \) and \( \vartheta_\varepsilon \) only depend on the choice of \( q_0, \delta_0 \) and the magnet interval \( U \). As \( \delta_0 \) and \( U \) are fixed but \( q_0 \) varies, \( C_\varepsilon \) and \( \vartheta_\varepsilon \) would also vary depending on the value of \( q_0 \).

Proof of Lemma 6.1 Without loss of generality, given a function \( f \in \mathcal{H}_{W, \gamma, t} \), we may assume that \( W = \{W_\alpha : \alpha \in A\} \) and \( f = \sum_{\alpha \in A} f_\alpha \) such that \( f_\alpha \neq 0 \) on each sub-interval \( W_\alpha \). We define on each \( W_\alpha \) two finite measures \( \tilde{\nu}_\alpha^1 \) and \( \tilde{\nu}_\alpha^2 \) such that their densities are given as follows:

\[
\frac{d\tilde{\nu}_\alpha^1}{dm} = f_\alpha + 2K_\alpha \quad \text{and} \quad \frac{d\tilde{\nu}_\alpha^2}{dm} = 2K_\alpha,
\]

where \( K_\alpha = \|f_\alpha\|_{W_\alpha, \gamma} > 0 \). Note that \( \frac{d\tilde{\nu}_\alpha^1}{dm} \in [K_\alpha, 3K_\alpha] \). Then we define two families \( \mathcal{S}^i = \sum_{\alpha \in A} \lambda_\alpha^i(W_\alpha, \nu_\alpha^i) \), \( i = 1, 2 \), by

\[
\nu_\alpha^i(.) = \tilde{\nu}_\alpha^i(\cdot | W_\alpha), \quad \text{and} \quad \lambda_\alpha^i = \frac{\tilde{\nu}_\alpha^i(W_\alpha)}{\sum_{\alpha \in A} \tilde{\nu}_\alpha^i(W_\alpha)}.
\]

We first show that \( \mathcal{S}^1 \) is a standard family in \( \mathcal{F} \). For any \( x, y \in W_\alpha \),

\[
\left| \frac{d\tilde{\nu}_\alpha^1}{dm}(x) - \frac{d\tilde{\nu}_\alpha^1}{dm}(y) \right| = \left| \log \frac{f_\alpha(x) + 2K_\alpha}{f_\alpha(y) + 2K_\alpha} \right|
\leq \log \left( 1 + \frac{|f_\alpha(x) - f_\alpha(y)|}{\min\{f_\alpha(x), f_\alpha(y)\} + 2K_\alpha} \right)
\leq \frac{|f_\alpha(x) - f_\alpha(y)|}{\min\{f_\alpha(x), f_\alpha(y)\}} + 2K_\alpha
\leq \frac{|f_\alpha|_{W_\alpha, \gamma} \gamma^{\mathcal{A}(x, y)} |x - y|}{\|f_\alpha\|_{W_\alpha, \gamma}} \leq C_\gamma \gamma^{\mathcal{A}(x, y)}.
\]

Hence each \( (W_\alpha, \nu_\alpha^1) \) is a standard pair, and thus \( \mathcal{S}^1 \) is a standard family. Further, since \( \tilde{\nu}_\alpha^1(W_\alpha) \leq 3K_\alpha \|W_\alpha\| \) and \( t \leq 1 - q_0 \), we have

\[
\mathcal{Z}(\mathcal{S}^1) = \sum_{\alpha \in A} \lambda_\alpha^1|W_\alpha|^{-q_0} = \frac{\sum_{\alpha \in A} \tilde{\nu}_\alpha^1(W_\alpha)|W_\alpha|^{-q_0}}{\sum_{\alpha \in A} \tilde{\nu}_\alpha^1(W_\alpha)} \leq \frac{3 \sum_{\alpha \in A} K_\alpha |W_\alpha|^{1-q_0}}{\sum_{\alpha \in A} \tilde{\nu}_\alpha^1(W_\alpha)} \leq \frac{3 \|f\|_{W_\gamma, t}}{\sum_{\alpha \in A} \tilde{\nu}_\alpha^1(W_\alpha)} < \infty.
\]

Similarly, we can show that \( \mathcal{S}^2 \) is a standard family and

\[
\mathcal{Z}(\mathcal{S}^2) \leq \frac{3 \|f\|_{W_\gamma, t}}{\sum_{\alpha \in A} \tilde{\nu}_\alpha^2(W_\alpha)} < \infty.
\]
By Theorem 3 \[ \| T^n \nu_q - \mu \|_{TV} \leq C_c \varphi^n \mathcal{Z}(S^i), \ i = 1, 2, \] which implies that
\[
\left\| T^n \left( \sum_{\alpha \in A} \bar{\nu}^q_{\alpha} \right) - \left( \sum_{\alpha \in A} \bar{\nu}^q_{\alpha} (W_{\alpha}) \right) \mu \right\|_{TV} \leq 3C_c \varphi^n \| f \|_{W_{\gamma,t}}.
\]

Therefore, for any \( g \in L^\infty \),

\[
\left| \int fg \circ T^n dm - \int dm \int gd\mu \right| = \left| T^n \left( \sum_{\alpha \in A} (\bar{\nu}^q_{\alpha} - \bar{\nu}^q_{\alpha}^2) \right) (g) - \left( \sum_{\alpha \in A} (\bar{\nu}^q_{\alpha} (W_{\alpha}) - \bar{\nu}^q_{\alpha}^2 (W_{\alpha})) \right) \mu(g) \right| \leq \left\| T^n \left( \sum_{\alpha \in A} \bar{\nu}^q_{\alpha} \right) - \left( \sum_{\alpha \in A} \bar{\nu}^q_{\alpha} (W_{\alpha}) \right) \mu \right\|_{TV} \| g \|_{\infty} \leq 6C_c \varphi^n \| f \|_{W_{\gamma,t}} \| g \|_{\infty}.
\]

This completes the proof of Lemma 6.1.

Now we are ready to prove Theorem 5. For any \( t \in [0, \bar{t}] \), again as the choice of \( q_0 \) is flexible, we may set
\[
q_0 := \frac{\bar{t} - t}{2} \quad \text{and} \quad s := 1 - \frac{\bar{t} + t}{2}.
\]

It is obvious that \( s \in (1 - \bar{t}, 1] \), then by Theorem 4, the invariant density \( h = dm/d\mu \in \mathcal{H}_{W_{\gamma,t}} \). We denote the collection \( W_h = \{ W_\beta : \beta \in \mathcal{B} \} \) and write \( h = \sum_{\beta \in \mathcal{B}} h_\beta \), where each \( h_\beta \in \mathcal{H}_{W_{\beta,\gamma}} \).

For any \( f \in \mathcal{H}_{W_{\gamma,t}} \) with a collection \( W = \{ W_\alpha : \alpha \in A \} \), we write \( f = \sum_{\alpha \in A} f_\alpha \), where each \( f_\alpha \in \mathcal{H}_{W_{\alpha,\gamma}} \). Set the joint collection by \( W \cup W_h := \{ W_\alpha \cap V_\beta : \alpha \in A, \ \beta \in \mathcal{B} \} \). Then we can write
\[
fh = \sum_{\alpha \in A} \sum_{\beta \in \mathcal{B}} |W_{\alpha} \cap V_{\beta}|^t \cdot |f_\alpha| \cdot \| h_\beta \|_{W_{\alpha \cap V_{\beta},\gamma}} = \sum_{\alpha \in A} |W_{\alpha}|^t \cdot |f_\alpha| \cdot \sum_{\beta \in \mathcal{B}} |V_{\beta}|^s \cdot \| h_\beta \|_{W_{\alpha,\gamma}} \leq \| f \|_{W_{\alpha,\gamma}} h \|_{L_{W_{\beta,\gamma}}}.
\]

In other words, \( fh \in \mathcal{H}_{W \cup W_h, \gamma,t+s} \) such that \( \| f \|_{W \cup W_h, \gamma,t+s} \leq \| f \|_{W_{\gamma,t}} h \|_{L_{W_{\beta,\gamma}}} \). Note that the scale is \( q_0 = 1 - (t + s) \), and note that the constants \( C_c \) and \( \varphi_c \) in Lemma 6.1 depend on \( q_0 \) and thus on \( t \). By Lemma 6.1

\[
\left| \int fg \circ T^n dm - \int dm \int gd\mu \right| = \left| \int fhg \circ T^n dm - \int dm \int fdh dm \int gd\mu \right| \leq 6C_c \varphi^n \| fh \|_{W \cup W_h, \gamma,t+s} \| g \|_{\infty} \leq C_c \varphi^n \| f \|_{W_{\gamma,t}} \| g \|_{\infty},
\]

where \( C_t = 6C_c \| h \|_{W_{h},\gamma,s} \) and \( \varphi_t = \varphi_c \). This finishes the proof of Theorem 5.
7 Proof of Theorem 6

Let \( f \in \mathcal{H}_{W,\gamma,t}^d \) be a function satisfying all the conditions in Theorem 6 and let \( \sigma_f^2 \) be given by (2.12). If \( \sigma_f^2 = 0 \), then it is well known that \( f \) is a coboundary up to a constant, i.e., \( f = g - g \circ T + \mathbb{E}(f) \) for some \( g \in L^2(\mu) \) (see e.g. Theorem 18.2.2 in [31]), and thus the ASIP is automatic. In the rest of the proof, we concentrate on the case when \( \sigma_f^2 > 0 \).

Given an integrable function \( f : M \to \mathbb{R} \) and a measurable partition \( \xi \) of \( M = [0, 1] \), we denote by \( \mathbb{E}(f|\xi) \) the conditional expectation of \( f \) with respect to \( \xi \). We also denote by \( \sigma(\xi) \) the Borel \( \sigma \)-algebra on \( M \) generated by \( \xi \).

We recall the following result in [41] (see also §7.9 in [16]).

**Proposition 7.1.** Suppose there exist constants \( \varepsilon \in (0, 2] \) and \( C > 0 \) such that

1. \( f \in L^{2+\varepsilon}(\mu) \);
2. for all \( m \geq 1 \), \( \|f - \mathbb{E}(f|\xi_m)\|_{L^{2+\varepsilon}(\mu)} \leq Cm^{-(2+\varepsilon)/\varepsilon} \);
3. \( \sigma_f^2 > 0 \) and \( \text{Var}\left(\sum_{k=0}^{n-1} f \circ T^k\right) = n\sigma_f^2 + O(n^{1-\varepsilon/30}) \);
4. For any \( n \geq 1 \) and \( \mu(A \cap B) - \mu(A)\mu(B) \leq Cn^{-168(1+2/\varepsilon)} \) for any \( A \in \sigma(\xi_m) \) and \( B \in \sigma(T^{-n+m}\xi_\infty) \).

Then the stationary process \( \{f \circ T^n\}_{n \geq 0} \) satisfies the ASIP.

Now we continue to prove Theorem 6 by verifying conditions in Proposition 7.1 as follows:

- Since \( f \in \mathcal{H}_{W,\gamma,t}^d \subset L^{1/t}(\mu) \), where \( t < \frac{1}{2} \), then Condition (1) holds by taking \( \varepsilon = \min\left\{\frac{2}{t} - 2\right\} \).

- To check Condition (2), we denote the adapted collection \( W = \{W_\alpha : \alpha \in \mathcal{A}\} \) of countably many intervals such that the endpoints of \( W_\alpha \) belong to \( S_{n(\alpha)} \) for some \( n(\alpha) \in \mathbb{N} \). In other words, \( W_\alpha \in \sigma(\xi_{n(\alpha)}) \). Then we rewrite \( f \in \mathcal{H}_{W,\gamma,t}^d \) as \( f = \sum_{\alpha \in \mathcal{A}} f_\alpha \), such that \( f_\alpha \in \mathcal{H}_{W,\gamma,t}^d \). Now for every interval \( W \in \xi_m \), and any two points \( x, y \in W \), we have \( s(x, y) \geq m \) and thus

\[
|f_\alpha(x) - f_\alpha(y)| \leq \begin{cases} 2\|f_\alpha\|_{\infty}, & \text{if } s(x, y) < n(\alpha), \\ \|f_\alpha|_{\mathcal{H}_{W,\gamma,t}^d}\gamma^{s(x,y)}, & \text{if } s(x, y) \geq n(\alpha), \\ \end{cases}
\]

which implies that \( \|f_\alpha - \mathbb{E}(f_\alpha|\xi_m)\|_{\infty} \leq 2\|f_\alpha\|_{\mathcal{H}_{W,\gamma,t}^d}\gamma^{m-n(\alpha)} \). Also, note that if \( n(\alpha) \leq m \), then both \( f_\alpha \) and \( \mathbb{E}(f_\alpha|\xi_m) \) are supported on \( W_\alpha \).

Note that \( 2 + 7/\varepsilon = \max\left\{\frac{11}{2}, \frac{2+3\varepsilon}{2+\varepsilon}\right\} \) < \( a \), where \( a \) is given by (2.10). Set \( b = a/(2+7/\varepsilon) \). By Minkowski’s inequality, as well as (2.9) and (2.11),

\[
\|f - \mathbb{E}(f|\xi_m)\|_{L^{2+\varepsilon}(\mu)} \leq \sum_{\alpha \in \mathcal{A}} \|f_\alpha - \mathbb{E}(f_\alpha|\xi_m)\|_{L^{1/t}(\mu)} \\
\leq \sum_{\alpha \in \mathcal{A}: n(\alpha) < m^{1/8}} \|f_\alpha - \mathbb{E}(f_\alpha|\xi_m)\|_{\infty} \mu(W_\alpha)^t \leq 2\|f\|_{\mathcal{H}_{W,\gamma,t}^d}\gamma^{m-m^{1/8}} + O\left(m^{-a/b}\right)
\]

26
conditions in Proposition 7.1, and hence the stationary process \( \{ f \circ T^n \} \).

Hence Condition (2) holds.

- Note that the series in (2.12) converges absolutely by Condition (2.11). By direct computation, we have

\[
\text{Var} \left( \sum_{k=0}^{n-1} f \circ T^k \right) = n\sigma_f^2 - 2 \sum_{k>n} n \text{Cov} (f, f \circ T^k) - 2 \sum_{k=1}^{n-1} k \text{Cov} (f, f \circ T^k)
\]

\[
= n\sigma_f^2 + O \left( n^{1-\frac{\alpha}{2}} \right) = n\sigma_f^2 + O \left( n^{1-\varepsilon/30} \right).
\]

Therefore, Condition (3) holds.

- By the \( T \)-invariance of \( \mu \), it suffices to show Condition (4) with \( m = 1 \). Note that any \( A \in \sigma(\xi_1) \) is a disjoint union of intervals in \( \xi_1 \). We take \( f_A = 1_A + 1 \), then \( f_A \in \mathcal{H}_{M, \gamma} = \mathcal{H}_{(M), \gamma, 0} \) such that

\[
\| f_A \|_{M, \gamma} = \| f_A \|_\infty + \| f_A \|_{\mathcal{H}_{M, \gamma}} \leq 2 + 1/\gamma.
\]

Also, \( B \in \sigma(T^{-(n+1)} \xi_\infty) \) means that there is a Borel measurable subset \( B' \subset M \) such that \( B = T^{n+1} B' \), and thus \( 1_B = 1_{B'} \circ T^{n+1} \). Therefore, by Theorem 6

\[
|\mu(A \cap B) - \mu(A)\mu(B)| = \left| \int f_A \cdot 1_{B'} \circ T^{n+1} d\mu - \int f_A d\mu \int 1_{B'} d\mu \right|
\]

\[
\leq C_0 o^{n+1} + o^M, \gamma \| 1_{B'} \|_\infty
\]

\[
= C_0 (2 + 1/\gamma) o^{n+1},
\]

which indicates Condition (4).

To sum up, any function \( f \in \mathcal{H}_{W, \gamma, t}^d \) satisfying all the conditions in Theorem 6 also satisfies the four conditions in Proposition 7.1 and hence the stationary process \( \{ f \circ T^n \}_{n \geq 0} \) satisfies the ASIP. The proof of Theorem 6 is complete.

8 Examples and Remarks

We shall revisit several examples which were previously studied in the literature. Applying our results to these examples, we could reinterpret some known results and make some generalizations.

8.1 A class of piecewise linear maps

In this subsection, we consider a class of piecewise linear map on \( M = [0, 1] \) with infinitely many inverse branches. More precisely, given a sequence of positive numbers \( \{ a_k \}_{k \geq 1} \) such that \( \sum_{k \geq 1} a_k = 1 \). Set \( b_0 = 1 \) and for \( k \geq 1 \),

\[
b_k = 1 - \sum_{m=1}^{k} a_m = \sum_{m=k+1}^{\infty} a_m.
\]

It is clear that \( \xi_1 := \{ W_k \}_{k \geq 1} \) is a partition of \( M = [0, 1] \), where \( W_k := (b_k, b_{k+1}] \). Pick another sequence \( \{ \Lambda_k \}_{k \geq 1} \) of positive numbers such that \( \Lambda_k \geq 2 \). Moreover, we assume that \( a_k \Lambda_k \geq b_1 \) for \( k = 1, 2 \), and \( a_k \Lambda_k \geq b_{k-2} \) for any \( k > 2 \). Then we define a piecewise linear map \( T : M \rightarrow M \) by setting

\[
T(x) = \begin{cases} 
0, & x = 0, \\
\Lambda_k(x - b_k), & x \in W_k.
\end{cases}
\]

(8.1)
Albeit $T$ is piecewise linear, the existence of acip (absolutely continuous $T$-invariant probability measure) heavily depends on the above parameters. We emphasize that even if the map is Markov, the "big image property" (i.e., \( \inf_{n \geq 1} |T W_k| > 0 \)) does not hold, and hence the classical theory of Gibbs-Markov systems is not applicable in our situation. We recall some results for these piecewise linear maps in earlier literature.

1. Rychlik [42] showed that if \( \sum_{k \geq 1} \Lambda_k^{-1} < \infty \), then $T$ admits an acip which enjoys the exponential mixing. Rychlik also constructed a counter-example, that is, $T$ does not admit an acip if \( a_k = 2^{-k} \) and \( \Lambda_k = 2 \).

2. Bruin and Todd studied in [8] a class of piecewise linear maps\(^2\) which is a simplified linear model of the induced map of the Fibonacci unimodal map. To be precise, given any \( \lambda \in (0, 1) \), we set \( a_k := \lambda^{k-1}(1 - \lambda) \) for all \( k \geq 1 \) and thus \( b_k := \lambda^k \) for all \( k \geq 0 \). Meanwhile, put \( \Lambda_1 := 1/a_1 \) and \( \Lambda_k := 1/a_2 \) for all \( k \geq 2 \). The corresponding map is denoted by $T_\lambda$. Bruin and Todd showed that $T_\lambda$ admits an acip if and only if \( \lambda \in (0, \frac{1}{2}) \). Moreover, whenever \( \lambda \in (0, \frac{1}{2}) \), they also showed that the invariant density restricting on each $W_k$ is a constant equals to

\[
\frac{v_k}{|W_k|} := \frac{1-2\lambda}{\Lambda_k} \cdot \frac{(\frac{\lambda}{1-\lambda})^k}{a_k} = \frac{(1-\lambda)(1-2\lambda)}{(1-\lambda)^k}.
\]

(8.2)

The following proposition provides a sufficient condition for the existence of acip, when $T$ is the piecewise linear map given by [8,1].

**Proposition 8.1.** If there is \( q \in (0, 1) \) such that

\[
\inf_{k \geq 2} \Lambda_k > 2, \quad \text{and} \quad \limsup_{N \to \infty} \sum_{k=N+1}^{\infty} \frac{1-q}{b_k^{1-q}} \frac{\Lambda_k^{-q}}{a_k} < 1,
\]

then the piecewise linear map $T$ admits an acip, which satisfies the exponential decay of correlation and almost sure invariant principle.

**Remark 12.** It is not hard to see that Rychlik’s condition $\sum_{k \geq 1} \Lambda_k^{-1} < \infty$ is stronger than Condition [8,3]. Therefore, the results in [30,12] are recovered by our coupling method. Also, Condition [8,3] never holds for any \( q \in (0, 1] \) if \( \Lambda_k = 2 \) for all \( k \), which corresponds to the absence of acip.

**Proof of Proposition [8,1].** It is obvious that Assumption (H2) holds since the log Jacobian $\log |T'|$ is constant on each interval $W_k \in \xi_1$.

We next verify that $T$ satisfies Assumption (H3) by showing the second branch $W_2$ is a magnet interval. By our definition, it is easy to see $TW_1 \supset W_2$, $TW_2 \supset W_2$, and $TW_k \supset \cup_{m=k-1}^{\infty} W_{m} \supset W_{k-1}$ for any $k > 2$. Hence a component of $T^n W_k$ must contain $W_2$ for any $n \geq k-2$. For any interval $W \subset M$, by the uniform expansion with rate $\Lambda_k > 2$, $T^n W$ must be cut by $S_1 = \{b_k\}_{k \geq 1}$ for some positive integer $n_0 \leq -\log_2 |W|$. We pick a component $V$ of $T^n W$ whose left endpoint belongs to $S_1$, then $TV \supset W_\ell$ for some $\ell > 2$. Therefore, at least one component of $T^n W$ contains $W_2$ for any $n \geq n_W := n_0 + \ell - 1$, which implies that $W_2$ is a magnet.

Finally, we focus on the validity of Assumption (H1). Indeed, let $W$ be an interval of length less than a sufficiently small $\delta > 0$.

- If $W$ is away from the accumulation point 0, then it only intersects two consecutive intervals in $\xi_1$, say $W_k$ and $W_{k+1}$, and thus

\[
\sum_{\alpha \in W/\xi_1} \left( \frac{|W|}{T W_{\alpha}} \right)^q |W_{\alpha}| \left( \frac{1}{\Lambda_k} + \frac{1}{\Lambda_{k+1}} \right)^q \leq \left( 1 + \frac{1}{\inf_{k \geq 2} \Lambda_k} \right)^q.
\]

\(^2\)By personal communication, Bruin and Todd named such map as the vSSV map, because it was introduced by van Strien to Stratmann and Vogt. This map has a bearing on the existence and nature of wild attractors in interval dynamics, see [4].
• Otherwise, if $W$ is close to $0$, without loss of generality, we may assume $W = [0, b_N] = \bigcup_{k=N+1}^{\infty} W_k$ for sufficiently large $N$. Then

$$
\sum_{\alpha \in W/\xi} \left( \frac{|W|}{|TW_\alpha|} \right)^q \frac{|W_\alpha|}{|W|} = \sum_{k=N+1}^{\infty} \left( \frac{|W|}{|TW_k|} \right)^q \frac{|W_k|}{|W|} = \frac{\sum_{k=N+1}^{\infty} \frac{a_k^{1-q} \Lambda_k^{-q}}{b_N^{1-q}}}{\sum_{k=N+1}^{\infty} a_k^{1-q} \Lambda_k^{-q}}
$$

In other words, Condition (8.3) guarantees Assumption (H1) - the Chernov’s one-step expansion holds at $q$-scale in either of the above cases.

Applying Theorems 8.1 and 8.2 we can deduce all the assertions of Proposition 8.1.

We now provide two particular examples of piecewise linear maps which satisfy Condition (8.3) and thus Proposition 8.1.

1. In spirit of Rychlik’s results and counter-example in [12], we consider the piecewise linear map with $a_k = 2^{-k}$ and $\Lambda_k = k$. It is straightforward that $b_N = 2^{-N}$ and for any $q \in (0, 1)$, we have

$$
\sum_{k=N+1}^{\infty} \frac{a_k^{1-q} \Lambda_k^{-q}}{b_N^{1-q}} = \frac{\sum_{k=N+1}^{\infty} \frac{2^{k(q-1)}k^{-q}}{2^{N(q-1)}}}{2^{q-1}N^{-q}} \leq \frac{2^{q-1}N^{-q}}{1 - 2q^{-1}} \to 0
$$

as $N \to \infty$, and hence Condition (8.3) holds.

2. Let $T_\lambda$ be the piecewise linear map that Bruin and Todd studied in [8]. Given any $\lambda \in (0, 1)$, we recall that $a_k := \lambda^{k-1}(1 - \lambda)$ for all $k \geq 1$ and thus $b_k := \lambda^k$ for all $k \geq 0$. Moreover, $\Lambda_k := 1/a_1$ and $\Lambda_k := 1/a_2$ for all $k \geq 2$. We claim that $T_\lambda$ satisfies Condition (8.3) and thus Proposition 8.1 if and only if $\lambda \in (0, \frac{1}{2})$, which agrees with the results of Bruin and Todd in [8]. Indeed, it is easy to see that for any $\lambda \in (0, 1)$,

$$
\inf_{k \geq 2} \Lambda_k = \frac{1}{\lambda(1 - \lambda)} \geq 4.
$$

Meanwhile,

$$
\sum_{k=N+1}^{\infty} \frac{a_k^{1-q} \Lambda_k^{-q}}{b_N^{1-q}} = \frac{\sum_{k=N+1}^{\infty} \frac{\lambda^{k-1}(1 - \lambda)\left[1 - \lambda^{1-\lambda}\right]^{-q}}{\lambda^{1-q}N}}{\lambda^{(1-q)N}} = \frac{\lambda^q(1 - \lambda)}{1 - \lambda^{1-q}}. \tag{8.4}
$$

It is not hard to check that (8.4) is less than 1 if and only if $\lambda^{1-q} < 1 - \lambda$, and hence (8.4) is less than 1 for some $q \in (0, 1)$ if and only if $\lambda \in (0, \frac{1}{2})$. In other words, Condition (8.3) holds if and only if $\lambda \in (0, \frac{1}{2})$.

We remark that when $\lambda \in (0, \frac{1}{2})$, the invariant density given by (8.2) is a dynamically Hölder series, which agrees with our Theorem 7. More precisely, it is straightforward to check that the invariant density belongs to $\mathcal{H}_{W, \gamma, s}$, where $W = \{W_k\}_{k \geq 1}$, for any $\gamma \in (0, 1)$, and for any $s \in (0, 1)$ such that $\lambda^s < 1 - \lambda$.

### 8.2 Certain unbounded observables

Let $T : M = [0, 1] \to M$ be a one-dimensional map satisfying Assumption (H1), i.e., the one-step expansion at $q$-scale, and recall that $\mathcal{C}$ is the supremum of such $q$ given in (2.3). It directly from (H1) that $T$ is uniformly expanding, i.e., there exists $\Lambda > 1$ such that $\inf_{x \in M \setminus S_{\infty}} |T'(x)| > \Lambda$. It is easy to see that the separation time $s(\cdot, \cdot)$ in Definition 7 induces a weaker metric on $M$, that is, there exists $C > 0$ such that

$$
|x - y| \leq CA^{-s(x, y)}, \quad \text{for any } x, y \in M.
$$

29
Let $\gamma$ be the constant given by (2.2), which can be taken arbitrarily close to 1, and set $\kappa := -\log_{\lambda} \gamma > 0$. If $f$ is a $\kappa$-H"older function on an interval $W \subset M$, then $f$ is also a dynamically H"older function on $W$ with parameter $\gamma$ such that $|f|_{W, \gamma} \leq C^{-\kappa} |f|_{C^\infty(W)}$. Note that in applications, we could always take $\gamma$ arbitrarily close to 1.

As pointed out in Remark 4 the space $\mathcal{H}_{W, \gamma, t}$ with $t > 0$ would contain some unbounded observables. For instance, for any $\tau \in (0, \frac{T}{2})$, we consider the unbounded function

$$f(x) = \begin{cases} x^{-\tau}, & 0 < x \leq 1, \\ 0, & x = 0. \end{cases}$$  \hspace{1cm} (8.5)$$

Remark 13. This function was studied in Gou"ezel’s note [23], in which $T$ is the doubling map. He showed that $f$ satisfies a stable law when $\tau \geq \frac{T}{2}$, and he also pointed out that $f$ satisfies a CLT when $\tau \in (0, \frac{T}{2})$, using the criteria by Dedecker [19]. We shall show below that the ASIP holds in the latter case.

It is clear that $f \in \mathcal{H}_{W, \gamma, t}$ for any $\tau \in (t, \frac{T}{2})$ and some $\gamma$ close to 1, where the collection is chosen to be $W = \{W_k := (2^{-k}, 2^{-k+1}]\}_{k \geq 1}$. Indeed, we set $\kappa := -\log_{\lambda} \gamma$, and write $f = \sum_{k \geq 1} f_k$ with $f_k = f 1_{W_k}$, then

$$\|f_k\|_{W_k, \gamma} \lesssim \|f_k\|_{L^\infty(W_k)} + |f_k|_{C^\infty(W_k)} \lesssim \|f_k\|_{L^\infty(W_k)} + |f_k|_{L^\infty(W_k)} |W_k|^{1-\kappa} \lesssim 2^{k\tau} + 2^{k(\tau+1)} 2^{-k(1-\kappa)} \lesssim 2^{k(\tau+\kappa)}.$$  \hspace{1cm} (8.6)

Thus, if we choose $\gamma$ close to 1 such that $\kappa := -\log_{\lambda} \gamma < t - \tau$, then

$$\|f\|_{\mathcal{H}_{W, \gamma, t}} = \sum_{k \geq 1} |W_k|^t \|f_k\|_{W_k, \gamma} \lesssim \sum_{k \geq 1} 2^{k(\tau+\kappa-t)} < \infty.$$  \hspace{1cm} (8.7)

By Theorem 6 the correlations between any unbounded function $f \in \mathcal{H}_{W, \gamma, t}$ and any bounded observable $g \in L^\infty(m)$ decays exponentially fast.

Finally, we discuss the space $\mathcal{H}_{W, \gamma, t}^{ad}$ with fast tail, for which the ASIP applies by Theorem 6. For simplicity, we consider the doubling map $T : x \mapsto 2x \pmod{1}$, with the partition $\xi_1 = \{[0, \frac{1}{2}), (\frac{1}{2}, 1]\}$ and invariant measure $\mu = m$. We claim that if $\tau \in (0, \frac{T}{2})$, then the unbounded function $f$ given by (8.6) satisfies the ASIP. Indeed,

- The collection $W = \{W_k := (2^{-k}, 2^{-k+1}]\}_{k \geq 1}$ is adapted such that $n(k) = k$. Pick any $t \in (\tau, \frac{T}{2})$, it follows from (8.6) that $f \in \mathcal{H}_{W, \gamma, t}^{ad}$.
- Moreover, $f$ has fast tail since

$$\sum_{k \geq n} \|f_k\|_{L^{1/t}(\mu)} \leq \sum_{k \geq n} \|f_k\|_{L^\infty} |W_k|^t = \sum_{k \geq n} 2^{k(\tau-t)} = O \left( 2^{n(\tau-t)} \right).$$

- The auto-correlations condition (2.11) holds since the Fourier coefficients of $f$ satisfy $a_k := \int_0^1 x^{-\tau} e^{i 2\pi k x} dx \lesssim k^{\tau-1}$, and thus

$$|\text{Cov}(f, f \circ T^n)| = \sum_{k=1}^\infty a_k a_{kn} 2^{n(\tau-1)} \sum_{k \geq 1} k^{2(\tau-1)} = O \left( (2^{\tau-1})^n \right).$$

Therefore, the unbounded function $f$ given by (8.5) satisfies the ASIP and thus the CLT.
A Proof of Lemma 4.1 and 4.2

A.1 Proof of Lemma 4.1

Proof. By the formula of $Z(\cdot)$ and (4.1), for any standard family $\mathcal{G}$,

$$
\left| \frac{Z(\tilde{\mathcal{G}})}{Z(\mathcal{G})} - 1 \right| \leq \frac{\sum_{\alpha \in \tilde{\mathcal{A}}} \frac{1}{\Delta^\alpha} \lambda_\alpha W_\alpha^{-q_0} + \sum_{\alpha \in \mathcal{A} \setminus \tilde{\mathcal{A}}} \frac{1}{\Delta^\alpha} \lambda_\alpha W_\alpha^{-q_0}}{\sum_{\alpha \in \mathcal{A}} \lambda_\alpha W_\alpha^{-q_0} + \sum_{\alpha \in \mathcal{A} \setminus \mathcal{A}} \lambda_\alpha W_\alpha^{-q_0}}
$$

$$
\leq \max \left\{ \frac{\rho(1 - \tilde{\rho})}{1 - \tilde{\rho}^\alpha}, \frac{\tilde{\rho}^\alpha}{1 - \tilde{\rho}^\alpha} \right\} \leq \frac{\rho}{1 - \rho}.
$$

Moreover, for any $\alpha \in \mathcal{A}$, we have $W_\alpha = U$, and the density of $\frac{\nu_\alpha - \tilde{\mu}_W}{1 - \rho}$ is given by $\frac{\rho_\alpha - \tilde{\rho}}{1 - \rho}$ By Lemma 3.1 for any $x, y \in W_\alpha$,

$$
\left| \log \left( \frac{\rho_\alpha(x) - \tilde{\rho}}{1 - \rho} - \log \frac{\rho_\alpha(y) - \tilde{\rho}}{1 - \rho} \right) \right| \leq \left| \log \rho_\alpha(x) - \log \rho_\alpha(y) \right| + \log \left( 1 + \frac{|\rho_\alpha(x)^{-1} - \rho_\alpha(y)^{-1}|}{1 - \rho} \max\{\rho_\alpha(x)^{-1}, \rho_\alpha(y)^{-1}\} \right)
$$

$$
\leq C_\rho \gamma^{s(x,y)} + \frac{C_\rho e^{C_\rho}}{(1 - \rho e^{C_\rho})|U|} \gamma^{s(x,y)}.
$$

We use the fact $\log(1 + z) \leq z$ for any $z \geq 0$ in the last inequality. Hence, for any $\alpha \in \mathcal{A}$,

$$
\left| \log \left( \frac{\rho_\alpha - \tilde{\rho}}{1 - \rho} \right) \right|_{W_\alpha, \gamma} \leq C_\rho + \frac{C_\rho e^{C_\rho}}{(1 - \rho e^{C_\rho})|U|}.
$$

Therefore, we can choose $\tilde{\rho}_e$ small enough such that for any $\tilde{\rho} \in (0, \tilde{\rho}_e)$ and any standard family $\mathcal{G}$, we have that $Z(\tilde{\mathcal{G}}) \leq Z(\mathcal{G})/\theta_0$, and the density of each pair in $\tilde{\mathcal{G}}$ satisfies (3.12) and (3.11). By Remarks 7 and 9 we have that $T\tilde{\mathcal{G}}$ is a standard family, and

$$
Z(T\tilde{\mathcal{G}}) \leq e^{4C_\rho} \left( Z(\tilde{\mathcal{G}}) \theta_0 + c_0 \right) \leq e^{4C_\rho} \left( Z(\mathcal{G}) + c_0 \right).
$$

The proof of this lemma is complete. \hfill \Box

A.2 Proof of Lemma 4.2

We first choose an integer $k \geq 1$ such that $(k/3)^{q_0} \geq 2C_\rho$, where $C_\rho$ is the proper constant that we choose in (2.6). We then divide $M = [0, 1]$ into $k$ sub-intervals $W_1, W_2, \ldots, W_k$ of equal length. For each Lebesgue standard pair $\mathcal{G}_i = (W_i, m_{W_i})$, by Assumption (H3), there exists $n_{W_i} \geq 1$ such that for any $n \geq n_{W_i}$, at least one component of $T^n(W_i)$ contains $U$, which means that $\delta(T^n\mathcal{G}_i) > 0$. We set

$$
n_e := \max\{n_p, \max_{1 \leq i \leq k} n_{W_i}\}, \quad d_e := \min_{1 \leq i \leq k} \delta(T^n\mathcal{G}_i).
$$

For any proper standard family $\mathcal{G} = \sum_{\alpha \in \mathcal{A}} \lambda_\alpha (W_\alpha, \nu_\alpha)$, we denote $\mathcal{A}_0 := \{ \alpha \in \mathcal{A} : |W_\alpha| \geq 3/k \}$, then

$$
\sum_{\alpha \in \mathcal{A}_0} \lambda_\alpha = 1 - \sum_{\alpha \not\in \mathcal{A}_0} \lambda_\alpha \geq 1 - \left( \frac{k}{3} \right)^{q_0} Z(\mathcal{G}) \geq \frac{1}{2}.
$$
For any $\alpha \in A_0$, there exists $1 \leq i_\alpha \leq N$ such that $W_\alpha$ contains $W_{i_\alpha}$. We then cut the Lebesgue standard pair $(W_\alpha, m_{W_\alpha})$ by the two endpoints of $W_{i_\alpha}$, and obtain a new standard family $S'_\alpha$. Note that the weight of $W_{i_\alpha}$ in $G'_\alpha$ is:

$$\frac{|W_{i_\alpha}|}{|W_\alpha|} \geq \frac{1}{k}.$$ 

By (4.2), (4.3) and (4.4), we have:

$$\delta(T^{n_e}G) \geq \sum_{\alpha \in A_0} \lambda_\alpha \delta(T^{n_e}(W_\alpha, \nu_\alpha)) \geq \sum_{\alpha \in A_0} \lambda_\alpha e^{-C_r} \delta(T^{n_e}(W_\alpha, m_{W_\alpha}))$$

$$\geq \sum_{\alpha \in A_0} \lambda_\alpha e^{-C_r} \frac{1}{k} \delta(T^{n_e}(G'_\alpha))$$

$$\geq \sum_{\alpha \in A_0} \lambda_\alpha e^{-C_r} \frac{1}{k} d'_c$$

$$\geq \frac{e^{-C_r} d'_c}{2k} =: d_c.$$

This completes the proof of the lemma.

Acknowledgement
J. Chen would like to thank Tianyuan Mathematical Center in Southwest China (TMCSC), where part of the work was done, for their hospitality and support (Grant 11826102).

Y.-W. Zhang is partially support by the NSFC grant 11701200 and 11871262. Y.-W. Zhang would also like to thank AMS China exchange program Ky and Yu-Fen Fan fund travel grant for visiting University of Massachusetts Amherst, where part of this work was carried out.

References
[1] M. Blank, G. Keller, and C. Liverani. Ruelle-Perron-Frobenius spectrum for Anosov maps. *Nonlinearity*, 15(6):1905–1973, 2002.

[2] X. Bressaud, R. Fernández, and A. Galves. Decay of correlations for non-Holderian dynamics. A coupling approach. *Electron. J. Probab.*, 4:no. 3, 19, 1999.

[3] X. Bressaud and C. Liverani. Anosov diffeomorphisms and coupling. *Ergodic Theory Dynam. Systems*, 22(1):129–152, 2002.

[4] H. Bruin. Induced maps, Markov extensions and invariant measures in one-dimensional dynamics. *Comm. Math. Phys.*, 168(3):571–580, 1995.

[5] H. Bruin and C. Kalle. Natural extensions for piecewise affine maps via Hofbauer towers. *Monatsh. Math.*, 175(1):65–88, 2014.

[6] H. Bruin, G. Keller, T. Nowicki, and S. van Strien. Wild Cantor attractors exist. *Ann. of Math. (2)*, 143(1):97–130, 1996.

[7] H. Bruin and M. Todd. Markov extensions and lifting measures for complex polynomials. *Ergodic Theory Dynam. Systems*, 27(3):743–768, 2007.

[8] H. Bruin and M. Todd. Transience and thermodynamic formalism for infinitely branched interval maps. *J. Lond. Math. Soc. (2)*, 86(1):171–194, 2012.

[9] J. Buzzi. Intrinsic ergodicity of smooth interval maps. *Israel J. Math.*, 100:125–161, 1997.
[10] J. Buzzi. Markov extensions for multi-dimensional dynamical systems. *Israel J. Math.*, 112:357–380, 1999.

[11] J. Chen and K. Nguyen. Invariance principles for ergodic systems with slowly $\alpha$-mixing inducing base. In *Differential equations and dynamical systems*, volume 268 of *Springer Proc. Math. Stat.*, pages 115–126. Springer, Cham, 2018.

[12] J. Chen, Y. Yang, and H.-K. Zhang. Non-stationary almost sure invariance principle for hyperbolic systems with singularities. *J. Stat. Phys.*, 172(6):1499–1524, 2018.

[13] N. Chernov. Decay of correlations and dispersing billiards. *J. Statist. Phys.*, 94(3-4):513–556, 1999.

[14] N. Chernov. Advanced statistical properties of dispersing billiards. *J. Stat. Phys.*, 122(6):1061–1094, 2006.

[15] N. Chernov and D. Dolgopyat. Brownian Brownian motion. I. *Mem. Amer. Math. Soc.*, 198(927):viii+193, 2009.

[16] N. Chernov and R. Markarian. *Chaotic billiards*, volume 127 of *Mathematical Surveys and Monographs*. American Mathematical Society, Providence, RI, 2006.

[17] N. Chernov and H.-K. Zhang. Billiards with polynomial mixing rates. *Nonlinearity*, 18(4):1527–1553, 2005.

[18] N. Chernov and H.-K. Zhang. On statistical properties of hyperbolic systems with singularities. *J. Stat. Phys.*, 136(4):615–642, 2009.

[19] J. Dedecker. Inégalités de covariance. *C. R. Math. Acad. Sci. Paris*, 339(7):503–506, 2004.

[20] M. F. Demers and C.Liverani. Stability of statistical properties in two-dimensional piecewise hyperbolic maps. *Trans. Amer. Math. Soc.*, 360(9):4777–4814, 2008.

[21] M. F. Demers and H.-K. Zhang. Spectral analysis of the transfer operator for the Lorentz gas. *J. Mod. Dyn.*, 5(4):665–709, 2011.

[22] M. F. Demers and H.-K. Zhang. A functional analytic approach to perturbations of the Lorentz gas. *Comm. Math. Phys.*, 324(3):767–830, 2013.

[23] M. F. Demers and H.-K. Zhang. Spectral analysis of hyperbolic systems with singularities. *Nonlinearity*, 27(3):379–433, 2014.

[24] D. Dolgopyat. On dynamics of mostly contracting diffeomorphisms. *Comm. Math. Phys.*, 213(1):181–201, 2000.

[25] S. Gouëzel. Stable laws for the doubling map. *Online*, 2008.

[26] S. Gouëzel. Almost sure invariance principle for dynamical systems by spectral methods. *Ann. Probab.*, 38(4):1639–1671, 2010.

[27] N. Haydn, M. Nicol, A. Török, and S. Vaienti. Almost sure invariance principle for sequential and non-stationary dynamical systems. *Trans. Amer. Math. Soc.*, 369(8):5293–5316, 2017.

[28] F. Hofbauer. On intrinsic ergodicity of piecewise monotonic transformations with positive entropy. *Israel J. Math.*, 34(3):213–237 (1980), 1979.

[29] F. Hofbauer. On intrinsic ergodicity of piecewise monotonic transformations with positive entropy. II. *Israel J. Math.*, 38(1-2):107–115, 1981.
[30] F. Hofbauer and G. Keller. Ergodic properties of invariant measures for piecewise monotonic transformations. *Math. Z.*, 180(1):119–140, 1982.

[31] I. A. Ibragimov and Y. V. Linnik. *Independent and stationary sequences of random variables*. Wolters-Noordhoff Publishing, Groningen, 1971. With a supplementary chapter by I. A. Ibragimov and V. V. Petrov, Translation from the Russian edited by J. F. C. Kingman.

[32] G. Keller. Lifting measures to Markov extensions. *Monatsh. Math.*, 108(2-3):183–200, 1989.

[33] G. Keller and C. Liverani. Stability of the spectrum for transfer operators. *Ann. Scuola Norm. Sup. Pisa Cl. Sci. (4)*, 28(1):141–152, 1999.

[34] A. Korepanov. Equidistribution for nonuniformly expanding dynamical systems, and application to the almost sure invariance principle. *Comm. Math. Phys.*, 359(3):1123–1138, 2018.

[35] T. Lindvall. *Lectures on the coupling method*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons, Inc., New York, 1992. A Wiley-Interscience Publication.

[36] C. Liverani. Personal notes for the beijing minicourses. Limit theorem for hyperbolic systems. 2009.

[37] I. Melbourne and M. Nicol. Almost sure invariance principle for nonuniformly hyperbolic systems. *Comm. Math. Phys.*, 260(1):131–146, 2005.

[38] I. Melbourne and M. Nicol. A vector-valued almost sure invariance principle for hyperbolic dynamical systems. *Ann. Probab.*, 37(2):478–505, 2009.

[39] Y. B. Pesin, S. Senti, and K. Zhang. Lifting measures to inducing schemes. *Ergodic Theory Dynam. Systems*, 28(2):553–574, 2008.

[40] Y. B. Pesin and Y. G. Sinaï. Gibbs measures for partially hyperbolic attractors. *Ergodic Theory Dynam. Systems*, 2(3-4):417–438 (1983), 1982.

[41] W. Philipp and W. Stout. Almost sure invariance principles for partial sums of weakly dependent random variables. *Mem. Amer. Math. Soc.*, 2(161):iv+140, 1975.

[42] M. Rychlik. Bounded variation and invariant measures. *Studia Math.*, 76(1):69–80, 1983.

[43] M. Stenlund. A vector-valued almost sure invariance principle for Sinai billiards with random scatterers. *Comm. Math. Phys.*, 325(3):879–916, 2014.

[44] H. Thorisson. *Coupling, stationarity, and regeneration*. Probability and its Applications (New York). Springer-Verlag, New York, 2000.

[45] S. Vaienti and H.-K. Zhang. Optimal bounds on correlation decay rates for nonuniform hyperbolic systems. *Submitted*, 2016.

[46] L.-S. Young. Recurrence times and rates of mixing. *Israel J. Math.*, 110:153–188, 1999.

[47] R. Zweimüller. Kuzmin, coupling, cones, and exponential mixing. *Forum Math.*, 16(3):447–457, 2004.