



**HAL**  
open science

## Rates in almost sure invariance principle for nonuniformly hyperbolic maps

C Cuny, J Dedecker, A Korepanov, F Merlevède

► **To cite this version:**

C Cuny, J Dedecker, A Korepanov, F Merlevède. Rates in almost sure invariance principle for nonuniformly hyperbolic maps. 2024. hal-04168255v2

**HAL Id: hal-04168255**

**<https://hal.science/hal-04168255v2>**

Preprint submitted on 30 Jun 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Rates in almost sure invariance principle for nonuniformly hyperbolic maps

C. Cuny<sup>\*</sup>, J. Dedecker<sup>†</sup>, A. Korepanov<sup>‡</sup>, F. Merlevède<sup>§</sup>

28 June 2024

## Abstract

We prove the Almost Sure Invariance Principle (ASIP) with close to optimal error rates for nonuniformly hyperbolic maps. We do not assume exponential contraction along stable leaves, therefore our result covers in particular slowly mixing invertible dynamical systems as Bunimovich flowers, billiards with flat points as in Chernov and Zhang (2005) and Wojtkowski' (1990) system of two falling balls.

For these examples, the ASIP is a new result, not covered by prior works for various reasons, notably because in absence of exponential contraction along stable leaves, it is challenging to employ the so-called Sinai's trick (Sinai 1972, Bowen 1975) of reducing a nonuniformly hyperbolic system to a nonuniformly expanding one.

Our strategy follows our previous papers on the ASIP for nonuniformly expanding maps, where we build a semiconjugacy to a specific renewal Markov shift and adapt the argument of Berkes, Liu and Wu (2014). The main difference is that now the Markov shift is *two-sided*, the observables depend on the full trajectory, both the future and the past.

## 1 Introduction

Let  $T: M \rightarrow M$  be a nonuniformly hyperbolic dynamical system in the sense of Young [34]. Notable examples of such systems are Axiom A or Hénon-type attractors and various chaotic billiards. Suppose that  $T$  preserves an ergodic physical (Sinai-Ruelle-Bowen) probability measure  $m$ . Let  $\varphi: M \rightarrow \mathbb{R}$  be a Hölder observable with  $\int \varphi dm = 0$  and let  $S_n = \sum_{j=0}^{n-1} \varphi \circ T^j$  be its corresponding Birkhoff sums.

It is common to expect that  $S_n$ , considered as a random process on the probability space  $(M, m)$ , behaves like a Brownian motion. A standard result, proved for a large class of nonuniformly hyperbolic maps under nonrestrictive assumptions, is the *Almost Sure Invariance Principle* (ASIP), see Melbourne and Nicol [25] or Gouëzel [19]. It holds if, without changing the distribution, the process  $(S_n)_{n \geq 0}$  can be defined on a probability space supporting a Brownian motion  $(W_t)_{t \geq 0}$  with variance  $c^2$ , such that with some  $\delta \in (0, 1/2)$ ,

$$S_n = W_n + o(n^\delta) \quad \text{almost surely.} \quad (1.1)$$

In this case we say that the process  $S_n$  satisfies the ASIP with rate  $o(n^\delta)$  and variance  $c^2$ .

---

<sup>\*</sup>Université de Brest, LMBA, UMR CNRS 6205. Email: christophe.cuny@univ-brest.fr

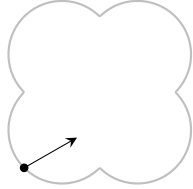
<sup>†</sup>Université Paris Cité, Laboratoire MAP5 (UMR 8145). Email: jerome.dedecker@u-paris.fr

<sup>‡</sup>Loughborough, UK. Email: a.korepanov@lboro.ac.uk

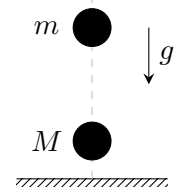
<sup>§</sup>LAMA, Univ Gustave Eiffel, Univ Paris Est Créteil, UMR 8050 CNRS. Email: florence.merlevede@univ-eiffel.fr

The ASIP has a range of useful implications, such as the (functional) central limit theorem or the (functional) law of iterated logarithm, see Berkes and Philipp [5]. The error rates received a significant amount of attention, see e.g. our paper [13] for an overview. See also Zaitsev [35] for a historical overview for sums of independent random vectors.

There are natural nonuniformly hyperbolic dynamical systems where the ASIP is expected but is not covered by existing proofs for one reason or another. Prime examples are Bunimovich flowers [8] and Wojtkowski's system of two falling balls [33].



Flower billiard



Two falling balls

(1.2)

The nonuniformly hyperbolic structure for these maps, under certain natural assumptions, is established by Bálint, Borbély and Varga [2] and Chernov and Zhang [11]. Although the ASIP has not been proven, other statistical properties such as the functional central limit theorem or (iterated) moment bounds are known, see Fleming-Vázquez [17] and Melbourne and Varandas [26].

Another prototypical example is the intermittent baker's map [26, Example 4.1]. It is defined as a transformation of the unit square  $M = [0, 1] \times [0, 1]$  by

$$T(x, y) = \begin{cases} (g(x), g^{-1}(y)) & x \in [0, 1/2], \\ (2x - 1, (y + 1)/2) & x \in (1/2, 1], \end{cases} \quad (1.3)$$

where  $g(x) = x(1 + 2^\alpha x^\alpha)$  and  $\alpha \in (0, 1)$ .

In this paper we prove the ASIP for a class of nonuniformly hyperbolic maps, including (1.2) and (1.3), with rates close to optimum (see Theorem 3.1). Applied to the above examples, our main result is:

**Theorem 1.1.** *Suppose that  $T: M \rightarrow M$  is either (1.3) or one of the maps (1.2) under the assumptions of [2, 11]. Let  $m$  be the unique physical invariant probability measure. Let  $v: M \rightarrow \mathbb{R}$  be Hölder with  $\int v dm = 0$  and  $S_n = \sum_{j < n} v \circ T^j$ . Then, for each  $\varepsilon > 0$ ,*

(a)  $S_n$  satisfies the ASIP with rate  $o(n^{1/3}(\log n)^{4/3+\varepsilon})$  for both maps (1.2). (Possibly, the logarithmic factor can be reduced, see Theorem 3.1 and Remark 3.3.)

(b)  $S_n$  satisfies the ASIP with rate  $o(n^\alpha(\log n)^{\alpha+\varepsilon})$  for the map (1.3) when  $\alpha < 1/2$ .

Our proofs are based on:

- A construction of an extension of  $T$ , which is similar to a Young tower in the sense that it is (topologically) Markov with a tower structure and that the semiconjugacy map is Lipschitz, but with the difference that our extension is also Markov from the measure-theoretical point of view (i.e. can be described by a matrix of transition probabilities).
- A representation of the process  $(v \circ T^j)_{j \geq 0}$ , without changing its distribution (but on another probability space), as

$$\left( \psi(\dots, g_{k-1}, g_k, g_{k+1}, \dots) \right)_{k \in \mathbb{Z}}, \quad (1.4)$$

where  $(g_n)$  is a particular Markov chain and  $\psi$  is a sufficiently regular function of its trajectories. A key tool in the study of the Markov chain is the *meeting time* which is, informally, the time when two independent copies of the Markov chain first meet. Moments of the meeting time are used to control approximations of  $\psi(\dots, g_{j-1}, g_j, g_{j+1}, \dots)$  by functions of finitely many coordinates.

- An adapted argument of Berkes, Liu and Wu [4] which allows to deal with functions of the whole trajectory of a Markov chain instead of a sequence of independent identically distributed random variables as in [4].

This is a continuation and extension of our previous works on nonuniformly expanding maps [13, 14].

*Remark 1.2.* As in [13, 14], our proof of ASIP happens on the level of a Markov chain (1.4). At the same time, in [13, 14] we had to deal with functions  $\psi$  which depend only on future trajectories, i.e.  $\sum_{j=0}^{n-1} \psi(g_j, g_{j+1}, \dots)$ . Working with nonuniformly hyperbolic dynamical systems, we necessarily have to deal with functions of the whole past and future trajectories. This creates problems which go beyond simple modifications of our previous arguments both in the reduction to the Markov chain (Section 2) and in the probabilistic part (Section 3). This can be contrasted with the proof of the ASIP in [4] which is written for a one-sided Bernoulli shift rather than for a two-sided one “to simplify the notation”.

*Remark 1.3.* It is curious that the degenerate case of zero variance is not covered by our general arguments and requires a special and very different treatment (Section 3.6).

*Remark 1.4.* There are prior ideas which may be useful in finding alternative proofs of the ASIP for dynamical systems such as (1.2) and (1.3), including the generalization to vector-valued observables. Notably:

- Melbourne and Nicol [24, 25] prove the ASIP for vector-valued Hölder observables on nonuniformly hyperbolic dynamical systems with uniform contraction along stable leaves. Their proofs work using the so-called Sinai’s trick (after Sinai [32] and Bowen [7], see [26, Introduction]), representing a Hölder observable  $v: M \rightarrow \mathbb{R}$  as  $v = u + \chi \circ T - \chi$ , where  $\chi$  is bounded and  $u$  is Hölder and constant on stable leaves. Then one can work with Birkhoff sums of  $u$  on a quotient system which is nonuniformly expanding, and can be in turn reduced to a Gibbs-Markov map by inducing. For the ASIP, Melbourne and Nicol adapt the results of Philipp and Stout [28] and Kuelbs and Philipp [22].

In the absence of uniform contraction along stable leaves, as is the case for the maps (1.2) and (1.3), Sinai’s trick works but  $\chi$  and  $u$  lose regularity and become challenging to control. Nevertheless, some control is possible. Melbourne and Varandas [26] prove that  $\chi, u \in L^p$  with some  $p > 1$ , and furthermore  $u = m + \psi \circ T - \psi$  where again  $m, \psi \in L^p$  and Birkhoff sums of  $m$  are a reverse martingale.

- Gouëzel [19] proves the ASIP for vector-valued random variables under a condition of certain exponential memory loss for characteristic functions. This condition does not hold for the maps (1.2) or (1.3), but for instance Gouëzel [19, Theorem 2.4] uses it to prove the ASIP for a class of  $L^p$  observables on Gibbs-Markov maps, and as a corollary on nonuniformly expanding maps via inducing.

However, using the above techniques, the error rates would be  $O(n^{-1/4}(\log n)^\alpha)$  for some  $\alpha > 0$  at best, whereas our approach is the only currently available one with better rates.

The paper is organized as follows. In Section 2 we state the assumptions on the nonuniformly hyperbolic maps and prove that they can be modelled using Markov chains of a particular type.

We call such systems Markov shift towers. In Section 3 we prove the ASIP for Hölder observables on nonuniformly hyperbolic maps through that on Markov shift towers. Theorem 1.1 is a corollary of the more general Theorem 3.1.

We use the notation  $\mathbb{N} = \{0, 1, \dots\}$  and  $\mathbb{N}_0 = \{1, 2, \dots\}$ .

## 2 Nonuniformly hyperbolic maps as Markov shift towers

In this section we give a standard definition of nonuniformly hyperbolic maps, introduce Markov shift towers and show that every nonuniformly hyperbolic map  $T: M \rightarrow M$  with invariant measure  $m$  has an extension  $f: \Delta \rightarrow \Delta$  with invariant measure  $\mathbb{P}$  which is a Markov shift tower, and the semiconjugacy  $\pi: \Delta \rightarrow M$  is Lipschitz.

$$\begin{array}{ccc} \Delta & \xrightarrow{f} & \Delta \\ \downarrow \pi & & \downarrow \pi \\ M & \xrightarrow{T} & M \end{array} \qquad \begin{array}{c} \mathbb{P} \\ \downarrow (\pi)_* \\ m \end{array}$$

We also quantify the *return time tails* of  $f$  depending on those of  $T$ . This result is stated formally in Theorem 2.4, but first we need a few pages of notation.

### 2.1 Notation and result

#### 2.1.1 Nonuniformly hyperbolic maps

Let  $(M, d_0)$  be a bounded metric space with a Borel probability measure  $m$  such that  $(M, m)$  is a Lebesgue space. Let  $T: M \rightarrow M$  be measurable for the Borel sigma-field. We assume that  $T$  preserves  $m$ , is ergodic and is nonuniformly hyperbolic in the following sense:

- There is a measurable subset  $Y \subset M$ ,  $m(Y) > 0$ , with an at most countable partition  $\alpha$  and a return time  $\tau: Y \rightarrow \{1, 2, \dots\}$  that is constant on each  $a \in \alpha$  with value  $\tau(a)$  and  $T^{\tau(y)}(y) \in Y$  for all  $y \in Y$ . We denote by  $F: y \mapsto T^{\tau(y)}(y)$  the induced map. There is an  $F$ -invariant probability measure  $\mu$  on  $Y$  such that  $\int \tau d\mu < \infty$  and which agrees with  $m$  in the usual for induced maps sense:

$$m = \left( \int \tau d\mu \right)^{-1} \sum_{a \in \alpha} \sum_{k=0}^{\tau(a)-1} (T^k)_* \mu_a,$$

where  $\mu_a$  is the restriction of  $\mu$  on  $a$ .

- Coding of orbits under  $F$  by elements of  $\alpha$  is non-pathological in the sense that the set of  $(\dots, a_{-1}, a_0, a_1, \dots) \in \alpha^{\mathbb{Z}}$ , for which there exists  $(\dots, y_{-1}, y_0, y_1, \dots) \in Y^{\mathbb{Z}}$  such that  $y_n \in a_n$  and  $F(y_n) = y_{n+1}$ , is measurable in  $\alpha^{\mathbb{Z}}$  with the product topology and Borel sigma-algebra ( $\alpha$  is equipped with the discrete topology).
- For  $y, z \in Y$  we define the separation time  $s(y, z)$  as the least  $n \geq 0$  such that  $F^n(y)$  and  $F^n(z)$  belong to different elements of  $\alpha$ , or  $\infty$  if such  $n$  does not exist. There are constants  $K \geq 1$  and  $\gamma \in (0, 1)$  such that:

- If  $y, z \in Y$ , then for all  $j \geq 0$ ,

$$d_0(F^j(y), F^j(z)) \leq K(\gamma^{s(y,z)-j} + \gamma^j). \quad (2.1)$$

– If  $y, z \in a$ ,  $a \in \alpha$ , then for all  $0 \leq j \leq \tau(a)$ ,

$$d_0(T^j y, T^j z) \leq K [d_0(y, z) + d_0(F(y), F(z))] \quad (2.2)$$

- There is a partition  $\mathcal{W}^s$  of  $Y$  into “stable leaves”. The stable leaves are invariant under  $F$  meaning that  $F(W_y^s) \subset W_{F(y)}^s$  for all  $y \in Y$ , where  $W_y^s$  is the stable leaf containing  $y$ . Each  $a \in \alpha$  is a union of stable leaves, i.e.  $\mathcal{W}_s$  is a refinement of  $\alpha$ .

Let  $\bar{Y} = Y / \sim$ , where  $y \sim z$  if  $y \in W_z^s$ , and let  $\bar{\pi}: Y \rightarrow \bar{Y}$  be the natural projection. Since the stable leaves are invariant under  $F$  and  $\mathcal{W}^s$  is a refinement of  $\alpha$ , we obtain a well defined quotient map  $\bar{F}: \bar{Y} \rightarrow \bar{Y}$  with a partition  $\bar{\alpha}$  of  $\bar{Y}$  and separation time  $s$ .

We suppose that the probability measure  $\bar{\mu} = \bar{\pi}_* \mu$  on  $\bar{Y}$  is invariant under  $\bar{F}$ . Moreover, for each  $a \in \bar{\alpha}$ , the restriction  $\bar{F}: a \rightarrow \bar{Y}$  is a bijection (modulo  $\bar{\mu}$  zero measure), and its inverse Jacobian  $\zeta_a = \frac{d\bar{\mu}}{d\bar{\mu} \circ \bar{F}}$  satisfies

$$|\log \zeta_a(y) - \log \zeta_a(z)| \leq K \gamma^{s(y,z)} \quad \text{for all } y, z \in a.$$

In other words, the quotient map  $\bar{F}$  is full branch Gibbs-Markov.

### 2.1.2 Markov shift towers

A closely related class of nonuniformly hyperbolic maps is what we call a *Markov shift tower*. The main difference is that the induced map, denoted by  $f_X$  below, is a Bernoulli shift, both topologically and measure-theoretically.

We say that  $f: \Delta \rightarrow \Delta$  is a *Markov shift tower* if it has the following structure.

- We are given a finite or countable probability space  $(\mathcal{A}, \mathbb{P}_{\mathcal{A}})$ , an integrable function  $h_{\mathcal{A}}: \mathcal{A} \rightarrow \{1, 2, \dots\}$  and a constant  $\xi \in (0, 1)$ . These define the rest of the construction.
- Let  $(X, \mathbb{P}_X) = (\mathcal{A}^{\mathbb{Z}}, \mathbb{P}_{\mathcal{A}}^{\mathbb{Z}})$  be the product probability space and let  $f_X: X \rightarrow X$  be the left shift

$$f_X(\dots, a_{-1}, a_0, a_1, \dots) = (\dots, a_0, a_1, a_2, \dots).$$

Define  $h: X \rightarrow \{1, 2, \dots\}$  by  $h(\dots, a_{-1}, a_0, a_1, \dots) = h_{\mathcal{A}}(a_0)$ .

- The map  $f: \Delta \rightarrow \Delta$  is a suspension over  $f_X: X \rightarrow X$  with a roof function  $h$ , i.e.

$$\begin{aligned} \Delta &= \{(x, \ell) \in X \times \mathbb{Z} : 0 \leq \ell < h(x)\} \\ f(x, \ell) &= \begin{cases} (x, \ell + 1), & \ell < h(x) - 1 \\ (f_X(x), 0), & \ell = h(x) - 1 \end{cases} \end{aligned} \quad (2.3)$$

- Let  $\mathbb{P}$  be the probability measure on  $\Delta$  defined by

$$\mathbb{P}(A \times \{\ell\}) = \left( \int h d\mathbb{P}_X \right)^{-1} \mathbb{P}_X(A) \quad \text{for all } \ell \geq 0 \quad \text{and} \quad A \subset \{x \in X : h(x) \geq \ell + 1\}.$$

Since  $\mathbb{P}_X$  is  $f_X$ -invariant, note that  $\mathbb{P}$  is  $f$ -invariant.

- Define a distance  $d$  on  $X$  by  $d(x, y) = \xi^{s(x, y)}$ , where  $s: X \times X \rightarrow \{0, 1, \dots\}$  is the separation time,

$$s((\dots, a_{-1}, a_0, a_1, \dots), (\dots, b_{-1}, b_0, b_1, \dots)) = \inf\{j \geq 0 : a_j \neq b_j \text{ or } a_{-j} \neq b_{-j}\}.$$

Let  $d$  also denote the related distance on  $\Delta$ :

$$d((x, k), (y, j)) = \begin{cases} 1, & k \neq j \\ d(x, y), & k = j \end{cases}. \quad (2.4)$$

*Remark 2.1.* As a part of the definition of a Markov shift tower,  $f: \Delta \rightarrow \Delta$  is Markov both topologically (as a suspension over a Bernoulli shift, it has a natural Markov partition) and probabilistically (the invariant measure  $\mathbb{P}$  is Markov, it can be described by transition probabilities). Also  $\Delta$  is a metric space with metric  $d$ .

*Remark 2.2.* We used similar Markov shift towers in our previous works [13, 14, 21]. The main difference is that they were *one-sided*: the base map was the left shift  $f_X: X \rightarrow X$  on  $X = \mathcal{A}^{\mathbb{N}}$  instead of  $X = \mathcal{A}^{\mathbb{Z}}$  here. A two-sided Markov shift tower is the natural extension [30, 31] of the corresponding one-side tower.

*Remark 2.3.* An important characteristic of both nonuniformly hyperbolic maps and Markov shift towers is tail of the return times, i.e. the asymptotics of  $\mu(\tau \geq n)$  and  $\mathbb{P}_X(h \geq n)$  respectively.

### 2.1.3 Main result

**Theorem 2.4.** *Suppose that  $T: M \rightarrow M$  is a nonuniformly hyperbolic map as in Section 2.1.1. Then there exists a Markov shift tower  $f: \Delta \rightarrow \Delta$  and a map  $\pi: \Delta \rightarrow M$ , defined  $\mathbb{P}$ -almost surely on  $\Delta$ , such that:*

- $\pi$  is a semiconjugacy, i.e.  $\pi \circ f = T \circ \pi$ ,
- $\pi$  is measure preserving, i.e.  $\pi_* \mathbb{P} = m$ ,
- $\pi$  is Lipschitz.

Moreover, the return time tails of  $f$  are closely related to those of  $T$ :

- if  $\mu(\tau \geq n) = O(n^{-\beta}L(n))$  with  $\beta > 0$  and  $L$  a slowly varying function at infinity, then  $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} \geq n) = O(n^{-\beta}L(n))$ ;
- if  $\int \tau^\beta d\mu < \infty$  with  $\beta > 0$ , then  $\int h_{\mathcal{A}}^\beta d\mathbb{P}_{\mathcal{A}} < \infty$ ;
- if  $\int e^{\beta\tau^\delta} d\mu < \infty$  with  $\beta > 0$  and  $\delta \in (0, 1]$ , then  $\int e^{\beta' h_{\mathcal{A}}^\delta} d\mathbb{P}_{\mathcal{A}} < \infty$  for some  $0 < \beta' \leq \beta$ .

## 2.2 Proof of Theorem 2.4

Our strategy is to reduce the proof to an application of [21, Theorem 3.4]. There, a one-sided Markov shift tower is constructed for a nonuniformly expanding map. Then  $T$  can be considered via the natural extension of such a map, and the corresponding natural extension of the one-sided Markov shift in [21] is a two-sided Markov shift, as required for Theorem 2.4.

Let  $\sigma: \alpha^{\mathbb{Z}} \rightarrow \alpha^{\mathbb{Z}}$  be the left shift. We supply  $\alpha^{\mathbb{Z}}$  with the product topology and the Borel sigma-algebra. For  $x = (x_k)_{k \in \mathbb{Z}}$  and  $y = (y_k)_{k \in \mathbb{Z}}$  in  $\alpha^{\mathbb{Z}}$ , define the separation time and distance by

$$\begin{aligned} s(x, y) &= \inf\{k \geq 0 : x_k \neq y_k \text{ or } x_{-k} \neq y_{-k}\}, \\ d(x, y) &= \gamma^{s(x, y)}. \end{aligned}$$

Here  $\gamma$  is the constant in (2.1).

### 2.2.1 Tower over full shift

Our first step is to model  $T: M \rightarrow M$  as a suspension tower over  $\sigma: \alpha^{\mathbb{Z}} \rightarrow \alpha^{\mathbb{Z}}$ . It is natural that this can be done, yet this requires some care.

**Proposition 2.5.** *There exists a probability measure  $\mu_\alpha$  on  $\alpha^{\mathbb{Z}}$  and a  $\mu_\alpha$ -almost everywhere defined map  $\chi_\alpha: \alpha^{\mathbb{Z}} \rightarrow Y$  such that on the domain of  $\chi_\alpha$ :*

- $\chi_\alpha$  is Lipschitz:  $d_0(\chi_\alpha(x), \chi_\alpha(y)) \leq 2Kd(x, y)$ ,
- $(\chi_\alpha)_* \mu_\alpha = \mu$ ,
- $F \circ \chi_\alpha = \chi_\alpha \circ \sigma$ ,
- $\chi_\alpha(\dots, a_{-1}, a_0, a_1, \dots) \in a_0$ .

*Proof.* In order to construct the natural extension [30, 31] of  $F: Y \rightarrow Y$ , we assume that  $F(Y) = Y$ . This comes without loss of generality: otherwise instead of  $Y$  we work with  $\bigcap_{n \geq 0} F^{-n}(Y)$ . This is a full measure subset of  $Y$  on which  $F$  is surjective, and the proof goes through with straightforward and minor changes.

Let  $\check{F}: \check{Y} \rightarrow \check{Y}$  be the natural extension [30, 31] of  $F: Y \rightarrow Y$ :

$$\check{Y} = \{(\dots, y_{-1}, y_0, y_1, \dots) \in Y^{\mathbb{Z}} : y_{n+1} = F(y_n)\},$$

and  $\check{F}$  is the left shift. The topology on  $\check{Y}$  is generated by open cylinders of the type  $\{y_i \in E\}$  with open  $E \subset Y$  (in the induced topology on  $Y$  as a subspace of  $M$ ), and we consider  $\check{Y}$  with the Borel sigma-algebra. Let  $\check{\pi}_\ell: \check{Y} \rightarrow Y$ ,  $(\dots, y_{-1}, y_0, y_1, \dots) \mapsto y_\ell$  be the natural projections, set  $\check{\pi} = \check{\pi}_0$ , and let  $\check{\mu}$  be the unique  $\check{F}$ -invariant probability measure on  $\check{Y}$  such that  $\check{\pi}_* \check{\mu} = \mu$ .

Define  $\iota: \check{Y} \rightarrow \alpha^{\mathbb{Z}}$  by  $(\dots, y_{-1}, y_0, y_1, \dots) \mapsto (\dots, a_{-1}, a_0, a_1, \dots)$  where  $y_n \in a_n$ . Then  $\iota$  is a measurable and injective map. Set  $\mu_\alpha = \iota_* \check{\mu}$ . Using (2.1), for  $0 \leq |\ell| \leq n$ ,

$$\begin{aligned} \text{diam}((\check{\pi}_\ell \circ \iota^{-1})([a_{-n}, \dots, a_n])) &= \text{diam}(F^{n+\ell}(\{y \in Y : F^k(y) \in a_{-n+k} \text{ for all } 0 \leq k \leq 2n\})) \\ &\leq 2K\gamma^{n-|\ell|} = 2K\gamma^{-|\ell|} \text{diam}([a_{-n}, \dots, a_n]), \end{aligned}$$

where  $[a_{-n}, \dots, a_n] \subset \alpha^{\mathbb{Z}}$  is a cylinder. Hence  $\check{\pi}_\ell \circ \iota^{-1}$  is Lipschitz with Lipschitz constant  $2K\gamma^{-|\ell|}$ . Since  $\iota^{-1}(x) = ((\check{\pi}_\ell \circ \iota^{-1})(x))_{\ell \in \mathbb{Z}}$  and each  $\check{\pi}_\ell \circ \iota^{-1}$  is continuous,  $\iota^{-1}: \iota(\check{Y}) \rightarrow \check{Y}$  is continuous. Recall that, as a part of the definition of  $T$  as a nonuniformly hyperbolic map we assumed that  $\iota(\check{Y})$  is measurable in  $\alpha^{\mathbb{Z}}$ . Then  $\iota^{-1}$  is measurable.



Observe that  $\iota^{-1}$  is a conjugacy between  $\check{F}$  and  $\sigma$ . Set  $\chi_\alpha = \check{\pi} \circ \iota^{-1}$ .

$$\begin{array}{ccc}
\alpha^{\mathbb{Z}} & \xrightarrow{\sigma} & \alpha^{\mathbb{Z}} \\
\downarrow \iota^{-1} & & \downarrow \iota^{-1} \\
\check{Y} & \xrightarrow{\check{F}} & \check{Y} \\
\downarrow \check{\pi} & & \downarrow \check{\pi} \\
Y & \xrightarrow{F} & Y
\end{array}
\qquad
\begin{array}{c}
\mu_\alpha \\
\downarrow (\iota^{-1})_* \\
\check{\mu} \\
\downarrow (\check{\pi})_* \\
\mu
\end{array}$$

Then  $\chi_\alpha$  is a measure preserving semiconjugacy between  $\sigma$  and  $F$  with the required properties.  $\square$

In the setup of Proposition 2.5, consider the suspension map over  $\sigma$  with roof function  $\tau_\alpha = \tau \circ \chi_\alpha$ , i.e. the space

$$M_\alpha = \{(x, \ell) : x \in \alpha^{\mathbb{Z}}, 0 \leq \ell < \tau_\alpha(x)\}$$

with the transformation

$$T_\alpha : (x, \ell) \mapsto \begin{cases} (x, \ell + 1), & \ell < \tau_\alpha(x) - 1, \\ (\sigma(x), 0), & \text{else.} \end{cases}$$

As in (2.4), we supply  $M_\alpha$  with the metric

$$d_\alpha((x, k), (y, \ell)) = \begin{cases} 1, & k \neq \ell, \\ d(x, y), & k = \ell. \end{cases}$$

Let  $m_\alpha = \mu_\alpha \times \text{counting/normalization}$  be the probability measure on  $M_\alpha$ , and let  $\pi_\alpha : M_\alpha \rightarrow M$ ,

$$\pi_\alpha : (x, \ell) \mapsto T^\ell(\chi_\alpha(x)).$$

*Remark 2.6.* Observe that:

- $T_\alpha : M_\alpha \rightarrow M_\alpha$  is a nonuniformly hyperbolic map with the same tails of return times as  $T$ , namely  $\mu_\alpha(\tau_\alpha \geq n) = \mu(\tau \geq n)$  for all  $n$ .
- $\pi_\alpha$  is defined  $m_\alpha$ -almost surely, but possibly not on the whole  $M_\alpha$ .
- $\pi_\alpha : M_\alpha \rightarrow M$  is a measure preserving semiconjugacy between  $T_\alpha$  and  $T$  (modulo zero measure).
- $\pi_\alpha : M_\alpha \rightarrow M$  is Lipschitz (on its domain). Indeed,

$$d_0(\pi_\alpha(x, \ell), \pi_\alpha(y, k)) \leq Cd((x, \ell), (y, k))$$

holds:

- with  $C = \text{diam } M$  when  $\ell \neq k$  by construction,
- with  $C = 2K$  when  $\ell = k = 0$  by Proposition 2.5,
- with  $C = 2K^2(1 + \gamma^{-1})$  when  $\ell = k > 0$ . Indeed, by (2.2),

$$d_0(\pi_\alpha(x, \ell), \pi_\alpha(y, \ell)) \leq K[d_0(\chi_\alpha(x), \chi_\alpha(y)) + d_0(F \circ \chi_\alpha(x), F \circ \chi_\alpha(y))].$$

Now, by Proposition 2.5,  $F \circ \chi_\alpha = \chi_\alpha \circ \sigma$  and  $\chi_\alpha$  is Lipschitz. Then

$$d_0(\pi_\alpha(x, \ell), \pi_\alpha(y, \ell)) \leq 2K^2[d(x, y) + d(\sigma(x), \sigma(y))].$$

To conclude, we use the fact that since  $s(x, y) \leq 1 + s(\sigma(x), \sigma(y))$ ,  $d(\sigma(x), \sigma(y)) \leq \gamma^{-1}d(x, y)$ .

### 2.2.2 Proof of Theorem 2.4 for $T_\alpha$

We prove Theorem 2.4 for  $T_\alpha$  instead of  $T$ , and Remark 2.6 guarantees that the result carries over to  $T$ . We construct a number of intermediate spaces.

Let  $\mathcal{A}$  denote the set of all finite words in the alphabet  $\alpha$ , not including the empty word. For  $w = a_0 \cdots a_{n-1} \in \mathcal{A}$  let  $h_{\mathcal{A}}(w) = \tau(a_0) + \cdots + \tau(a_{n-1})$ .

Define  $\pi_{\mathcal{A}}: \mathcal{A}^{\mathbb{Z}} \rightarrow \alpha^{\mathbb{Z}}$  by  $\pi_{\mathcal{A}}(\dots, w_{-1}, w_0, w_1, \dots) = \cdots w_{-1}w_0w_1 \cdots$ , which is a concatenation with the first letter of  $w_0$  at index 0. Similarly define  $\bar{\pi}_{\mathcal{A}}: \mathcal{A}^{\mathbb{N}} \rightarrow \alpha^{\mathbb{N}}$  by  $\bar{\pi}_{\mathcal{A}}(w_0, w_1, \dots) = w_0w_1 \cdots$ .

Let  $\pi_\alpha^+: \alpha^{\mathbb{Z}} \rightarrow \alpha^{\mathbb{N}}$  be the projection on nonnegative coordinates. Using that  $\tau_\alpha$  depends only on the ‘‘future’’ coordinates, define  $\bar{\tau}_\alpha: \alpha^{\mathbb{N}} \rightarrow \mathbb{N}_0$  so that  $\tau(x) = \bar{\tau}_\alpha(\pi_\alpha^+(x))$ .

Let  $\bar{M}_\alpha = \{(x, \ell) \in \alpha^{\mathbb{N}} \times \mathbb{Z} : 0 \leq \ell < \bar{\tau}_\alpha(x)\}$ , and let

$$\bar{T}_\alpha: (x, \ell) \mapsto \begin{cases} (x, \ell + 1) & \ell < \bar{\tau}_\alpha(y) - 1, \\ (\sigma(x), 0) & \ell = \bar{\tau}_\alpha(y) - 1. \end{cases}$$

Then  $\bar{T}_\alpha: \bar{M}_\alpha \rightarrow \bar{M}_\alpha$  is a nonuniformly *expanding* system. It preserves the probability measure  $\bar{m}_\alpha = \bar{\mu}_\alpha \times \text{counting/normalization}$ , where  $\bar{\mu}_\alpha = (\pi_\alpha^+)_* \mu_\alpha$  is the projection of  $\mu_\alpha$  on nonnegative coordinates.

Let  $\psi: M_\alpha \rightarrow \bar{M}_\alpha$  be the natural projection,

$$\psi: (x, \ell) \mapsto (\pi_\alpha^+(x), \ell).$$

*Remark 2.7.*  $T_\alpha$  is the natural extension of  $\bar{T}_\alpha$ . In particular,  $\psi$  is a semiconjugacy, and  $\bar{m}_\alpha$  is the unique  $\bar{T}_\alpha$ -invariant probability measure so that  $\psi_* m_\alpha = \bar{m}_\alpha$ .

To use the same notations as in [21], let  $\xi = \gamma$ . By [21, Theorem 3.4, Section 4.1], there exists a probability measure  $\mathbb{P}_{\mathcal{A}}$  on  $\mathcal{A}$ , such that the *one-sided* Markov shift tower (see Remark 2.2)  $\bar{f}: \bar{\Delta} \rightarrow \bar{\Delta}$ , defined by  $\{(\mathcal{A}, \mathbb{P}_{\mathcal{A}}), h_{\mathcal{A}}, \xi\}$ , is an extension of  $\bar{T}_\alpha$  with the measure preserving and Lipschitz semiconjugacy  $\bar{\pi}_\alpha: \bar{\Delta} \rightarrow \bar{M}_\alpha$ ,

$$\bar{\pi}_\alpha: (x, \ell) \mapsto \bar{T}_\alpha^\ell(\bar{\pi}_{\mathcal{A}}(x), 0).$$

Next let  $\bar{\mathbb{P}}$  denote be the  $\bar{f}$ -invariant probability measure on  $\bar{\Delta}$ .

Now let  $f: \Delta \rightarrow \Delta$  be the *two-sided* (i.e. as in Section 2.1) Markov shift tower defined by the same objects  $\{(\mathcal{A}, \mathbb{P}_{\mathcal{A}}), h_{\mathcal{A}}, \xi\}$ , with the invariant probability measure  $\mathbb{P}$ .

Let  $\bar{\pi}_\Delta: \Delta \rightarrow \bar{\Delta}$  be the natural projection,

$$\bar{\pi}_\Delta: (x, \ell) \mapsto (\bar{\pi}_{\mathcal{A}}^+(x), \ell),$$

where  $\bar{\pi}_{\mathcal{A}}^+: \mathcal{A}^{\mathbb{Z}} \rightarrow \alpha^{\mathbb{N}}$  is defined by  $\bar{\pi}_{\mathcal{A}}^+(\dots, w_{-1}, w_0, w_1, \dots) = w_0w_1w_2 \cdots$ . Observe that  $\bar{\pi}_\Delta$  is a measure preserving and Lipschitz semiconjugacy between  $f$  and  $\bar{f}$ .

*Remark 2.8.*  $\bar{\pi}_\alpha \circ \bar{\pi}_\Delta: \Delta \rightarrow \bar{M}_\alpha$  is a measure preserving semiconjugacy, in particular

$$(\bar{\pi}_\alpha \circ \bar{\pi}_\Delta)_* \mathbb{P} = \bar{m}_\alpha.$$

Define  $\pi_\alpha: \Delta \rightarrow M_\alpha$ ,

$$\pi_\alpha: (x, \ell) \mapsto T_\alpha^\ell(\pi_{\mathcal{A}}(x), 0).$$

Observe that  $\pi_\alpha$  is a measure preserving semiconjugacy too.

Now,  $f: (\Delta \rightarrow \Delta)$  preserving the probability measure  $\mathbb{P}$ , with the semiconjacy  $\pi_\alpha$  to  $T_\alpha$  ( $M_\alpha \rightarrow M_\alpha$ ), is the Markov shift tower we are after. It remains to verify that it has the required properties: that  $\pi_\alpha$  is measure preserving and Lipschitz, and to bound the return time tails.

**Proposition 2.9.**  $\pi_\alpha$  is measure preserving:  $(\pi_\alpha)_* \mathbb{P} = m_\alpha$ .

*Proof.* We have constructed four dynamical systems:  $f: \Delta \rightarrow \Delta$  with invariant measure  $\mathbb{P}$ ,  $\bar{f}: \bar{\Delta} \rightarrow \bar{\Delta}$  with invariant measure  $\bar{\mathbb{P}}$ ,  $T_\alpha: M_\alpha \rightarrow M_\alpha$  with invariant measure  $m_\alpha$ , and  $\bar{T}_\alpha: \bar{M}_\alpha \rightarrow \bar{M}_\alpha$  with invariant measure  $\bar{m}_\alpha$ . They are connected by semiconjugacies  $\pi_\alpha: \Delta \rightarrow M_\alpha$ ,  $\bar{\pi}_\alpha: \bar{\Delta} \rightarrow \bar{M}_\alpha$ ,  $\bar{\pi}_\Delta: \bar{\Delta} \rightarrow \Delta$  and  $\psi: M_\alpha \rightarrow \bar{M}_\alpha$ , where  $\psi$  is the natural projection.

The left diagram commutes, and we have to justify the dashed arrow in the right diagram:

$$\begin{array}{ccc}
 \Delta & & \mathbb{P} \\
 \pi_\alpha \downarrow & \searrow \bar{\pi}_\Delta & \downarrow (\bar{\pi}_\Delta)_* \\
 & \bar{\Delta} & \bar{\mathbb{P}} \\
 & \searrow \bar{\pi}_\alpha & \downarrow (\bar{\pi}_\alpha)_* \\
 M_\alpha & \xrightarrow{\psi} & \bar{M}_\alpha \\
 & & \downarrow (\pi_\alpha)_* \\
 & & m_\alpha \\
 & & \xrightarrow{\psi_*} \bar{m}_\alpha
 \end{array} \tag{2.5}$$

By Remark 2.7,  $m_\alpha$  is the *unique* probability measure on  $M_\alpha$  so that  $\psi_* m_\alpha = \bar{m}_\alpha$ . On the other hand,  $\psi \circ \pi_\alpha = \bar{\pi}_\Delta \circ \bar{\pi}_\alpha$ , and therefore  $\psi_*((\pi_\alpha)_* \mathbb{P}) = \bar{m}_\alpha$ , so  $(\pi_\alpha)_* \mathbb{P} = m_\alpha$  as required.  $\square$

By the same argument as in [21, Proposition 4.18],  $\pi_\alpha$  is Lipschitz.

The return time tails on  $\Delta$  and  $\bar{\Delta}$  are equal by construction, and [21, Theorem 3.4] proves that on  $\bar{\Delta}$  they are as required for Theorem 2.4 except for the more general than in [21] polynomial case  $\mu_\alpha(\tau_\alpha \geq n) = O(n^{-\beta} L(n))$  with  $\beta > 0$  and  $L$  a slowly varying function at infinity. In turn, the relation between the tails in [21] is taken from [20, Section 4]. To extend the results as we require, it is sufficient to notice that, by taking into account the properties of slowly varying functions (see for instance [6, Chapter 1]), the following version of [20, Proposition 4.4] holds:

**Proposition 2.10** (in notation of [20]). *Suppose that there exist  $C_\tau > 0$ ,  $\beta > 0$ ,  $L$  a slowly varying function at infinity and  $\ell_0 \geq 1$ , such that  $m(\tau \geq \ell) \leq C_\tau \ell^{-\beta} L(\ell)$  for all  $\ell \geq \ell_0$ . Then  $\mathbb{P}(t \geq \ell) \leq C \ell^{-\beta} L(\ell)$ .*

The proof of Theorem 2.4 for  $T_\alpha$  is complete, and following Remark 2.6 we recover the full result.

### 3 ASIP for Hölder observables of nonuniformly hyperbolic maps

Theorem 1.1 is an application of the following general result, which is the goal of this section:

**Theorem 3.1.** *Let  $T: M \rightarrow M$  be an ergodic, measure-preserving transformation defined on a bounded metric space  $(M, d)$  with Borel probability measure  $m$ . Let  $\varphi: M \rightarrow \mathbb{R}$  be Hölder continuous and centered, i.e.  $\int \varphi dm = 0$ . Let  $S_n(\varphi) = \sum_{k=0}^{n-1} \varphi \circ T^k$ . Suppose that  $T$  is nonuniformly hyperbolic (in the sense of Section 2.1.1) with uniformly hyperbolic induced map  $F$ , return time  $\tau$  and  $F$ -invariant measure  $\mu$  associated with a subset  $Y$  of  $M$ . Assume that  $\int \tau^2 d\mu < \infty$ . Then the limit  $c^2 = \lim_{n \rightarrow \infty} n^{-1} \int |S_n(\varphi)|^2 d\mu$  exists. In addition,*

- (a) *If  $\mu(\tau \geq n) = O(n^{-\beta} (\log n)^\gamma)$ , with  $\beta > 2$  and  $\gamma \in \mathbb{R}$ , then for each  $\varepsilon > 0$  the process  $S_n$  satisfies the ASIP with variance  $c^2$  and rate  $o(n^{1/\beta} (\log n)^{(\gamma+1)/\beta+\varepsilon})$ .*
- (b) *If  $\int \tau^\beta d\mu < \infty$  with  $\beta > 2$ , then  $S_n$  satisfies the ASIP with variance  $c^2$  and rate  $o(n^{1/\beta})$ .*

(c) If  $\int e^{\beta\tau^\delta} d\mu < \infty$  with  $\beta > 0$  and  $\delta \in (0, 1]$ , then  $S_n$  satisfies the ASIP with variance  $c^2$  and rate  $O(\log n)^{1+1/\delta}$ .

*Remark 3.2.* The dependence of the ASIP on  $\varepsilon$  in item (a) should be understood as follows: for every  $\varepsilon > 0$ , there exists a Brownian motion  $(W_t)_{t>0}$  with variance  $c^2$  such that  $S_n = W_n + o(n^{1/\beta}(\log n)^{(\gamma+1)/\beta+\varepsilon})$  almost surely.

For the flower billiard map and the two falling balls map defined in (1.2), it has been proved in [11] that they are examples of nonuniformly hyperbolic maps with tails of the return times of the form  $O(n^{-3}(\log n)^3)$ . On another hand, for the intermittent Baker's map as described in (1.3), it has been proved in [26] that the associated return times have tails  $\sim n^{-\beta}$  with  $\beta = 1/\alpha$ . These considerations together with Theorem 3.1 prove Theorem 1.1 of the introduction.

*Remark 3.3.* It is expected [3] that for the maps (1.2) the tails of the return times can be improved to  $O(n^{-3})$ , dropping the logarithmic factor, similarly to what was done for Buminovich stadia and related systems [12]. This would improve our result.

Let  $f: \Delta \rightarrow \Delta$  be a Markov shift tower constructed for  $T$  as in Theorem 2.4, built with a probability space  $(\mathcal{A}, \mathbb{P}_{\mathcal{A}})$ , roof function  $h_{\mathcal{A}}$  and the metric constant  $\xi \in (0, 1)$ . Recall the associated notations, in particular that  $(\Delta, d)$  is a metric space and that  $f$  preserves the probability measure  $\mathbb{P}$ .

By Theorem 2.4, the process  $(\varphi \circ T^k)_k$  defined on  $(M, m)$  and the process  $(v \circ f^k)_k$ , where  $v = \varphi \circ \pi$ , defined on  $(\Delta, \mathbb{P})$  have the same law. Moreover, if  $\varphi$  is Hölder continuous then so is  $v$  since  $\pi$  is Lipschitz. Hence Theorem 3.1 will follow from its equivalent result stated for Markov shift towers as given in Proposition 3.4 below.

### 3.1 ASIP for Markov shift towers

**Proposition 3.4.** *Suppose that  $v: \Delta \rightarrow \mathbb{R}$  is Hölder continuous and centered, i.e.  $\int v d\mathbb{P} = 0$ . Let  $S_n = \sum_{k < n} v \circ f^k$ . Assume that  $\int h_{\mathcal{A}}^2 d\mathbb{P}_{\mathcal{A}} < \infty$ . Then  $c^2 = \lim_{n \rightarrow \infty} n^{-1} \int |S_n(\varphi)|^2 d\mu$  exists. In addition,*

- (a) *If  $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} \geq n) = O(n^{-\beta}(\log n)^\gamma)$ , with  $\beta > 2$  and  $\gamma \in \mathbb{R}$ , then  $S_n$  satisfies the ASIP with variance  $c^2$  and rate  $o(n^{1/\beta}(\log n)^{(\gamma+1)/\beta+\varepsilon})$  for each  $\varepsilon > 0$ .*
- (b) *If  $\int h_{\mathcal{A}}^\beta d\mathbb{P}_{\mathcal{A}} < \infty$  with  $\beta > 2$ , then  $S_n$  satisfies the ASIP with variance  $c^2$  and rate  $o(n^{1/\beta})$ .*
- (c) *If  $\int e^{\beta h_{\mathcal{A}}^\delta} d\mathbb{P}_{\mathcal{A}} < \infty$  with  $\beta > 0$  and  $\delta \in (0, 1]$ , then  $S_n$  satisfies the ASIP with variance  $c^2$  and rate  $O(\log n)^{1+1/\delta}$ .*

The rest of this section is devoted to the proof of Proposition 3.4.

### 3.2 The associated Markov chain

It is convenient to represent the dynamics on a Markov shift tower as a Markov chain in the conventional sense. For this, let  $G = \{(a, \ell) \in \mathcal{A} \times \mathbb{Z} : 0 \leq \ell < h_{\mathcal{A}}(a)\}$  and let  $\mathcal{G} \subset G^{\mathbb{Z}}$  be the set of admissible symbolic trajectories:

$$\mathcal{G} = \{g = (g_n)_{n \in \mathbb{Z}} \in G^{\mathbb{Z}} : \text{if } g_n = (a, \ell) \text{ with } \ell < h(a) - 1, \text{ then } g_{n+1} = (a, \ell + 1)\}.$$

Let now  $(g_n)_{n \geq 0}$  be a Markov chain with state space  $G$  and transition probabilities

$$\begin{aligned} & \mathbb{P}(g_{n+1} = (a, \ell) \mid g_n = (a', \ell')) \\ &= \begin{cases} 1, & \ell = \ell' + 1 \text{ and } \ell' + 1 < h(a) \text{ and } a = a', \\ \mathbb{P}_{\mathcal{A}}(a), & \ell = 0 \text{ and } \ell' + 1 = h(a'), \\ 0, & \text{else.} \end{cases} \end{aligned}$$

The Markov chain has a unique invariant probability measure  $\nu$  with respect to which the Markov chain  $(g_n)_{n \geq 0}$  is stationary. By the Kolmogorov existence theorem, there exists a stationary Markov chain indexed by  $\mathbb{Z}$  that we still denote by  $(g_n)_{n \in \mathbb{Z}}$  with transition probabilities given above and which defines a probability measure  $\mathbb{P}_{\mathcal{G}}$  on the space  $\mathcal{G}$ .

Define  $\pi_G: \Delta \rightarrow \mathcal{G}$  by  $\pi_G(x) = (a_n, \ell_n)_n$  such that  $f^n(x) \in ([a_n], \ell_n)$  for all  $n$ , where  $[a_n] \subset \mathcal{A}^{\mathbb{Z}}$  is the cylinder  $\{x \in \mathcal{A}^{\mathbb{Z}} : x_0 = a_n\}$ . Note that  $\pi_G$  is a bijection; moreover, it is a measure preserving conjugacy. It follows that if we define

$$\psi: \mathcal{G} \rightarrow \mathbb{R} \text{ by } \psi = v \circ \pi_G^{-1}, \quad (3.1)$$

then setting for all  $k \in \mathbb{Z}$ ,

$$X_k = \psi((g_{k+n})_{n \in \mathbb{Z}}), \quad (3.2)$$

the stationary process  $(X_k)_{k \in \mathbb{Z}}$  has the same law as  $(v \circ f^k)_{k \in \mathbb{Z}}$ . To give the regularity properties of  $\psi$ , we need to equip  $\mathcal{G}$  with a suitable metric.

### 3.3 Metric on $\mathcal{G}$

Since  $\pi_G: \Delta \rightarrow \mathcal{G}$  is a bijection and  $\Delta$  is a metric space, it is natural to define a metric on  $\mathcal{G}$  so that  $\pi_G$  is an isometry. For  $g, g' \in \mathcal{G}$ , let

$$\tilde{s}^{\pm}(g, g') = \inf\{\ell \geq 0 : g_{\pm\ell} \neq g'_{\pm\ell}\}.$$

Let  $G_0 = \{(a, 0) : a \in \mathcal{A}\} \subset G$ ,

$$s^-(g, g') = \#\{0 \leq \ell < \tilde{s}^-(g, g') : g_{-\ell} \in G_0\} \text{ and } s^+(g, g') = \#\{0 < \ell \leq \tilde{s}^+(g, g') : g_{\ell} \in G_0\}.$$

Let

$$s(g, g') = \min\{s^-(g, g'), s^+(g, g')\}.$$

Then  $s$  is a kind of separation time (cf. the separation time on  $X$  defined in Section 2.1). Recall that  $\xi \in (0, 1)$  and define

$$d(g, g') = \xi^{s(g, g')}.$$

This is the metric on  $\mathcal{G}$  which agrees with that on  $\Delta$  in the sense that  $d(x, y) = d(\pi_G(x), \pi_G(y))$  for all  $x, y \in \Delta$ .

This allows to infer that if  $v$  is Hölder with index  $\eta \in ]0, 1]$  then so is  $\psi$  defined by (3.1).

Next, to prove the ASIP for the partial sums associated with  $(X_k)_{k \in \mathbb{Z}}$  defined by (3.2) with  $\psi$  bounded and Hölder continuous, it is convenient, as in [13, 14], to represent the Markov chain with the help of its innovations and to introduce a particular meeting time.

### 3.4 Innovations and meeting time of the underlying Markov chain

Extending the approach in [13, Section 3], without changing the distribution of  $(g_n)_{n \in \mathbb{Z}}$ , we assume that there is a sequence of iid random elements  $(\varepsilon_n)_{n \in \mathbb{Z}}$  sampled from  $(\mathcal{A}, \mathbb{P}_{\mathcal{A}})$  such that each  $\varepsilon_{n+1}$  is independent from  $(\varepsilon_{\ell}, g_{\ell})_{\ell \leq n}$ , and  $g_{n+1} = U(g_n, \varepsilon_{n+1})$ , where

$$U((a, \ell), \varepsilon) = \begin{cases} (a, \ell + 1), & \ell < h(a) - 1, \\ (\varepsilon, 0), & \ell = h(a) - 1. \end{cases}$$

We refer to  $(\varepsilon_n)_{n \in \mathbb{Z}}$  as *innovations*.

For  $g_0, g'_0 \in G$  and a sequence  $(\varepsilon_n)_{n \in \mathbb{Z}}$  of random elements in  $\mathcal{A}$ , let  $T_{g_0, g'_0}((\varepsilon_n))$  be the meeting time of two Markov chains with respective initial states  $g_0, g'_0$  and common innovations  $(\varepsilon_n)_{n \in \mathbb{Z}}$ :

$$T_{g_0, g'_0}((\varepsilon_n)) = \inf\{\ell \geq 0 : g_\ell = g'_\ell\},$$

where  $g_{n+1} = U(g_n, \varepsilon_{n+1})$  and  $g'_{n+1} = U(g'_n, \varepsilon_{n+1})$ .

**Proposition 3.5.** *Suppose that  $g'_0$  is an independent copy of  $g_0$ , also independent from  $(g_n, \varepsilon_n)_{n \geq 0}$ . Set  $T = T_{g_0, g'_0}((\varepsilon_n))$ .*

1. *Assume that there exists a sequence  $\varepsilon(n)$  of positive reals tending to 0 as  $n \rightarrow \infty$ , such that  $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} \geq n) \geq \exp(-n\varepsilon(n))$ . If in addition  $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} \geq n) = O(L(n)n^{-\beta})$  with  $\beta > 1$  and  $L(n)$  a slowly varying function at infinity, then  $\mathbb{P}(T \geq n) = O(L(n)n^{-(\beta-1)})$ .*
2. *If  $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} \geq n) = O(L(n)n^{-\beta})$  with  $\beta > 1$  and  $L(n)$  a slowly varying function at infinity, then  $\mathbb{E}(g_\beta(T)) < \infty$  for any  $\eta > 1$  where  $g_\beta(x) = x^{\beta-1}(\log(1+x))^{-\eta}/L(x)$ .*
3. *If  $\mathbb{E}_{\mathcal{A}}(h_{\mathcal{A}}^\beta) < \infty$  for some  $\beta > 1$ , then  $\mathbb{E}(T^{\beta-1}) < \infty$ .*
4. *If  $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} \geq n) = O(e^{-cn^\delta}) < \infty$  for some  $c > 0$  and  $\delta \in (0, 1]$ , then there exists  $\kappa > 0$  such that  $\mathbb{P}(T \geq n) = O(e^{-\kappa n^\delta})$ .*

Items 2 and 3 are proved in [13] whereas Item 4 is proved in [14]. Item 1 of Proposition 3.5 follows straightforwardly from Lemma 3.6 below.

**Lemma 3.6.** *Suppose that  $g'_0$  is an independent copy of  $g_0$ , also independent from  $(g_n, \varepsilon_n)_{n \geq 0}$ . Set  $T = T_{g_0, g'_0}((\varepsilon_n))$ . Suppose that there exists a sequence  $\varepsilon(n)$  of positive reals tending to 0 as  $n \rightarrow \infty$ , such that  $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} \geq n) \geq \exp(-n\varepsilon(n))$ . Then*

$$\mathbb{P}(T \geq n) = O(\mathbb{E}_{\mathcal{A}}((h_{\mathcal{A}} - n)_+)).$$

*Proof.* As in Lindvall [23] (see also Rio [29, Proposition 9.6]), let  $\Lambda_0$  be the class of nondecreasing functions  $\psi$  from  $\mathbb{N}$  to  $[1, \infty[$  such that  $\log(\psi(n))/n$  decreases to 0 as  $n \rightarrow \infty$ . Let

$$\psi^{(0)}(k) = \sum_{i=0}^{k-1} \psi(i) \quad \text{and, for } k \geq 1, \quad \varphi(k) = \psi(k) - \psi(k-1).$$

Proceeding exactly as in the proof of [13, Lemma 3.1] and applying the result of [23], we infer that

$$\mathbb{E}_{\mathcal{A}}(\psi^{(0)}(h_{\mathcal{A}})) < \infty \Rightarrow \mathbb{E}(\psi(T)) < \infty.$$

This last assertion is in fact equivalent to

$$\sum_{k=1}^{\infty} \varphi(k) \mathbb{E}_{\mathcal{A}}((h_{\mathcal{A}} - k)_+) < \infty \Rightarrow \sum_{k=1}^{\infty} \varphi(k) \mathbb{P}(T \geq k) < \infty. \quad (3.3)$$

Now, assume that the conclusion of Lemma 3.6 is not true. Then there exists an increasing sequence  $(n_k)_{k \geq 1}$  such that  $\mathbb{P}(T \geq n_k) \geq k \mathbb{E}_{\mathcal{A}}((h_{\mathcal{A}} - n_k)_+)$ . Define then the function  $\varphi$  as follows:  $\varphi(i) = 0$  if  $i \notin \{n_k, k \geq 1\}$  and  $\varphi(n_k) = k^{-3/2} / \mathbb{E}_{\mathcal{A}}((h_{\mathcal{A}} - n_k)_+)$ . For such a  $\varphi$  it is clear that the sum on left hand in (3.3) is finite, while the selection of  $n_k$  implies that the sum on right hand is  $+\infty$ . This leads to a contradiction and proves the Lemma, provided we check that the function  $\psi$  defined by  $\psi(k) = \psi(0) + \varphi(1) + \dots + \varphi(k)$  belongs to  $\Lambda_0$ . Hence, it remains to check that  $\log(\psi(n_k))/n_k$  tends to 0 as  $k \rightarrow \infty$ . Now, by definition of  $\varphi$ ,  $\psi(n_k) - \psi(0) \leq C\sqrt{k} / \mathbb{E}_{\mathcal{A}}((h_{\mathcal{A}} - n_k)_+)$ . On another hand, the assumption on  $h_{\mathcal{A}}$  implies that  $\mathbb{E}_{\mathcal{A}}((h_{\mathcal{A}} - n_k)_+) \geq \exp(-n_k \varepsilon'(n_k))$  for some sequence  $\varepsilon'(n)$  of positive reals tending to 0 as  $n \rightarrow \infty$ . This implies that  $\log(\psi(n_k))/n_k$  tends to 0 as  $k \rightarrow \infty$ , and the proof is complete.  $\square$

### 3.5 Proof of Proposition 3.4 when $c^2 > 0$

By the previous considerations, it suffices to prove the ASIP for the partial sums associated with  $(X_k)_{k \in \mathbb{Z}}$  defined by (3.2) with  $\psi$  bounded and Hölder continuous, by taking into account Proposition 3.5.

Without loss of generality, one can assume that  $v: \Delta \rightarrow \mathbb{R}$  is Lipschitz and that  $\|v\|_{\text{Lip}} \leq 1$ , where  $\|v\|_{\text{Lip}} = \sup_x |v(x)| + \sup_{x \neq y} |v(x) - v(y)|/d(x, y)$ . Recall that  $\psi: \mathcal{G} \rightarrow \mathbb{R}$ ,  $\psi = v \circ \pi_G^{-1}$ . Then  $\psi$  is also Lipschitz, in particular

$$|\psi(g) - \psi(g')| \leq \xi^{s(g, g')}.$$

We assume also in the rest of this section that the Markov chain  $(g_n)_{n \in \mathbb{Z}}$  is aperiodic. If it is not the case, we modify the proof as in [13, Appendix A].

The proof follows the lines of [13, Section 4.2] or [14, Section 3] with the same notations, but with the following changes. [13, Proposition 3.2] and [14, Proposition 2.3] are replaced by Proposition 3.7 below whereas [13, Inequality 4.10] is replaced by our Proposition 3.8, whereas [13, Lemma 3.3] is replaced by our Lemma 3.9. Moreover in the case where  $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} \geq n) = O(n^{-\beta}(\log n)^\gamma)$ , with  $\beta > 2$  and  $\gamma \in \mathbb{R}$ , we take  $m = m_\ell = \lceil 3^{\ell/\beta} \ell^\kappa \rceil$  with  $\kappa > (1 + \gamma)/\beta$ . In case  $\int h_{\mathcal{A}}^\beta d\mathbb{P}_{\mathcal{A}} < \infty$  with  $\beta > 2$ , we select  $m = m_\ell = \lceil 3^{\ell/\beta} \rceil$ , and if  $\int e^{\beta h_{\mathcal{A}}} d\mathbb{P}_{\mathcal{A}} < \infty$  with  $\beta > 0$  and  $\delta \in (0, 1]$ , we take  $m = m_\ell = \lceil \kappa \ell^{1/\delta} \rceil$ , for a suitable  $\kappa$ .

To state the key propositions 3.7 and 3.8, we need to introduce some notations.

Fix  $k \in \mathbb{Z}$  and  $m > 0$ , and let  $(\varepsilon'_n)_n$  be a copy of  $(\varepsilon_n)_n$  independent of  $(g_n, \varepsilon_n)_n$ . Define

$$\tilde{g}_\ell = \begin{cases} g_\ell, & \ell \leq k + m, \\ U(g_{\ell-1}, \varepsilon'_\ell), & \ell = k + m + 1, \\ U(\tilde{g}_{\ell-1}, \varepsilon'_\ell), & \ell \geq k + m + 2 \end{cases} \quad \text{and} \quad X_{m,k} = \mathbb{E}_g \psi((\tilde{g}_{n+k})_{n \in \mathbb{Z}}),$$

where  $\mathbb{E}_g$  is the conditional expectation given  $(g_n, \varepsilon_n)$ .

Let

$$T = T_{g_0, g'_0}((\varepsilon_n)_{n \geq 1}) \tag{3.4}$$

be the meeting time between  $(g_n)_{n \geq 0}$  and  $(\tilde{g}_n)_{n \geq 0}$  with the same innovation  $(\varepsilon_n)_{n \geq 1}$  but starting from two independent starting points  $g_0$  and  $g'_0$ .

#### Proposition 3.7.

1. Assume that  $\mathbb{E}(T) < \infty$ . Then, for every  $r \geq 1$ ,

$$\mathbb{E}|X_{m,k} - X_k| \ll m^{-r/2} + \mathbb{P}(T \geq \lfloor m/r \rfloor).$$

2. Assume that there exist  $\delta > 0$  and  $\gamma \in ]0, 1]$  such that  $\mathbb{P}(T \geq n) = O(e^{-\delta n^\gamma})$ . Then there exists  $\delta' > 0$  such that

$$\mathbb{E}|X_{m,k} - X_k| = O((e^{-\delta' m^\gamma}).$$

*Proof.* See [13, Proposition 3.2] and [14, Proposition 2.3]. □

Now define

$$\tilde{X}_{m,k} = \mathbb{E}(X_{m,k} \mid \varepsilon_{k-m}, \dots, \varepsilon_{k+m}). \tag{3.5}$$

**Proposition 3.8.** Let  $\theta_u^- = \sum_{k=0}^{u-1} \mathbf{1}_{\{g_{-k} \in G_0\}}$  and  $\theta_u^+ = \sum_{k=0}^{u-1} \mathbf{1}_{\{g_k \in G_0\}}$ . Then

$$\mathbb{E}|\tilde{X}_{m,k} - X_{m,k}| \ll \|\psi\|_\infty \mathbb{P}(T > m/2) + \mathbb{E} \xi^{\min(\theta_{\lfloor m/2 \rfloor}^-, \theta_m^+)},$$

where  $T$  is defined by (3.4). Consequently:

1. If  $\mathbb{E}(T) < \infty$ , then, for every  $r \geq 1$ ,

$$\mathbb{E}|\tilde{X}_{m,k} - X_k| \ll m^{-r/2} + \mathbb{P}(T \geq \lfloor m/r \rfloor).$$

2. If there exist  $\delta > 0$  and  $\gamma \in ]0, 1]$  such that  $\mathbb{P}(T \geq n) = O(e^{-\delta n^\gamma})$ , then there exists  $\delta' > 0$  such that

$$\mathbb{E}|\tilde{X}_{m,k} - X_k| = O(e^{-\delta' m^\gamma}).$$

*Proof.* Since  $X_{m,k}$  is determined by  $(g_n)_{n \leq k+m}$ , we write  $X_{m,k} = h_m((g_n)_{n \leq k+m})$  with some function  $h_m$ . Let  $(g'_n, \varepsilon'_n)_{n \in \mathbb{Z}}$  be an independent copy of  $(g_n, \varepsilon_n)_{n \in \mathbb{Z}}$ . Note that  $g'_n = U(g'_{n-1}, \varepsilon'_n)$ . Define  $(\hat{g}_\ell)$  by

$$\hat{g}'_\ell = \begin{cases} g'_\ell & \ell < k-m, \\ U(g'_{k-m-1}, \varepsilon_{k-m}), & \ell = k-m, \\ U(\hat{g}'_{\ell-1}, \varepsilon_\ell), & \ell \leq k-m+1. \end{cases}$$

Let  $\mathcal{B}_{k,m}$  be the sigma-algebra generated by  $((g_\ell)_{\ell < k-m}, \varepsilon_{k-m}, \dots, \varepsilon_{k+m})$ . By the properties of conditional expectation, we have

$$X_{m,k} = \mathbb{E}(h_m((g_n)_{n \leq k+m}) \mid \mathcal{B}_{k,m})$$

and

$$\tilde{X}_{m,k} = \mathbb{E}(h_m((g'_u)_{u < k-m}, (\hat{g}'_u)_{k-m \leq u \leq k+m}) \mid \mathcal{B}_{k,m}).$$

Set

$$\hat{T} = T_{g_{k-m-1}, \hat{g}'_{k-m-1}}((\varepsilon_\ell)_{\ell \geq k-m})$$

This is the meeting time of the chains  $(g_n)_{n \geq k-m-1}$  and  $(\hat{g}'_n)_{n \geq k-m-1}$  with innovations  $(\varepsilon_\ell)_{\ell \leq k-m}$  and independent starting points  $g_{k-m-1}$  and  $\hat{g}'_{k-m-1}$ . It has the same law as  $T$  defined in (3.4). Note that

$$\begin{aligned} I &= \left| h_m((g_n)_{n \leq k+m}) - h_m((g'_u)_{u \leq k-m-1}, (\hat{g}'_u)_{k-m \leq u \leq k+m}) \right| \\ &\leq 2 \|h_m\|_\infty \mathbf{1}_{\hat{T} > m/2} + \sum_{\ell=0}^{\lfloor m/2 \rfloor} \mathbf{1}_{\hat{T}=\ell} \left| h_m((g_n)_{n \leq k+m}) \right. \\ &\quad \left. - h_m((g'_u)_{u \leq k-m-1}, (\hat{g}'_u)_{k-m \leq u < k-m+\ell}, g_{k-m+\ell}, \dots, g_k, \dots, g_{k+m}) \right|. \end{aligned}$$

But, from the definition of the separation distance and using  $\|v\|_{\text{Lip}} \leq 1$ , for any  $\ell$  in  $\{0, \dots, \lfloor m/2 \rfloor\}$ ,

$$\begin{aligned} &\left| h_m((g_n)_{n \leq k+m}) - h_m((g'_u)_{u \leq k-m-1}, (\hat{g}'_u)_{k-m \leq u < k-m+\ell}, g_{k-m+\ell}, \dots, g_k, \dots, g_{k+m}) \right| \\ &\leq \max(\xi^{\#\{k-m+\ell \leq i \leq k: g_i \in G_0\}}, \xi^{\#\{k \leq i \leq k+m: g_i \in G_0\}}). \end{aligned}$$

Taking the expectation, we get

$$\begin{aligned} \mathbb{E}|X_{m,k} - \tilde{X}_{m,k}| &\leq \mathbb{E}(I) \\ &\leq 2 \mathbb{E}(\|h_m\|_\infty \mathbf{1}_{\hat{T} > m/2}) + \mathbb{E} \max(\xi^{\#\{k-[m/2] \leq i \leq k: g_i \in G_0\}}, \xi^{\#\{k \leq i \leq k+m: g_i \in G_0\}}). \end{aligned}$$

The first part of the proposition follows by stationarity and the fact that  $\|h_m\|_\infty \leq \|\psi\|_\infty$ . To end the proof, we used the same arguments as in the proofs of [13, Proposition 3.2] and [14, Proposition 2.3].  $\square$



**Lemma 3.9.**

1. If  $\mathbb{E}(T) < \infty$ , for every  $\alpha \geq 1$  and every  $k \geq 1$ ,

$$|\text{Cov}(X_0, X_k)| \ll k^{-\alpha/2} + \mathbb{P}(T \geq \lfloor k/(4\alpha) \rfloor).$$

2. If there exist  $\delta > 0$  and  $\gamma \in ]0, 1]$  such that  $\mathbb{P}(T \geq n) = O(e^{-\delta n^\gamma})$ , then for any  $k \geq 0$ , there exists  $\delta' > 0$  such that

$$|\text{Cov}(X_0, X_k)| = O((e^{-\delta' k^\gamma})).$$

*Proof.* For every positive integers  $m$  and  $n$ , we have

$$\begin{aligned} |\text{Cov}(X_0, X_k)| &\leq \|X_k\|_\infty \|X_0 - \tilde{X}_{m,0}\|_1 + |\text{Cov}(\tilde{X}_{m,0}, X_k)| \\ &\leq \|X_k\|_\infty \|X_0 - \tilde{X}_{m,0}\|_1 + \|\tilde{X}_{m,0}\|_\infty \|X_k - \tilde{X}_{n,k}\|_1 + |\text{Cov}(\tilde{X}_{m,0}, \tilde{X}_{n,k})|. \end{aligned}$$

For every  $k \geq 2$ , we select  $m = \lfloor k/2 \rfloor$  and  $n = \lfloor k/2 \rfloor - 1$ . In this case  $\tilde{X}_{m,0}$  and  $\tilde{X}_{n,k}$  are independent, implying that for any  $k \geq 2$ ,

$$|\text{Cov}(X_0, X_k)| \leq \|\psi\|_\infty (\|X_0 - \tilde{X}_{\lfloor k/2 \rfloor, 0}\|_1 + \|X_k - \tilde{X}_{\lfloor k/2 \rfloor - 1, k}\|_1).$$

This upper bound combined with Proposition 3.8 proves the lemma.  $\square$

### 3.6 Proof of Proposition 3.4 when $c^2 = 0$

Like in the nonuniformly expanding case [13], our strategy is to represent  $X_k$  as a coboundary:  $X_k = z_{k-1} - z_k$ , where  $z_i$  is a sufficiently nice stationary sequence, and to bound the growth of  $z_k$  almost surely. Our proof follows well established ideas [7, 15, 26, 32] but unlike in the nonuniformly expanding case, we have not found a result that we can conveniently cite.

The representation  $X_k = z_{k-1} - z_k$  is provided by the following general result which is essentially due to Gordin [18] (see also [15, Lemma 7.1] in case of adapted r.v.'s but there the condition c) is actually unnecessary).

**Lemma 3.10.** *Suppose that  $(X_k)_{k \in \mathbb{Z}}$  is a strictly stationary sequence of real-valued random variables that are in  $L^2$ , and let  $(\mathcal{F}_k)_{k \in \mathbb{Z}}$  be a stationary filtration. Let  $\mathbb{E}_i$  denote the conditional expectation given  $\mathcal{F}_i$  and set  $S_n = X_1 + \dots + X_n$ . Assume that*

(a)  $\mathbb{E}_0(S_n)$  and  $\sum_{k=0}^n (X_{-k} - \mathbb{E}_0(X_{-k}))$  converge in  $L^1$  as  $n \rightarrow \infty$ ,

(b)  $\liminf_{n \rightarrow \infty} \frac{\mathbb{E}|S_n|}{\sqrt{n}} = 0$ .

*Then, almost surely,  $X_i = z_{i-1} - z_i$ , where  $z_i = g_i - h_i$  with  $g_i = \sum_{k \geq i+1} \mathbb{E}_i(X_k)$  and  $h_i = \sum_{k \leq i} (X_k - \mathbb{E}_i(X_k))$ . All  $z_i, g_i, h_i$  are in  $L^1$ .*

*Proof.* Set  $d_i = \sum_{k \in \mathbb{Z}} P_i(X_{k+i})$  where  $P_i = \mathbb{E}_i - \mathbb{E}_{i-1}$ . Then

$$X_i = d_i + z_{i-1} - z_i. \tag{3.6}$$

By assumption, all random variables in the above decomposition are in  $L^1$ . Moreover,  $(d_i)_{i \in \mathbb{Z}}$  is a stationary martingale difference sequence. Let's prove that the  $d_i$ 's are almost surely equal to zero. With this aim, we shall proceed as in [16]. By Burkholder's deviation inequality for the

martingale square function [9] (see [16, Lemma 6] for an easy reference), there exists a positive constant  $C$  such that, for any  $\lambda > 0$ ,

$$\mathbb{P}\left(n^{-1} \sum_{i=1}^n d_i^2 > \lambda^2\right) \leq C\lambda^{-1} \frac{\mathbb{E}\left|\sum_{i=1}^n d_i\right|}{\sqrt{n}}.$$

But, by (3.6),  $\mathbb{E}\left|\sum_{i=1}^n d_i\right| \leq \mathbb{E}|S_n| + 2\mathbb{E}|z_0|$ . Using part (b), there exists a subsequence  $(n_k)_{k>0}$  tending to infinity such that  $n_k^{-1} \sum_{i=1}^{n_k} d_i^2$  converges to zero in probability as  $k \rightarrow \infty$ . On the other hand, by the ergodic theorem, since  $(d_i^2)_{i \in \mathbb{Z}}$  is a stationary sequence of nonnegative r.v.'s,  $n^{-1} \sum_{i=1}^n d_i^2$  converges almost surely to  $\mathbb{E}(d_0^2|\mathcal{I})$  where  $\mathcal{I}$  is the so-called invariant  $\sigma$ -field. So, overall,  $\mathbb{E}(d_0^2|\mathcal{I}) = 0$  almost surely and  $\mathbb{E}(d_0^2) = \mathbb{E}(\mathbb{E}(d_0^2|\mathcal{I})) = 0$ . This proves that  $d_0 = 0$  almost surely.  $\square$

Now we return to the specific process we are interested in:  $X_k = \psi((g_{k+\ell})_{\ell \in \mathbb{Z}})$  as in (3.2). Let  $\mathcal{F}_k = \sigma(\varepsilon_i, i \leq k)$ .

**Proposition 3.11.** *Suppose that  $\mathbb{E}(T) < \infty$  and  $c^2 = 0$ . Then the assumptions of Lemma 3.10 are satisfied.*

*Proof.* Recall the construction of  $X_{m,k}$  and  $\tilde{X}_{m,k} = \mathbb{E}(X_{m,k}|\varepsilon_{k-m}, \dots, \varepsilon_{k+m})$ . For  $k \geq 0$ , let  $X_k^* = \tilde{X}_{m_k,k}$  with  $m_k = \lfloor k/2 \rfloor$ . Clearly  $X_k^*$  is independent of  $\mathcal{F}_0$  and is centered. To prove the first part of assumption (a), that is the convergence of  $\mathbb{E}_0(S_n)$ , write

$$\sum_{k \geq 2} \|\mathbb{E}_0(X_k)\|_1 \leq \sum_{k \geq 2} \|X_k - X_k^*\|_1,$$

and use Proposition 3.8 to prove that the right hand side above converges.

Next we prove the second part of assumption (a). For  $k \geq 0$  let  $X_{-k}^* = \tilde{X}_{\lfloor k/2 \rfloor, -k}$ . Clearly  $X_{-k}^*$  is  $\mathcal{F}_0$ -measurable. Hence

$$\sum_{k \geq 0} \|X_{-k} - \mathbb{E}_0(X_{-k})\|_1 = \sum_{k \geq 0} \|X_{-k} - X_{-k}^* - \mathbb{E}_0(X_{-k} - X_{-k}^*)\|_1 \leq 2 \sum_{k \geq 0} \|X_{-k} - X_{-k}^*\|_1.$$

By another application of Proposition 3.8, for any  $r \geq 1$ ,

$$\sum_{k \geq 2} \|X_{-k} - \mathbb{E}_0(X_{-k})\|_1 \ll \sum_{k \geq 2} (k^{-r/2} + \mathbb{P}(T \geq \lfloor k/(2r) \rfloor)).$$

Choosing  $r > 1$ , the second part of assumption (a) follows.

We turn now to the proof of assumption (b). By Lemma 3.9, since  $\mathbb{E}(T) < \infty$ ,  $\sum_{k \geq 0} |\text{Cov}(X_0, X_k)| < \infty$  and then  $c^2 = \lim_{n \rightarrow \infty} n^{-1} \mathbb{E}(S_n^2)$  is well defined and is assumed to be zero. It follows that assumption (b) is satisfied.  $\square$

At this point, as long as  $c^2 = 0$ , we have the representation

$$S_n = X_1 + \dots + X_n = z_0 - z_n$$

with  $z_n$  as in Lemma 3.10. Thus the proof of Proposition 3.4 is reduced to estimating  $z_n$  almost surely, and is completed by:

**Proposition 3.12.** *Suppose  $\mathbb{E}(T) < \infty$  and  $c^2 = 0$  and  $z_i$  are as in Lemma 3.10.*

1. *If  $\mathbb{E}(T^p) < \infty$  for some  $p > 1$ , then  $\|z_0\|_p < \infty$  and  $S_n = o(n^{1/(p+1)})$  a.s.*

2. If  $\mathbb{E}(\psi_p(T)) < \infty$  with  $\psi_p(x) = x^p(\log x)^{-(\gamma+1+\varepsilon)}$  for some  $\varepsilon > 0$ , then  $S_n = o(n^{1/(p+1)}(\log n)^{(\gamma+1+\varepsilon)/(p+1)})$  a.s.

3. If  $\mathbb{P}(T \geq n) = O(e^{-\delta n^\gamma})$  with some  $\delta > 0$  and  $\gamma \in ]0, 1]$ , then  $S_n = O((\log n)^{1/\gamma})$  a.s.

*Proof.* Denote  $M = \|X_0\|_\infty = \|\psi\|_\infty$ .

We begin with part 1: let  $p > 1$  and suppose that  $\mathbb{E}(T^p) < \infty$ . By Proposition 3.8,

$$\sum_{k>0} k^{p-1} (\|X_k - \tilde{X}_{[k/2],k}\|_1 + \|X_{-k} - \tilde{X}_{[k/2],-k}\|_1) < \infty, \quad (3.7)$$

First we use (3.7) to prove that  $z_0$  is in  $L^p$ . Recall that  $z_0 = g_0 - h_0$  where  $g_0 = \sum_{k \geq 1} \mathbb{E}_0(X_k)$  and  $h_0 = \sum_{k \leq 0} (X_k - \mathbb{E}_0(X_k))$ . Observe that for  $k \geq 2$ , these summands can be bounded in  $L^1$  by those in (3.7):  $\|\mathbb{E}_0(X_k)\|_1 \leq \|X_k - \tilde{X}_{[k/2],k}\|_1$  and  $\|X_{-k} - \mathbb{E}_0(X_{-k})\|_1 \leq 2\|X_{-k} - \tilde{X}_{[k/2],-k}\|_1$ . Now,  $g_0$ ,  $h_0$  and consequently  $z_0$  are in  $L^p$  by an application of Lemma 3.13 with  $V = g_0$  and  $V = h_0$  respectively. (To prove that  $g_0$  is in  $L^p$  we can also use the arguments given in the proof of [15, Proposition 2.1].)

To prove that  $S_n = o(n^{1/(p+1)})$  a.s., it suffices to show that, for every  $\varepsilon > 0$ ,  $\sum_{\ell > 0} \mathbb{P}(\max_{k \leq 2^\ell} |S_k| > \varepsilon 2^{\ell/(p+1)}) < \infty$ , which is equivalent to proving that

$$\sum_{n>1} n^{-1} \mathbb{P}(\max_{k \leq n} |S_k| > \varepsilon n^{1/(p+1)}) < \infty. \quad (3.8)$$

Let  $k \leq n$  and write

$$S_k = S_k - \mathbb{E}_k(S_k) + \mathbb{E}_k(S_k - S_n) + \mathbb{E}_k(S_n).$$

Let  $q \geq 1$ . Using  $\mathbb{E}_k(X_i - \mathbb{E}_{i-q}(X_i)) = 0$  if  $i - q \geq k$ , we have

$$\begin{aligned} |\mathbb{E}_k(S_n - S_k)| &\leq \left| \sum_{i=k+1}^n \mathbb{E}_k(X_i - \mathbb{E}_{i-q}(X_i)) \right| + \left| \sum_{i=k+1}^n \mathbb{E}_k(\mathbb{E}_{i-q}(X_i)) \right| \\ &\leq 2qM + \sum_{i=k+1}^n |\mathbb{E}_k(\mathbb{E}_{i-q}(X_i))|. \end{aligned}$$

Then, using  $\mathbb{E}_k(\mathbb{E}_{\ell+q}(X_\ell)) = \mathbb{E}_{\ell+q}(X_\ell)$  for  $\ell \leq k - q$ ,

$$\begin{aligned} |S_k - \mathbb{E}_k(S_k)| &\leq \left| \sum_{\ell=1}^{k-q} (X_\ell - \mathbb{E}_{\ell+q} X_\ell - \mathbb{E}_k(X_\ell - \mathbb{E}_{\ell+q} X_\ell)) \right| + 2qM \\ &\leq \sum_{\ell=1}^{k-q} |X_\ell - \mathbb{E}_{\ell+q}(X_\ell)| + \sum_{\ell=1}^{k-q} \mathbb{E}_k |X_\ell - \mathbb{E}_{\ell+q}(X_\ell)| + 2qM. \end{aligned}$$

Overall,

$$\begin{aligned} \max_{1 \leq k \leq n} |S_k| &\leq 4qM + \max_{1 \leq k \leq n} |\mathbb{E}_k(S_n)| + \max_{1 \leq k \leq n} \sum_{i=1}^n |\mathbb{E}_k(\mathbb{E}_{i-q}(X_i))| \\ &\quad + \sum_{\ell=1}^n |X_\ell - \mathbb{E}_{\ell+q}(X_\ell)| + \max_{1 \leq k \leq n} \sum_{\ell=1}^n \mathbb{E}_k |X_\ell - \mathbb{E}_{\ell+q}(X_\ell)|. \end{aligned}$$

Accordingly, if  $x > 0$  and  $qM \leq x$ , then

$$\begin{aligned} \mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| > 8x\right) &\leq \mathbb{P}\left(\max_{1 \leq k \leq n} |\mathbb{E}_k(S_n)| > x\right) + \mathbb{P}\left(\max_{1 \leq k \leq n} \sum_{i=1}^n |\mathbb{E}_k(\mathbb{E}_{i-q}(X_i))| > x\right) \\ &+ \mathbb{P}\left(\sum_{\ell=1}^n |X_\ell - \mathbb{E}_{\ell+q}(X_\ell)| > x\right) + \mathbb{P}\left(\max_{1 \leq k \leq n} \sum_{\ell=1}^n \mathbb{E}_k |X_\ell - \mathbb{E}_{\ell+q}(X_\ell)| > x\right). \end{aligned} \quad (3.9)$$

Let

$$\theta(q) = \|X_0 - \tilde{X}_{q,0}\|_1.$$

Starting from (3.9) and using Doob's maximal inequality, as done in the proof of inequality (A.42) in [27], we infer that for any nondecreasing, non negative and convex function  $\varphi$  and any  $x > 0$ ,

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| > 8x\right) \ll \frac{\mathbb{E}(\varphi(S_n))}{\varphi(x)} + nx^{-1}\theta(\lfloor x/M \rfloor). \quad (3.10)$$

Then, taking  $\varphi(x) = x^p$ ,

$$\begin{aligned} \sum_{n>1} n^{-1} \mathbb{P}\left(\max_{k \leq n} |S_k| > \varepsilon n^{1/(p+1)}\right) &\ll \sum_{n>1} n^{-1-\frac{p}{p+1}} \|S_n\|_p^p + \sum_{n>1} n^{-1/(p+1)} \theta(\lfloor \varepsilon n^{1/(p+1)} / M \rfloor) \\ &\ll \sup_n \|S_n\|_p^p + \sum_{n>1} n^{p-1} \theta(n). \end{aligned}$$

The expression above is finite:  $\|S_n\|_p \leq 2\|z_0\|_p < \infty$  and  $\sum_{n>1} n^{p-1} \theta(n) < \infty$  by Proposition 3.8. This verifies (3.8) and completes the proof of part 1.

We now turn to part 2. By part 1,  $\|S_n\|_r < \infty$  for all  $r \in [1, p[$ . Hence, with  $b_n = n^{1/(p+1)}(\log n)^{(\varepsilon+\gamma+1)/(p+1)}$ , applying (3.10) with  $\varphi(x) = x^r$ , it follows that for any  $\eta > 0$ ,

$$\mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| > \eta b_n\right) \ll b_n^{-r} + \min\left(1, \frac{n}{\eta b_n} \mathbb{P}(T > c\eta b_n)\right),$$

where  $c$  is a positive constant which does not depend on  $n$ . Part 2 is proved if we show that  $\sum_{n \geq 1} b_n^{-1} \mathbb{P}(T > c\eta b_n) < \infty$ . By a change of variable this is equivalent to  $\sum_{n \geq 2} n^{p-1} (\log n)^{\varepsilon+\gamma+1} \mathbb{P}(T > n) < \infty$  which holds by Proposition 3.8 because  $\mathbb{E}(\psi_p(T)) < \infty$ .

We turn now to the proof of part 3. Let  $K > 0$ , to be chosen later. By (3.10) applied with  $x = K(\log n)^{1/\gamma}$  and  $\varphi(u) = u^2$ , using Lemma (3.13), our assumption and part 2 of Proposition 3.8, we see that

$$\begin{aligned} \frac{1}{n} \mathbb{P}\left(\max_{1 \leq k \leq n} |S_k| > 8K(\log n)^{1/\gamma}\right) &\ll \frac{1}{n(\log n)^{2/\gamma}} + \frac{1}{K(\log n)^{1/\gamma}} e^{-\delta'(K(\log n)^{1/\gamma}/M)^\gamma} \\ &\ll \frac{1}{n(\log n)^2} + \frac{1}{n^{\delta'(K/M)^\gamma}}. \end{aligned}$$

We conclude by taking  $K = 2M\delta'^{-1/\gamma}$ . □

**Lemma 3.13.** *Let  $(V_k)_{k \in \mathbb{N}}$  be a sequence of non negative random variables uniformly bounded by  $M$ . Set  $V = \sum_{k \in \mathbb{N}} V_k$ . Then, for every  $p > 1$ , we have*

$$\|V\|_p \leq M + \sum_{k \in \mathbb{N}} k^{p-1} \|V_k\|_1. \quad (3.11)$$

*Proof.* Recall that  $\|V\|_p = \sup_{\|Z\|_q=1} \mathbb{E}(ZV)$ , where  $q = p/(p-1)$ . Let  $Z$  be a random variable with  $\|Z\|_q = 1$ . Then

$$\begin{aligned} \mathbb{E}(|ZV|) &\leq \sum_{k \in \mathbb{N}} k^{p-1} \mathbb{E}(|V_k|) + \sum_{k \in \mathbb{N}} M \mathbb{E}(|Z| \mathbf{1}_{\{|Z| \geq k^{p-1}\}}) \\ &= \sum_{k \in \mathbb{N}} k^{p-1} \mathbb{E}(|V_k|) + M \mathbb{E}\left(|Z| \sum_{1 \leq k \leq |Z|^{1/(p-1)}} 1\right) \\ &= \sum_{k \in \mathbb{N}} k^{p-1} \mathbb{E}(|V_k|) + M \mathbb{E}(|Z|^q), \end{aligned}$$

and the lemma is proved. □

## Acknowledgements

The authors are grateful to E. Rio for useful discussions concerning the proof of Lemma 3.6. The authors are also grateful to the three referees for their careful reading and comments, which improved the presentation of the paper.

## Statements and declarations

**Funding.** A.K. has been partially supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No 787304) and by EPSRC grant EP/V053493/1.

**Conflict of interest.** The authors declare that they have no financial interests.

**Data availability.** There is no associated data.

## References

- [1] V. Araujo, M. J. Pacifico, E. R. Pujals and M. Viana, *Singular-hyperbolic attractors are chaotic*, Trans. Amer. Math. Soc. **361** (2009), 2431–2485.
- [2] P. Bálint, G. Borbély, A. N. Varga, *Statistical properties of the system of two falling balls*, Chaos **22** (2012), 026104.
- [3] P. Bálint, H.K. Zhang, private communication.
- [4] I. Berkes, W. Liu, W. Wu, *Komlós-Major-Tusnády approximation under dependence*, Ann. Probab. **42** (2014), 794–817.
- [5] I. Berkes, W. Philipp, *Approximation theorems for independent and weakly dependent random vectors*, Ann. Probab. **7** (1979), 29–54.
- [6] N. H. Bingham, C. M. Goldie, J. L. Teugels, *Regular variation*, Cambridge University Press (1987).
- [7] R. Bowen, *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms*, Lecture Notes in Math. **470**, Springer, Berlin, 1975.
- [8] L. A. Bunimovich, *The ergodic properties of billiards that are nearly scattering*, Soviet Math. Dokl. **14** (1973), 1136–1139.
- [9] D.L. Burkholder, *A sharp inequality for martingale transforms*, Ann. Probab. **7** (1979), no. 5, 858–863.

- [10] N. Chernov, H.-K. Zhang, *A family of chaotic billiards with variable mixing rates*, Stoch. Dyn. **5** (2005), 535–553.
- [11] N. Chernov, H.-K. Zhang, *Billiards with polynomial mixing rates*, Nonlinearity **18** (2005), 1527–1553.
- [12] N. Chernov, H.-K. Zhang, *Improved estimates for correlations in billiards*, Comm. Math. Phys. **277** (2008), 305–321.
- [13] C. Cuny, J. Dedecker, A. Korepanov, F. Merlevède, *Rates in almost sure invariance principle for slowly mixing dynamical systems*, Ergodic Theory Dynam. Systems **40** (2020), 2317–2348.
- [14] C. Cuny, J. Dedecker, A. Korepanov, F. Merlevède, *Rates in almost sure invariance principle for quickly mixing dynamical systems*, Stoch. Dyn. **20** (2020), no. 1, 2050002.
- [15] J. Dedecker, F. Merlevède, E. Rio, *Deviation inequalities for dependent sequences with applications to strong approximations*, Stochastic Process. Appl. **174** (2024), Paper No. 104377.
- [16] C.G. Esseen, S. Janson, *On moment conditions for normed sums of independent variables and martingale differences*, Stochastic Process. Appl. **19** (1985), no. 1, 173–182.
- [17] N. Fleming-Vázquez, *Functional Correlation Bounds and Optimal Iterated Moment Bounds for Slowly-Mixing Nonuniformly Hyperbolic Maps*, Comm. Math. Phys. **391** (2022), 173–198.
- [18] M.I. Gordin, *Central limit theorems for stationary processes without the assumption of finite variance*, Abstracts of Communications, T.I:A-K. Pages 173–174, International Conference on Probability Theory and Mathematical Statistics, June 25–30, 1973, Vilnius.
- [19] S. Gouëzel, *Almost sure invariance principle for dynamical systems by spectral methods*, Ann. Probab. **38** (2010), 1639–1671.
- [20] A. Korepanov, *Equidistribution for nonuniformly expanding dynamical systems, and application to the almost sure invariance principle*, Comm. Math. Phys. **359** (2018), 1123–1138.
- [21] A. Korepanov, *Rates in almost sure invariance principle for dynamical systems with some hyperbolicity*, Comm. Math. Phys. **363** (2018), 173–190.
- [22] J. Kuelbs, W. Philipp, *Almost sure invariance principles for partial sums of mixing  $B$ -valued random variables*, Ann. Probab. **8** (1980), 1003–1036.
- [23] T. Lindvall, *On Coupling of Discrete Renewal Processes*, Z. Wahrscheinlichkeitstheor. verw. Geb. **48** (1979), 57–70.
- [24] I. Melbourne, M. Nicol, *Almost sure invariance principle for nonuniformly hyperbolic systems*, Comm. Math. Phys. **260** (2005), 131–146.
- [25] I. Melbourne, M. Nicol, *A vector-valued almost sure invariance principle for hyperbolic dynamical systems*, Ann. Probab. **37** (2009), 478–505.
- [26] I. Melbourne, P. Varandas, *A note on statistical properties for nonuniformly hyperbolic systems with slow contraction and expansion*, Stoch. Dyn. **16** (2016).

- [27] F. Merlevède, E. Rio, *Strong approximation of partial sums under dependence conditions with application to dynamical systems*, Stochastic Process. Appl. **122** (2012), no. 1, 386–417.
- [28] W. Philipp, W. Stout, *Almost sure invariance principles for partial sums of weakly dependent random variables*, Amer. Math. Soc. Mem. **161** (1975).
- [29] A. Rio, Asymptotic theory of weakly dependent random processes. Translated from the 2000 French edition. Probability Theory and Stochastic Modelling, 80. Springer, Berlin, 2017. xviii+204 pp.
- [30] V.A. Rokhlin, *Exact endomorphisms of a Lebesgue space*, Amer. Math. Soc. **39** (1964), 1–36.
- [31] O. Sarig, Lecture notes on ergodic theory. <https://www.weizmann.ac.il/math/sarigo/lecture-notes>
- [32] Y.G. Sinai, *Gibbs measures in ergodic theory*, Russ. Math. Surv. **27** (1972), 21–70.
- [33] M. P. Wojtkowski, *A system of one dimensional balls with gravity*, Comm. Math. Phys. **126** (1990), 507–533.
- [34] L.-S. Young, *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. **147** (1998), 585–650.
- [35] A.Yu. Zaitsev, *The accuracy of strong Gaussian approximation for sums of independent random vectors*, Russian Math. Surveys **68** (2013), 721–761.