Commun. Math. Phys. 363, 173–190 (2018)
Digital Object Identifier (DOI) https://doi.org/10.1007/s00220-018-3234-5

Communications in Mathematical Physics

Rates in Almost Sure Invariance Principle for Dynamical Systems with Some Hyperbolicity

Alexey Korepanov

University of Exeter, Exeter, UK. E-mail: kaikaikai@yandex.ru

Received: 1 September 2017 / Accepted: 18 June 2018
Published online: 22 August 2018 –

Abstract: We prove the almost sure invariance principle with rate $o(n^\varepsilon)$ for every $\varepsilon > 0$ for Hölder continuous observables on nonuniformly expanding and nonuniformly hyperbolic transformations with exponential tails. Examples include Gibbs–Markov maps with big images, Axiom A diffeomorphisms, dispersing billiards and a class of logistic and Hénon maps. The best previously proved rate is $O(n^{1/4}(\log n)^{1/2}(\log\log n)^{1/4})$. As a part of our method, we show that nonuniformly expanding transformations are factors of Markov shifts with simple structure and natural metric (similar to the classical Young towers). The factor map is Lipschitz continuous and probability measure preserving. For this we do not require the exponential tails.

1. Introduction

Definition 1.1. We say that a random process $X_0, X_1, \dots$ satisfies the *Almost Sure Invariance Principle* (ASIP) with rate, say $o(n^\varepsilon)$ with $\varepsilon \in (0, 1/2)$, if without changing the distribution, $\{X_n, n \geq 0\}$ can be redefined on a new probability space with a Brownian motion W_t such that

$$X_n = W_n + o(n^\varepsilon) \qquad \text{almost surely.}$$

The ASIP is a strong statistical property. It implies directly the functional central limit theorem, the functional law of iterated logarithm and other statistical laws, see Philipp and Stout [22, Chapter 1]. The rate in the ASIP has additional powerful implications, see Berkes, Liu and Wu [1] and references therein.

Suppose that $T : \Lambda \to \Lambda$ is a nonuniformly expanding or nonuniformly hyperbolic transformation as in Young [26,27] with *exponential tails* (see Sect. 2), such as Gibbs–Markov maps with big images, Axiom A diffeomorphisms, dispersing billiards, and a class of logistic and Hénon maps.

Suppose that ν is the unique T-invariant ergodic physical measure, $v\colon \Lambda \to \mathbb{R}$ is a Hölder continuous observable with $\int v\,d\nu = 0$ and $v_n = \sum_{k=0}^{n-1} v \circ T^k$. Then v_n, $n \geq 0$ is a random process with stationary increments on the probability space (Λ, ν).

We prove that v_n satisfies the ASIP with rate $o(n^\varepsilon)$ for every $\varepsilon > 0$. Our results strongly improve the best previously available rates.

Remark 1.2. Our analysis is restricted to discrete time $\mathbb{R}$-valued processes. The ASIP with good rates for flows and $\mathbb{R}^d$-valued processes with dependent increments, such as those in dynamical systems, is an important open problem.

Remark 1.3. We only consider processes with bounded increments. This is automatic for dynamical systems with Hölder continuous observables as above.

The ASIP has been introduced by Strassen [24,25], proved for processes with independent increments and martingales using the Skorokhod embedding. Approximations with martingales turned out to be very robust, see Philipp and Stout [22]; they have been used to prove the ASIP for various dynamical systems [6–9,11,18,20], including the nonuniformly expanding and nonuniformly hyperbolic maps in [18].

A downside of the martingale method is that the best rate in the ASIP which the Skorokhod embedding can produce is $O(n^{1/4}(\log n)^{1/2}(\log\log n)^{1/4})$, see Kiefer [12]. For nonuniformly expanding and nonuniformly hyperbolic systems, this rate has been achieved by Cuny and Merlevède [7] and Korepanov, Kosloff, and Melbourne [16].

For processes with independent and identically distributed increments, Komlós, Major, and Tusnády in their celebrated work [13] proved the ASIP with a much better rate $O(\log n)$, which is in fact unimprovable. Their proof is based on the so-called Hungarian construction and uses the quantile transform rather than the Skorokhod embedding.

For processes with dependent increments, it is also possible to prove the ASIP without relying on the Skorokhod embedding, but getting good rates proved to be challenging. For instance, Gouëzel [10] used blocking techniques to construct an approximation with a process with independent increments, for which the ASIP with the optimal rate $O(\log n)$ is known. However, an efficient control of the approximation error is tricky, and the best rate he could reach was $o(n^{1/4+\varepsilon})$ for every $\varepsilon > 0$, roughly the same as in the martingale method. For different reasons, $o(n^{1/4+\varepsilon})$ was not surpassed by various other methods [2,17,19,20].

In the dependent setting, the rate $O(n^{1/4}(\log n)^{1/2}(\log\log n)^{1/4})$ was unbeaten until very recently. First, Berkes, Liu, and Wu [1] proved the ASIP with rate $o(n^\varepsilon)$ for every $\varepsilon > 0$ for processes generated by a Bernoulli shift:

$$X_n = \sum_{k=0}^{n-1} \psi(\ldots, \xi_{k-1}, \xi_k, \xi_{k+1}, \ldots),$$

where $\{\xi_k\}$ is a sequence of independent identically distributed random variables and ψ is a sufficiently nice function. Their result is based on an insightful approximation by a process with independent increments and a Komlós–Major–Tusnády type result for processes with independent but not identically distributed increments by Sakhanenko [23]. Soon after, Merlevède and Rio [21] obtained the rate $O(\log n)$ for Harris recurrent geometrically ergodic Markov chains, strongly using the Markovian structure and in particular the regeneration technique.

The result of [1] readily covers some smooth dynamical systems such as the doubling map $x \mapsto 2x \pmod 1$, whose natural symbolic coding is a Bernoulli shift. But such systems are rare: for instance, they do not include smooth perturbations of the doubling map.

In the present work we extend the result of [1] to a large class of widely studied dynamical systems. Our strategy is to construct a semiconjugacy between a dynamical system in question and a Bernoulli shift. The semiconjugacy preserves the probability measure and sufficient structure for verification of assumptions of [1].

Remark 1.4. The historical overview above focuses on what is immediately relevant to our goals, without any attempt to describe the vast literature on the ASIP. For a thorough description of rates related results, see [1] and the review by Zaitsev [28].

Remark 1.5. Chernov and Haskell [4] prove the *Bernoulli property* for K-mixing nonuniformly hyperbolic maps. That is, such maps are measure-theoretically isomorphic to Bernoulli shifts. They remark that even though the Bernoulli property is a characterization of extreme chaotic behavior, it is not helpful in proving statistical properties like the central limit theorem. This is because a measure-theoretic isomorphism alone does not have to preserve any useful information about the structure of the space, such as metric or coordinates.

In contrast, we build a semiconjugacy to a Bernoulli shift which preserves enough information to prove the ASIP.

Remark 1.6. As an essential part of our proof, for a nonuniformly expanding dynamical system we construct an extension which is a renewal Markov shift, so that the factor map is Lipschitz with respect to a natural metric. Our construction is inspired by the coupling lemma for dispersing billiards as it appears in Chernov and Markarian [5, Lemma 7.24].

After circulating the first version of this paper, the author has been made aware of the work by Zweimüller [29], where he shows that nonuniformly expanding dynamical systems are similar to renewal Markov shifts.

Two dynamical systems are similar if they are factors of a common extension. All probability measure preserving systems are trivially similar, but in infinite ergodic theory the similarity is a highly nontrivial relation. The focus of [29] is on infinite measure preserving systems.

Our construction is remarkably similar to the one in [29], although we draw rather different conclusions: we make observations which allow us to treat probability measure preserving systems.

2. Statement of the Result

We use notation $\mathbb{N} = \{1, 2, \ldots\}$ and $\mathbb{N}_0 = \{0, 1, \ldots\}$. All functions, subsets, and partitions are assumed to be measurable. When we work with metric spaces, the default sigma algebra is Borel, and for finite and countable spaces the sigma algebra is discrete.

Let (Λ, d_Λ) be a bounded metric space and $T \colon \Lambda \to \Lambda$ be a transformation. Let Y be a subset of Λ and m be a probability measure on Y. Let α be an at most countable partition of Y (modulo a zero measure set) such that $m(a) > 0$ for all $a \in \alpha$.

Let $\tau \colon Y \to \mathbb{N}$ be an integrable function which is constant on each $a \in \alpha$ with value $\tau(a)$ such that $T^{\tau(a)}(y) \in Y$ for every $y \in a$, $a \in \alpha$. Let $F \colon Y \to Y$, $F(y) = T^{\tau(y)}(y)$ be the induced map.

We assume that for each $a \in \alpha$, the map F restricts to a (measure-theoretic) bijection from a to Y. Further, there are constants $0 < \eta \leq 1$, $\lambda > 1$ and $K, K_\tau \geq 1$ such that for all $a \in \alpha$ and $x, y \in a$:

- $d_\Lambda(F(x), F(y)) \geq \lambda d(x, y)$,
- $d_\Lambda(T^\ell(x), T^\ell(y)) \leq K_\tau \, d_\Lambda(F(x), F(y))$ for all $0 \leq \ell < \tau(a)$,
- the restriction $F \colon a \to Y$ is nonsingular and its inverse Jacobian $\zeta_a = \frac{dm}{dm \circ F}$ satisfies

$$|\log \zeta_a(x) - \log \zeta_a(y)| \leq K d_\Lambda^\eta(F(x), F(y)). \tag{1}$$

Finally, we assume that the induced map $F \colon Y \to Y$ allows a non-pathological coding by elements of α. We require that the set

$$\{(a_0, a_1, \dots) \in \alpha^{\mathbb{N}_0} : \text{there exists } y \in Y \text{ with } F^k(y) \in a_k \text{ for all } k\}$$

is measurable in $\alpha^{\mathbb{N}_0}$ (in the product topology with Borel sigma algebra).

We say that $T \colon \Lambda \to \Lambda$ as above is a *nonuniformly expanding* map. We say that it has *exponential (return time) tails,* if $\int_Y e^{\beta\tau} \, dm < \infty$ with some $\beta > 0$.

It is standard [27] that there is a unique T-invariant ergodic probability measure on Λ, with respect to which m is absolutely continuous. We denote this measure by ν.

For on observable $v \colon \Lambda \to \mathbb{R}$, denote

$$|v|_\infty = \sup_{x \in \Lambda} |v(x)|, \quad |v|_\eta = \sup_{x \neq y \in \Lambda} \frac{|v(x) - v(y)|}{d^\eta(x, y)} \quad \text{and} \quad \|v\|_\eta = |v|_\infty + |v|_\eta.$$

We say that v is *centered*, if $\int v \, d\nu = 0$, and that v is Hölder, if $\|v\|_\eta < \infty$.

Our main result is:

Theorem 2.1. *Suppose that there exists $\beta > 0$ such that $\int_Y e^{\beta\tau} \, dm < \infty$. If $v \colon \Lambda \to \mathbb{R}$ is a Hölder centered observable, then the process $v_n = \sum_{k=0}^{n-1} v \circ T^k$, defined on the probability space (Λ, ν), satisfies the ASIP with rate $o(n^\varepsilon)$ for every $\varepsilon > 0$.*

Remark 2.2. For maps which are naturally a Bernoulli shift, such as the doubling map, Theorem 2.1 follows directly from [1]. Our result is new for smooth perturbations of the doubling map and, for example, for:

- smooth expanding circle maps,
- Gibbs–Markov maps with big images,
- unimodal maps such as logistic with Collet–Eckmann parameters [3].

Remark 2.3. In nonuniformly hyperbolic maps with exponential tails and uniform contraction along stable leaves, as in Young [26], Hölder observables reduce to Hölder observables on a nonuniformly expanding quotient system through a bounded coboundary. A detailed exposition can be found in [16, Section 5]. Thus, Theorem 2.1 implies the ASIP with rate $o(n^\varepsilon)$ for every $\varepsilon > 0$ for maps such as:

- Anosov and Axiom A diffeomorphisms,
- dispersing billiards,
- Hénon maps with Benedicks-Carleson parameters,
- Lozi maps.

The paper is organized as follows: in Sect. 3 we introduce the notion of *Markov Young towers* and state Theorem 3.4 which establishes a semiconjugacy between $T \colon \Lambda \to \Lambda$ and a Markov Young tower. Theorem 3.4 is proved in Sect. 4. We prove Theorem 2.1 in Sect. 5.

3. Markov Young Towers

Suppose that

- $(\mathcal{A}, \mathbb{P}_{\mathcal{A}})$ is a finite or countable probability space,
- $h_{\mathcal{A}}: \mathcal{A} \to \mathbb{N}$ is an integrable function,
- $0 < \xi < 1$ is a constant.

Define a probability space $(X, \mathbb{P}_X) = (\mathcal{A}^{\mathbb{N}}, \mathbb{P}_{\mathcal{A}}^{\mathbb{N}})$ and let $f_X: X \to X$ be the left shift,

$$f_X(a_0, a_1, \dots) = (a_1, a_2, \dots).$$

Define $h: X \to \mathbb{N}$, $h(a_0, a_1, \dots) = h_{\mathcal{A}}(a_0)$. Let $f: \Delta \to \Delta$ be a suspension over $f_X: X \to X$ with a roof function h, i.e.

$$\Delta = \{(x, \ell) \in X \times \mathbb{Z} : 0 \le \ell < h(x)\}$$

$$f(x, \ell) = \begin{cases} (x, \ell + 1), & \ell < h(x) - 1, \\ (f_X(x), 0), & \ell = h(x) - 1 \end{cases}. \tag{2}$$

Define a distance d on X by $d(x, y) = \xi^{s(x,y)}$, where $s: X \times X \to \mathbb{N}_0$ is the separation time,

$$s((a_0, a_1, \dots), (b_0, b_1, \dots)) = \inf\{j \ge 0 : a_j \ne b_j\}.$$

Let d also denote the natural compatible distance on Δ:

$$d((x, k), (y, j)) = \begin{cases} 1, & k \ne j \\ d(x, y), & k = j \end{cases}. \tag{3}$$

Let $\bar{h} = \int h \, dm$. Let $\mathbb{P}$ be the probability measure on Δ given by $\mathbb{P}(A \times \{\ell\}) = \bar{h}^{-1} m(A)$ for all $\ell \ge 0$ and $A \subset \{y \in Y : h(y) \ge \ell + 1\}$. Note that $\mathbb{P}$ is f-invariant.

Let $\Delta_k = \{(y, \ell) \in \Delta : \ell = k\}$. Then X is naturally identified with Δ_0, which we refer to as the *base* of the suspension, and $\mathbb{P}_X$, f_X have their counterparts on Δ_0, which we also denote $\mathbb{P}_X$, f_X.

Definition 3.1. We call the map $f: \Delta \to \Delta$ as above a (non-invertible) *Markov Young tower.*

Remark 3.2. To define a Markov Young tower, we need an at most countable probability space $(\mathcal{A}, \mathbb{P}_{\mathcal{A}})$, an integrable function $h_{\mathcal{A}}: \mathcal{A} \to \mathbb{N}$ and a constant $0 < \xi < 1$. Further, we always use notation for Markov Young towers as above, i.e. with the symbols f, Δ, $\mathcal{A}$, $\mathbb{P}_{\mathcal{A}}$, X, $\mathbb{P}_X$, f_X, h, $\mathbb{P}$, d, ξ.

Remark 3.3. Similar to the classical Young towers, our Markov Young towers are very simple objects on their own, studied by various people under different names. We chose the term "Markov Young tower" because for both the key property is not just their own structure but the relation to a large class of dynamical systems (see Theorem 3.4 below). It allowed Young [26] to prove the exponential decay of correlations for dispersing billiards among other maps, and it is an essential ingredient in the proof of our main result.

Our key technical result is:

Theorem 3.4. *Suppose that* $T: \Lambda \to \Lambda$ *is a nonuniformly expanding map. Then there exists a Markov Young tower* $f: \Delta \to \Delta$ *and a map* $\pi: \Delta \to \Lambda$*, defined* $\mathbb{P}$*-almost everywhere, such that*

- π *is Lipschitz:*

$$d_\Lambda(\pi(x), \pi(y)) \le C_\Lambda d(x, y),$$

  where $C_\Lambda = \lambda K_\tau \operatorname{diam} \Lambda$*,*
- π *is a semiconjugacy:* $\mathbb{P}$*-almost surely,* $T \circ \pi = \pi \circ f$*,*
- π *preserves the probability measures:* $\pi_* \mathbb{P}_X = m$ *and* $\pi_* \mathbb{P} = \nu$*.*

In addition, moments of h *are closely related to those of* τ*:*

- *(Weak polynomial moments) If there exist* $C_\tau > 0$ *and* $\beta > 1$ *such that* $m(\tau \ge \ell) \le C_\tau \ell^{-\beta}$ *for all* $\ell \ge 1$*, then* $\mathbb{P}_X(h \ge \ell) \le C\ell^{-\beta}$ *for all* $\ell \ge 1$*, where the constant* C *continuously depends on* C_τ*,* β*,* λ*,* K *and* η*.*
- *(Strong polynomial moments) If there exist constants* $C_\tau > 0$ *and* $\beta > 1$ *such that* $\int \tau^\beta \, dm \le C_\tau$*, then* $\int h^\beta \, d\mathbb{P}_X \le C$*, where the constant* C *continuously depends on* C_τ*,* β*,* λ*,* K *and* η*.*
- *(Exponential and stretched exponential moments) If there exist constants* $C_\tau > 0$*,* $\beta > 0$ *and* $\gamma \in (0, 1]$ *such that* $\int e^{\beta\tau^\gamma} \, dm \le C_\tau$*, then* $\int e^{\beta' h^\gamma} \, d\mathbb{P}_X \le C$*, where the constants* $\beta' \in (0, \beta]$ *and* $C > 0$ *depend continuously on* C_τ*,* β*,* γ*,* λ*,* K *and* η*.*
- *(Exactly exponential moments) If* $\int e^{\beta\tau} \, dm < \infty$ *for some* $\beta > 0$*, then* $f: \Delta \to \Delta$ *can be constructed so that*

$$\mathbb{P}_X(h = n) = \begin{cases} \theta(1-\theta)^{-1}(1-\theta)^{n/N}, & n \in \{N, 2N, 3N, \dots\} \\ 0, & \textit{else} \end{cases}$$

  with some $0 < \theta < 1$ *and* $N \ge 1$*.*

Remark 3.5. The exact exponential moments in Theorem 3.4 allow us to represent in a natural way $f: \Delta \to \Delta$ as a factor of a Bernoulli shift and use [1] to prove the ASIP, see Sect. 5. Our results are limited to $\int e^{\beta\tau} \, dm < \infty$, because without the exact exponential moments such a representation does not work.

4. Proof of Theorem 3.4

For $v: \Delta \to \mathbb{R}$ and $\eta \in (0, 1]$, denote

$$|v|_\infty = \sup_{x \in \Delta} |v(x)|, \quad |v|_\eta = \sup_{x \ne y \in \Delta} \frac{|v(x) - v(y)|}{d^\eta(x, y)} \quad \text{and} \quad \|v\|_\eta = |v|_\infty + |v|_\eta.$$

4.1. Construction of Markov Young tower. We define $\mathcal{A}$ as the set of all finite words in the alphabet α (not including the empty word). For $w = a_0 \dots a_{n-1} \in \mathcal{A}$ we define

$$|w| = n \qquad \text{and} \qquad h_{\mathcal{A}}(w) = \tau(a_0) + \dots + \tau(a_{n-1}).$$

Let

$$Y_w = \{y \in Y : T^k \in a_k \text{ for all } 0 \le k \le n-1\}.$$

We use the measure $\mathbb{P}_{\mathcal{A}}$ from the following lemma:

Lemma 4.1. *There exists a probability measure $\mathbb{P}_{\mathcal{A}}$ on $\mathcal{A}$ and a disintegration $m = \sum_{w \in \mathcal{A}} \mathbb{P}_{\mathcal{A}}(w) m_w$, where m_w are probability measures on Y, such that for every $w \in \mathcal{A}$,*

- *m_w is supported on Y_w,*
- *$(T^{h_{\mathcal{A}}(w)})_* m_w = m$.*

In addition,

- *If there exist $C_\tau > 0$ and $\beta > 1$ such that $m(\tau \ge \ell) \le C_\tau \ell^{-\beta}$ for all $\ell \ge 1$, then $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} \ge \ell) \le C \ell^{-\beta}$ for all $\ell \ge 1$, where the constant C continuously depends on C_τ, β, λ, K and η.*
- *If there exist constants $C_\tau > 0$ and $\beta > 1$ such that $\int \tau^\beta \, dm \le C_\tau$, then $\int h_{\mathcal{A}}^\beta \, d\mathbb{P}_{\mathcal{A}} \le C$, where the constant C continuously depends on C_τ, β, λ, K and η.*
- *If there exist constants $C_\tau > 0$, $\beta > 0$ and $\gamma \in (0, 1]$ such that $\int e^{\beta \tau^\gamma} \, dm \le C_\tau$, then $\int e^{\beta' h_{\mathcal{A}}^\gamma} \, d\mathbb{P}_{\mathcal{A}} \le C$, where the constants $\beta' \in (0, \beta]$ and $C > 0$ depend continuously on C_τ, β, γ, λ, K and η.*

Remark 4.2. Our Lemma 4.1 corresponds to [29, Theorem 2], where the disintegration of m is called a *regenerative partition of unity.* For the ease of citation and explicit tail estimates, we refer to [14].

Proof of Lemma 4.1. Such a decomposition is constructed in [14, Section 4]. It is implicit in [14] that m_w is supported on Y_w.

We remark that in [14], the set $\mathcal{A}$ contains the empty word, while here we do not allow it. This, however, does not cause problems, because if w is the empty word, then $\mathbb{P}_{\mathcal{A}}(w)$ is uniformly bounded away from 1 and $m_w = m$. Thus the decomposition with the empty word translates to one without, with the same moment bounds. □

For the exactly exponential moments in Theorem 3.4, we obtain a special version of Lemma 4.1:

Lemma 4.3. *Suppose that $\int e^{\beta\tau} \, dm < \infty$ with some $\beta > 0$. Then the measure $\mathbb{P}_{\mathcal{A}}$ in Lemma 4.1 can be chosen so that*

$$\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} = \ell) = \begin{cases} \theta^{-1}(1-\theta)\theta^{\ell/N}, & \ell \in N\mathbb{N} \\ 0, & \text{else} \end{cases}$$

with some $N \in \mathbb{N}$ and $0 < \theta < 1$.

Our proof of Lemma 4.3 uses a rather delicate technical adaptation of the argument in [15, Section 4]. It is carried out in Appendix A.

Let $\mathbb{P}_{\mathcal{A}}$ and $\{m_w\}$ be as in Lemmas 4.1 or 4.3. Let $\xi = \lambda^{-1}$. According to Remark 3.2, $\mathcal{A}$, $\mathbb{P}_{\mathcal{A}}$, $h_{\mathcal{A}}$ and ξ define a Markov Young tower $f \colon \Delta \to \Delta$. To prove Theorem 3.4, it remains to construct the semiconjugacy $\pi \colon \Delta \to \Lambda$.

4.2. Semiconjugacy. Let $\iota \colon Y \to \alpha^{\mathbb{N}_0}$ be the natural embedding, $\iota(y) = (a_0, a_1, \ldots)$ if $F^k(y) \in a_k$ for all k. (Technically, ι is defined on a full measure subset of Y.) The space $\alpha^{\mathbb{N}_0}$ is supplied with the product topology and Borel sigma algebra.

Remark 4.4. The map ι is measurable and injective by construction; in addition we assumed that $\iota(Y)$ is measurable in $\alpha^{\mathbb{N}_0}$. It is straightforward to check that ι^{-1} is continuous on $\iota(Y)$, and that $\iota(A)$ is measurable for all measurable $A \subset Y$. Hence ι is *bimeasurable:* both images and preimages of measurable sets are measurable.

Let $m_\alpha = \iota_* m$. This is a Borel probability measure on $\alpha^{\mathbb{N}_0}$ with $m_\alpha(\alpha^{\mathbb{N}_0} \setminus \iota(Y)) = 0$.

For words $w_0, \dots, w_n \in \mathcal{A}$, let $w_0 \cdots w_n$ denote their concatenation. Then $|w_0 \cdots w_n| = |w_0| + \cdots + |w_n|$ and $h_{\mathcal{A}}(w_0 \cdots w_n) = h_{\mathcal{A}}(w_0) + \cdots + h_{\mathcal{A}}(w_n)$.

For $x = (w_0, w_1, \dots) \in X$, let $\pi_\alpha(x) \in \alpha^{\mathbb{N}_0}$ denote the sequence of elements of α obtained by concatenating all w_k, $k \geq 0$. It is clear that thus defined $\pi_\alpha \colon X \to \alpha^{\mathbb{N}_0}$ is continuous.

Proposition 4.5. $(\pi_\alpha)_* \mathbb{P}_X = m_\alpha$.

Proof. Recall that we have the disintegration $m = \sum_{w \in \mathcal{A}} \mathbb{P}_{\mathcal{A}}(w) m_w$.

Let $w_0 \in \mathcal{A}$. Since $F^{|w_0|} \colon Y_{w_0} \to Y$ is a bijection and $F_*^{|w_0|} m_{w_0} = m$, we can write $m_{w_0} = \sum_{w_1 \in \mathcal{A}} \mathbb{P}_{\mathcal{A}}(w_1) m_{w_0, w_1}$, where m_{w_0,w_1} are probability measures supported on $Y_{w_0 w_1}$ such that $F_*^{|w_0 w_1|} m_{w_0,w_1} = m$. Continuing with m_{w_0,w_1} and further recursively, we obtain for each $n \geq 1$ a disintegration

$$m = \sum_{w_0, \dots, w_n \in \mathcal{A}} \mathbb{P}_{\mathcal{A}}(w_0) \cdots \mathbb{P}_{\mathcal{A}}(w_n) m_{w_0, \dots, w_n},$$

where $m_{w_0,\dots,w_n}$ are probability measures supported on $Y_{w_0 \cdots w_n}$ such that $F_*^{|w_0 \cdots w_n|} m_{w_0,\dots,w_n} = m$.

Taking images under $\iota \colon Y \to \alpha^{\mathbb{N}_0}$, we obtain a similar disintegration in $\alpha^{\mathbb{N}_0}$:

$$m_\alpha = \sum_{w_0, \dots, w_n \in \mathcal{A}} \mathbb{P}_{\mathcal{A}}(w_0) \cdots \mathbb{P}_{\mathcal{A}}(w_n) m_{\alpha;\, w_0, \dots, w_n},$$

where $m_{\alpha;\, w_0,\dots,w_n} = \iota_* m_{\alpha;\, w_0,\dots,w_n}$ are probability measures supported on the cylinders $\alpha^{\mathbb{N}_0}_{w_0 \cdots w_n}$ with

$$\alpha^{\mathbb{N}_0}_w = \{(a_0, a_1, \dots) \in \alpha^{\mathbb{N}_0} : a_0 \dots a_{|w|-1} = w\}.$$

Let $w \in \mathcal{A}$ and $n = |w|$. Then for all $w_0, \dots, w_n \in \mathcal{A}$, either $\alpha^{\mathbb{N}_0}_{w_0 \cdots w_n} \subset \alpha^{\mathbb{N}_0}_w$ or $\alpha^{\mathbb{N}_0}_{w_0 \cdots w_n} \cap \alpha^{\mathbb{N}_0}_w = \emptyset$. Thus

$$m_\alpha(\alpha^{\mathbb{N}_0}_w) = \sum_{\substack{w_0, \dots, w_n \in \mathcal{A}: \\ \alpha^{\mathbb{N}_0}_{w_0 \cdots w_n} \subset \alpha^{\mathbb{N}_0}_w}} \mathbb{P}_{\mathcal{A}}(w_0) \cdots \mathbb{P}_{\mathcal{A}}(w_n) = \mathbb{P}_X(\pi_\alpha^{-1}(\alpha^{\mathbb{N}_0}_w)).$$

Thus $(\pi_\alpha)_* \mathbb{P}_X$ agrees with m_α on all cylinders in $\alpha^{\mathbb{N}_0}$. By Carathéodory's extension theorem, $(\pi_\alpha)_* \mathbb{P}_X = m_\alpha$. □

Let $\pi_X \colon X \to Y$, $\pi_X = \iota^{-1} \circ \pi_\alpha$.

Proposition 4.6. *π_X is well defined $\mathbb{P}_X$ almost everywhere on X and is measurable. Also, $(\pi_X)_* \mathbb{P}_X = m$.*

Proof. The map ι is injective, which allows us to define π_X on $X' = (\pi_\alpha^{-1} \circ \iota)(Y)$.

Recall that $\iota(Y)$ is measurable in $\alpha^{\mathbb{N}_0}$ and $m_\alpha(\iota(Y)) = 1$. The map π_α is continuous, so X' is a measurable subset of X and, by Proposition 4.5, $\mathbb{P}_X(X') = 1$. Hence π_X is defined almost everywhere.

Using the bimeasurability of ι and Proposition 4.5, for every measurable $A \subset Y$,

$$\mathbb{P}_X(\pi_X^{-1}(A)) = \mathbb{P}_X((\pi_\alpha^{-1} \circ \iota)(A)) = m_\alpha(\iota(A)) = m(A).$$

In other words, $(\pi_X)_* \mathbb{P}_X = m$. □

Remark 4.7. Further we silently ignore the zero measure subset of X, on which π_X is not defined, and the corresponding subset of Δ, which also has zero measure.

Define $\pi : \Delta \to \Lambda$ by

$$\pi((w_0, w_1, \ldots), \ell) = T^\ell(\pi_X(w_0, w_1, \ldots)). \tag{4}$$

Then $\pi_X : \Delta_0 \to Y$ is a restriction of π.

Proposition 4.8. *π is Lipschitz: for all $a, b \in \Delta$,*

$$d_\Lambda(\pi(a), \pi(b)) \le C_\Lambda\, d(a, b),$$

where $C_\Lambda = \lambda K_\tau$ diam Λ.

Proof. Let $a = (x_1, j)$ and $b = (x_2, k)$, where

$$x_1 = (w_{1,0}, w_{1,1}, \ldots) \quad \text{and} \quad x_2 = (w_{2,0}, w_{2,1}, \ldots).$$

If $j \ne k$ or $w_{1,0} \ne w_{2,0}$, then $d(a, b) = 1$ and the statement is trivial.

Suppose now that $j = k$ and $w_{1,0} = w_{2,0}$. Let $n = s(x_1, x_2)$. Note that $n \ge 1$ and

$$j = k < h(x_1) = h(x_2) = h_{\mathcal{A}}(w_{1,0}) = h_{\mathcal{A}}(w_{2,0}).$$

Observe that $\pi_X(x_i) \in Y_{w_{1,0}\cdots w_{1,n-1}}$ and $F(\pi_X(x_i)) \in Y_{w_{1,1}\cdots w_{1,n-1}}$ for $i = 1, 2$. Also, diam $Y_{w_{1,1}\cdots w_{1,n-1}} \le \lambda^{-(n-1)}$ diam Y. Then

$$\begin{aligned} d_\Lambda(\pi(a), \pi(b)) &= d_\Lambda\big(T^j(\pi_X(x_1)), T^j(\pi_X(x_2))\big) \le K_\tau \operatorname{diam} Y_{w_{1,1}\cdots w_{1,n-1}} \\ &\le K_\tau \lambda^{-(n-1)} \operatorname{diam} Y = \lambda K_\tau \operatorname{diam} Y\, d(a, b). \end{aligned}$$

□

Proposition 4.9. $T \circ \pi = \pi \circ f$.

Proof. Suppose that $a = (x, \ell) \in \Delta$, and $x = (w_0, w_1, \ldots)$. If $\ell < h(x) - 1$, then $f(a) = (x, \ell + 1)$ and

$$\pi(f(a)) = T^{\ell+1}(\pi_X(x)) = T(\pi(a)).$$

If $\ell = h(x) - 1$, then

$$\pi(f(a)) = \pi_X(f_X(x)) = F(\pi_X(x)) = T^{\ell+1}(\pi_X(x)) = T(\pi(a)).$$

Thus $\pi(f(a)) = T(\pi(a))$. □

Proposition 4.10. $\pi_* \mathbb{P} = \nu$.

Proof. We use the fact that ν is the unique T-invariant ergodic probability measure on Λ, with respect to which m is absolutely continuous.

Since $\mathbb{P}$ is f-invariant and ergodic, it follows from Proposition 4.9 that $\pi_* \mathbb{P}$ is T-invariant and ergodic. Since $\mathbb{P}_X$ is absolutely continuous with respect to $\mathbb{P}$ and $\pi_* \mathbb{P}_X = m$, using Proposition 4.6 we obtain that m is absolutely continuous with respect to $\pi_* \mathbb{P}$. Thus $\pi_* \mathbb{P} = \nu$. □

5. Proof of Theorem 2.1

5.1. ASIP for Bernoulli shift. Suppose that $\{\varepsilon_k\}_{k\in\mathbb{Z}}$ is a sequence of independent identically distributed random variables, and X_k are real valued random variables with mean zero given by

$$X_k = G(\dots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \dots)$$

for some function G.

Let $\{\varepsilon'_k\}$ be an independent copy of $\{\varepsilon_k\}$ and for $\ell \in \mathbb{Z}$ define $\{\varepsilon^\ell_k\}_{k\in\mathbb{Z}}$ by

$$\varepsilon^\ell_k = \begin{cases} \varepsilon_k, & k \neq \ell, \\ \varepsilon'_\ell, & k = \ell. \end{cases}$$

Define

$$X^\ell_k = G(\dots, \varepsilon^\ell_{k-1}, \varepsilon^\ell_k, \varepsilon^\ell_{k+1}, \dots).$$

Let $p > 4$,

$$\delta_{\ell,p} = \|X_0 - X^\ell_0\|_p \qquad \text{and} \qquad \Theta_{\ell,p} = \sum_{|k|\geq \ell} \delta_{k,p},$$

where $\|\cdot\|_p = \big(\mathbb{E}|\cdot|^p\big)^{1/p}$.

We use the following result [1, Theorem 2.1] (with [1, Corollary 2.1] to verify the assumptions):

Theorem 5.1. *If $\|X_k\|_p < \infty$ and $\Theta_{\ell,p} = o(\ell^{-p})$, then the partial sum process $\sum_{k=0}^{n-1} X_k$ satisfies the ASIP with rate $o(n^{1/p})$.*

Remark 5.2. Theorem 5.1 is proved in [1] under a more relaxed condition on $\Theta_{\ell,p}$. We use intentionally a suboptimal but easy to state result.

5.2. Construction of Bernoulli shift. Suppose that $f : \Delta \to \Delta$ is a Markov Young tower as in Sect. 3 with

$$\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} = n) = \begin{cases} \theta(1-\theta)^{-1}(1-\theta)^{n/N}, & n \in \{N, 2N, 3N, \dots\} \\ 0, & \text{else} \end{cases}$$

with $N \in \mathbb{N}$ and $0 < \theta < 1$. Let $v : \Delta \to \mathbb{R}$ be a centered Hölder observable and $v_n = \sum_{k=0}^{n-1} v \circ f^k$ be the corresponding random process on $(\Delta, \mathbb{P})$.

By Theorem 3.4, to prove Theorem 2.1 it is enough to show the ASIP for v_n.

The map f is N-periodic. For simplicity we assume that $N = 1$. We show how to remove this assumption in Sect. 5.4.

In the rest of this subsection we construct a suitable Bernoulli shift $\sigma : D \to D$ with a measure preserving semiconjugacy $g : D \to \Delta$. The random process $\sum_{k=0}^{n-1} X_k$ with $X_k = v \circ g \circ \sigma^k$ has the same distribution as v_n. If $\{\varepsilon_k\}$ are the coordinates of D, they are independent and identically distributed, and $X_k = (v \circ g)(\dots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \dots)$. This sets up a ground for the application of Theorem 5.1.

Let $(\Omega, \mathbb{P}_\Omega)$ be a probability space supporting random variables $A_n : \Omega \to \mathcal{A}, n \geq 1$, such that for $a \in \mathcal{A}$,

$$\mathbb{P}(A_n = a) = \begin{cases} 0, & h_{\mathcal{A}}(a) \neq n, \\ \frac{\mathbb{P}_{\mathcal{A}}(a)}{\mathbb{P}_{\mathcal{A}}(\bigcup\{a \in \mathcal{A} : h_{\mathcal{A}}(a) = n\})}, & h_{\mathcal{A}}(a) = n \end{cases}.$$

That is, A_n is a random element of $\mathcal{A}$ chosen among those with $h_{\mathcal{A}} = n$ with respect to the appropriately conditioned measure $\mathbb{P}_{\mathcal{A}}$.

Let $Z = \{0, 1\}$ and $\mathbb{P}_Z$ be the probability measure on Z given by $\mathbb{P}_Z(0) = 1 - \theta$ and $\mathbb{P}_Z(1) = \theta$.

Define $D = (\Omega \times Z)^{\mathbb{Z}}$ with the product probability measure $\mathbb{P}_D = (\mathbb{P}_\Omega \times \mathbb{P}_Z)^{\mathbb{Z}}$. Let $\varepsilon_k = (\omega_k, z_k)$ be the coordinates in D and $\sigma : D \to D$ be the left shift.

Let

$$t_0 = \sup\{k \leq 0 : z_k = 1\} \quad \text{and} \quad t_n = \inf\{k > t_{n-1} : z_k = 1\}, \quad n \geq 1.$$

Note that t_n are finite $\mathbb{P}_D$-almost surely.

Define $g : D \to \Delta$ by $g(\{\varepsilon_k\}) = (y, -t_0)$, where $y = (A_{t_1 - t_0}(\omega_{t_0}), A_{t_2 - t_1}(\omega_{t_1}), \ldots)$. Observe that g is a probability measure preserving semiconjugacy between $\sigma : D \to D$ and $T : \Delta \to \Delta$.

5.3. Weak dependence. Here we verify the assumptions of Theorem 5.1. As above, we set

$$X_k = (v \circ g)(\ldots, \varepsilon_{k-1}, \varepsilon_k, \varepsilon_{k+1}, \ldots).$$

Let $p > 4$. The observable v is Hölder continuous, thus $\|X_k\|_p \leq \|X_k\|_\infty < \infty$. It remains to prove that $\Theta_{\ell,p} = o(\ell^{-p})$.

Proposition 5.3. *There exists* $0 < \theta_\delta < 1$ *such that* $\delta_{\ell,p} = O(\theta_\delta^{|\ell|})$.

Proof. Let $\ell \in \mathbb{Z}$,

$$x = ((w_0, w_1, \ldots), r) = g(\{\varepsilon_k\}) \quad \text{and} \quad x^\ell = ((w_0^\ell, w_1^\ell, \ldots), r^\ell) = g(\{\varepsilon_k^\ell\}).$$

Suppose first that $\ell \geq 1$. Let $c_\ell = \sum_{j=1}^{\ell-1} z_j$. Then $w_j = w_j^\ell$ for all $0 \leq j \leq c_\ell - 1$. If $c_\ell \geq 1$, then $r = r^\ell$ and $d(x, x^\ell) \leq \xi^{c_\ell}$. If $c_\ell = 0$, then $d(x, x^\ell) \leq 1$. In either case,

$$d(x, x^\ell) \leq \xi^{c_\ell}.$$

Since $\{z_k\}$ are independent identically distributed,

$$\mathbb{E}\, d(x, x^\ell)^p \leq \mathbb{E}\, \xi^{p c_\ell} = \left(\mathbb{E}\, \xi^{p z_3}\right)^{\ell-1} = \left(1 - \theta + \theta \xi^p\right)^{\ell-1}.$$

Since v is Hölder continuous, $|X_0 - X_0^\ell| \leq |v|_\eta d(x, x^\ell)^\eta$ and the result for $\ell \geq 1$ follows.

Suppose now that $\ell \leq 0$. Then $x \neq x^\ell$ only when $t_0 \leq \ell$. The result follows from Hölder continuity of v and

$$\mathbb{P}(t_0 \leq \ell) = \mathbb{P}(z_0 = z_{-1} = \cdots = z_{\ell-1} = 0) = (1 - \theta)^\ell.$$

□

Finally, $\Theta_{\ell,p}$ decays exponentially in ℓ, because so does $\delta_{\ell,p}$. The proof of Theorem 2.1 is complete.

5.4. Periodic tower. In Sect. 5.2 we assumed that the tower $f\colon \Delta \to \Delta$ is aperiodic, namely that $N = 1$. Here we give a sketch of proof for $N > 1$.

Let

$$\Delta_N = \{(x, \ell) \in \Delta : \ell = 0 \pmod N\}.$$

We supply Δ_N with a probability measure $\mathbb{P}_N$, which is a (normalized) restriction of $\mathbb{P}$. Define a projection $\pi_N\colon \Delta \to \Delta_N$, $\pi_N(x, \ell) = \big(x, N\lfloor \frac{\ell}{N} \rfloor\big)$. Observe that $(\pi_N)_* \mathbb{P} = \mathbb{P}_N$.

Let

$$u_n = \sum_{k=0}^{n-1} u \circ f^{kN} \qquad \text{with} \qquad u = \sum_{k=0}^{N-1} v \circ f^k$$

be a process on the probability space $(\Delta_N, \mathbb{P}_N)$.

Since π_N is measure preserving and $|v_n - u_{\lfloor n/N \rfloor} \circ \pi_N| \le 2N|v|_\infty$, the processes v_n and u_n are naturally defined on a common probability space with $v_n = u_{\lfloor n/N \rfloor} + O(1)$ almost surely.

By the method which works for $N = 1$, we show the ASIP for the process u_n on the probability space $(\Delta_N, \mathbb{P}_N)$. So, there exists a Brownian motion W_n with $u_n = W_n + o(n^\varepsilon)$ almost surely for every $\varepsilon > 0$.

Thus, almost surely and for every $\varepsilon > 0$,

$$v_n = W_{\lfloor n/N \rfloor} + o(n^\varepsilon) = W'_n + o(n^\varepsilon),$$

where $W'_n = W_{n/N}$ is a Brownian motion. We used that $W_{n/N} - W_{\lfloor n/N \rfloor} = O(\log n)$ almost surely. This is the desired ASIP for v_n.

Acknowledgements. This research was supported in part by a European Advanced Grant *StochExtHomog* (ERC AdG 320977) at the University of Warwick and an Engineering and Physical Sciences Research Council grant EP/P034489/1 at the University of Exeter. The author is grateful to Mark Holland and Ian Melbourne for support. The author has been very lucky with the referees who were aware of the history of the problem and made invaluable suggestions for improvement of the manuscript. The author is thankful to Christophe Cuny, Jérôme Dedecker and Florence Merlevède for helpful comments and for pointing out a mistake in the construction of semiconjugacy.

A. Proof of Lemma 4.3

Our argument is based on [15, Section 4], and here we work in their notations, which are different from the rest of the paper.

In this section, $T\colon \Lambda \to \Lambda$ is a nonuniformly expanding map as in Sect. 2, $F\colon Y \to Y$ is the induced map and $f\colon \Delta \to \Delta$ is the Young tower,

$$\Delta = \{(y, \ell) \in Y \times \mathbb{Z} : 0 \le \ell < \tau(y)\},$$

$$f(y, \ell) = \begin{cases} (y, \ell + 1), & \ell < \tau(y) - 1, \\ (Fy, 0), & \ell = \tau(y) - 1. \end{cases}$$

Let $\bar\tau = \int_Y \tau \, dm$. Let m_Δ be the probability measure on Δ given by $m_\Delta(A \times \{\ell\}) = \bar\tau^{-1} m(A)$ for all $\ell \ge 0$ and $A \subset \{y \in Y : \tau(y) \ge \ell + 1\}$.

Let $L\colon L^1(m_\Delta)\to L^1(m_\Delta)$ be the transfer operator corresponding to f and m_Δ, so $\int L\phi\,\psi\,dm_\Delta=\int\phi\,\psi\circ f\,dm$ for all $\phi\in L^1$ and $\psi\in L^\infty$.

Without loss of generality we assume that f is mixing (otherwise we switch to a power of f which is mixing).

Remark A.1. The proof of decay of correlations in [15] is based on a construction of a probability space $(\mathbb{W},\mathbb{P}_\mathbb{W})$ and a random variable $r\colon\mathbb{W}\to\mathbb{N}$ such that each sufficiently regular observable $\psi\colon\Delta\to[0,\infty)$ with $\int\psi\,dm_\Delta=1$ can be decomposed into a sum $\psi=\sum_{w\in\mathbb{W}}\mathbb{P}_\mathbb{W}(w)\psi_w$ with $\int\psi_w\,dm_\Delta=1$ and $L^{r(w)}\psi_w=\bar\tau 1_{\Delta_0}$. In particular, this applies to $\psi=\bar\tau 1_{\Delta_0}$.

The distribution of r depends on the tails $m(\tau>n)$. There is quite a lot of flexibility in the construction. We show that if the tails decay exponentially, we can construct r distributed geometrically up to a period, as required for Lemma 4.3. Moreover, we can take $\mathbb{W}=\mathcal{A}$, $r=h_\mathcal{A}$ and ensure that the observables ψ_w are supported on the respective $Y_w\times\{0\}$. This yields the desired result, with m_w given by the densities ψ_w.

Remark A.2. Unfortunately, there is no easy way to point out what needs to be changed in [15]. We present a complete proof, referring to [15] in the proofs where possible.

Let $\Delta_\ell=\{(y,k)\in\Delta:k=\ell\}$. Recall that $\eta\in(0,1]$ is the exponent in (1). For $\psi\colon\Delta\to[0,\infty)$, define

$$|\psi|_{\eta,\ell}=\sup_{n\ge 0}\ \sup_{(y,n)\ne(y',n)\in\Delta_n}\frac{|\log\psi(y,n)-\log\psi(y',n)|}{d(y,y')^\eta},$$

where $\log 0=-\infty$ and $\log 0-\log 0=0$. Note that for a countable collection ψ_k of nonnegative functions, $\left|\sum_k\psi_k\right|_{\eta,\ell}\le\max_k|\psi_k|_{\eta,\ell}$.

For $a\in\alpha$, let $S_a=\{(y,k)\in\Delta:y\in a\text{ and }k=\tau(y)-1\}$, and let $\varkappa$ be the partition of Δ generated by $\{S_a\}_{a\in\alpha}$ and $\{\Delta_\ell\}_{\ell\ge 0}$. Let $\varkappa^n=\vee_{k=0}^{n-1}f^{-k}\varkappa$. Then $\varkappa^0$ is the trivial partition, and for every $n\ge 1$ and $a\in\varkappa^n$, there exists $\ell\ge 0$ such that $f^n\colon a\to\Delta_\ell$ is a bijection.

Fix constants $R>0$ and $\xi\in(0,e^{-R})$ such that $R(1-\xi e^R)\ge K+\lambda^{-1}R$.

Proposition A.3. *Suppose that* $\psi:\Delta\to[0,\infty)$ *with* $|\psi|_{\eta,\ell}\le R$. *Let* $n\ge 1$, $a\in\varkappa^n$ *and* $\psi_a=\psi 1_a$. *Then*

1. $e^{-R}\bar\tau\int_{\Delta_0}\psi\,dm_\Delta\le\psi\,1_{\Delta_0}\le e^R\bar\tau\int_{\Delta_0}\psi\,dm_\Delta$.
2. $|L^n\psi_a|_{\eta,\ell}\le R$.
3. *If* $t\in[0,\xi]$, *then* $\psi'_a=L^n\psi_a-t\,\bar\tau\int_{\Delta_0}L^n\psi_a\,dm_\Delta\,1_{\Delta_0}$ *is nonnegative and* $|\psi'_a|_{\eta,\ell}\le R$.

Proof. This is a minor modification of [15, Proposition 4.1]. □

Let $\mathcal{R}$ be the set of observables $\psi:\Delta\to[0,\infty)$ such that $|\psi|_\infty\le e^R\bar\tau\int_\Delta\psi\,dm_\Delta$ and $|\psi|_{\eta,\ell}\le R$.

For $n\ge 0$, let $\mathcal{R}^n$ denote the set of observables $\psi\colon\Delta\to[0,\infty)$ such that $L^n\psi\in\mathcal{R}$ and $|L^n(\psi 1_a)|_{\eta,\ell}\le R$ for every $a\in\varkappa^n$.

Corollary A.4. (a) *If* $\psi\colon\Delta\to[0,\infty)$ *is supported on* Δ_0 *and* $|\psi|_{\eta,\ell}\le R$, *then* $\psi\in\mathcal{R}$.
(b) *If* $\psi\in\mathcal{R}$, *then* $L\psi\in\mathcal{R}$.
(c) *If* $\psi\in\mathcal{R}^n$, *then* $\psi\in\mathcal{R}^k$ *for all* $k\ge n$.
(d) *If* $\psi,\psi'\in\mathcal{R}^n$ *and* $t\ge 0$, *then* $\psi+\psi'$ *and* $t\psi$ *belong in* $\mathcal{R}^n$.

Proof. See [15, Corollary 4.2]. □

Lemma A.5. *There exist* $N \ge 1$ *and* $\varepsilon > 0$ *such that*

(a) $\int_{\Delta_0} \psi \, dm_\Delta \ge \varepsilon \int_\Delta \psi \, dm_\Delta$ *for all* $\psi \in L^N \mathcal{R}$,
(b) $(1 - \xi\varepsilon)\big(\frac{1-\xi}{1-\xi\varepsilon}\big)^n \ge e^R \bar\tau m_\Delta\big(\bigcup_{\ell=Nn}^\infty \Delta_\ell\big)$ *for all* $n \ge 1$.

Proof. (a) is proved in [15, Lemma 4.5]. Following the proof, we are free to choose ε as small as needed and N as large as needed. By assumptions of Lemma 4.3, $m_\Delta\big(\bigcup_{\ell=n}^\infty \Delta_\ell\big)$ decays exponentially in n, thus we can choose N and ε so that (b) is satisfied. □

Further we assume that N and ε are as in Lemma A.5. Define $\mathcal{B} = L^N \mathcal{R}$. Note that $L\mathcal{B} \subset \mathcal{B} \subset \mathcal{R}$. For $n \ge 0$ let $\mathcal{B}^n$ denote the set of observables $\psi \colon \Delta \to [0, \infty)$ such that $L^n \psi \in \mathcal{B}$ and $|L^n(\psi 1_a)|_{\eta,\ell} \le R$ for every $a \in \varkappa^n$.

Remark A.6. If $\psi \in \mathcal{B}$, then $L\psi \in \mathcal{B}$. If $\psi \in \mathcal{B}^n$, then $\psi \in \mathcal{B}^k$ for all $k \ge n$.

Define a sequence p_n, $n \ge -1$ by

$$p_{-1} = \xi\varepsilon \quad \text{and} \quad p_n = \begin{cases} (1-\xi)\varepsilon\big(\frac{1-\varepsilon}{1-\xi\varepsilon}\big)^{n/N}, & n \in N\mathbb{Z} \\ 0, & n \notin N\mathbb{Z} \end{cases} \quad \text{for } n \ge 0.$$

Let $t_n = 1 - \sum_{k=-1}^{n-1} p_k$ for $n \ge 1$. Then $\sum_{k=-1}^\infty p_k = 1$, $t_1 = 1 - \varepsilon$ and for $n \ge 2$ using Lemma A.5 we obtain

$$t_{Nn} = 1 - \sum_{k=-1}^{Nn-1} p_k = (1 - \xi\varepsilon)\Big(\frac{1-\varepsilon}{1-\xi\varepsilon}\Big)^n \ge \min\big\{t_1,\, e^R \bar\tau m_\Delta\big(\textstyle\bigcup_{\ell=Nn}^\infty \Delta_\ell\big)\big\}. \quad (5)$$

Let $E_0 = \Delta_0$ and $E_k = \{(y, \ell) \in \Delta : \ell = \tau(y) - k,\ \ell \ge 1\}$ for $k \ge 1$. Then $\{E_0, E_1, \ldots\}$ defines a partition of Δ and $m_\Delta(E_k) = m_\Delta(\Delta_k)$ for all k.

Proposition A.7. *If* $\psi \in \mathcal{B}$ *with* $\int_\Delta \psi \, dm_\Delta = 1$, *then* $\int_{\bigcup_{\ell=n}^\infty E_\ell} \psi \, dm_\Delta \le t_n$, *for* $n \ge 1$.

Proof. See [15, Proposition 4.6]. □

Proposition A.8. *Let* $p_j, q_j \in [0, \infty)$ *be sequences such that* $\sum_{j=0}^\infty p_j = \sum_{j=0}^\infty q_j < \infty$ *and* $\sum_{j=0}^k q_j \ge \sum_{j=0}^k p_j$ *for all* $k \ge 0$. *Then there exist* $s_{k,j} \in [0, 1]$, $0 \le j \le k$, *such that* $\sum_{j=0}^k s_{k,j} q_j = p_k$ *for all* $k \ge 0$ *and* $\sum_{k=j}^\infty s_{k,j} = 1$ *for all* $j \ge 0$.

Proof. See [15, Proposition 4.7]. □

Lemma A.9. *Let* $\psi \in \mathcal{B}^n$ *for some* $n \ge 0$. *Then* $\psi = \sum_{k=-1}^\infty \psi_k$, *where* $\psi_k : \Delta \to [0, \infty)$ *are such that*

(a) $L^n(\psi_{-1} 1_a) = c_a 1_{\Delta_0}$ *for all* $a \in \varkappa^n$, *where* c_a *are nonnegative constants,*
(b) $\sum_{a \in \varkappa^n} c_a = p_{-1} \bar\tau \int_\Delta \psi \, dm_\Delta$,
(c) $\psi_k \in \mathcal{R}^{n+k}$ *for all* $k \ge 0$,
(d) $\int_\Delta \psi_k \, dm_\Delta = p_k \int_\Delta \psi \, dm_\Delta$ *for all* $k \ge -1$.

Proof. We follow the proof of [15, Lemma 4.8]. Suppose without loss of generality that $\int_\Delta \psi \, dm_\Delta = 1$.

Define

$$t = p_{-1}/\int_{\Delta_0} L^n \psi \, dm_\Delta = \xi\varepsilon/\int_{\Delta_0} L^n \psi \, dm_\Delta .$$

By Lemma A.5, $\int_{\Delta_0} L^n \psi \, dm_\Delta \geq \varepsilon$, so $t \in [0, \xi]$.

Under convention that $0/0 = 0$, let

$$\psi_{-1} = t\bar{\tau} \sum_{a \in \varkappa^n} \left(\frac{\int_{\Delta_0} L^n(\psi 1_a) \, dm_\Delta}{L^n(\psi 1_a)} \circ f^n \right) \psi 1_a .$$

Then properties (a) and (b) are satisfied.

Let $g = \psi - \psi_{-1}$ and $g_k = g 1_{T^{-n} E_k}$ for $k \geq 0$. Then $L^{n+k} g_k$ is supported on Δ_0 and $|L^{n+k}(g_k 1_a)|_{\eta,\ell} \leq R$ for every $a \in \varkappa^n$. By Corollary A.4, $g_k \in \mathcal{R}^{n+k}$.

Let $q_k = \int_\Delta g_k \, dm_\Delta$. Then $\sum_{k=0}^\infty q_k = \sum_{k=0}^\infty p_k$ and by Proposition A.7, $\sum_{k=0}^n q_k \geq \sum_{k=0}^n p_k$ for all $n \geq 0$. Choose $s_{k,j} \in [0, 1]$ as in Proposition A.8, and define $\psi_k : \Delta \to [0, \infty)$, $k \geq 0$, by

$$\psi_k = \sum_{j=0}^k s_{k,j} g_j .$$

Then (d) holds for all k. Corollary A.4 implies (c). □

Let $\mathbb{W}$ be the countable set of all finite words in the alphabet $\mathbb{N}_0$ including the zero length word, and let $\mathbb{W}_k$ be the subset consisting of words of length k. Let $\mathbb{P}_\mathbb{W}$ be the probability measure on $\mathbb{W}$ given for $w = w_1 \cdots w_k \in \mathbb{W}_k$ by $\mathbb{P}_\mathbb{W}(w) = p_{-1} p_{w_1} \cdots p_{w_k}$. Define $r \colon \mathbb{W} \to \mathbb{N}_0$ by $r(w) = \Sigma w + N|w|$, where $\Sigma w = w_1 + \cdots + w_k$ and $|w| = k$ for $w = w_1 \cdots w_k$.

Proposition A.10. *Let* $\psi \in \mathcal{B}$ *with* $\int_\Delta \psi \, dm_\Delta = 1$. *Then* $\psi = \sum_{w \in \mathbb{W}} \psi_w$, *where* $\psi_w : \Delta \to [0, \infty)$ *are such that*

(a) $\int_\Delta \psi_w \, dm_\Delta = \mathbb{P}_\mathbb{W}(w)$,
(b) $L^{r(w)} \psi_w = \mathbb{P}_\mathbb{W}(w) \bar{\tau} 1_{\Delta_0}$,
(c) $L^{r(w)}(\psi_w 1_a) = c_{w,a} 1_{\Delta_0}$ *for all* $a \in \varkappa^{r(w)}$, *where* $c_{w,a}$ *are nonnegative constants.*

Proof. Proof is identical to [15, Proposition 4.9] except for condition (c), which is guaranteed by Lemma A.9. □

Definition A.11. We say that a random variable X has geometric distribution with parameter $\theta \in (0, 1)$ (or $X \sim \text{Geom}(\theta)$), if X takes values in $\mathbb{N}_0$ and $\mathbb{P}(X = n) = (1 - \theta)^n \theta$ for $n \geq 0$.

Proposition A.12. *Suppose that* $Y = \sum_{k=1}^M (1 + X_k)$, *where* $M \sim \text{Geom}(\theta_M)$ *and* $X_k \sim \text{Geom}(\theta_X)$ *are independent random variables. Let* $\eta_2 = \theta_X \theta_M$ *and* $\eta_1 = \frac{\theta_M - \eta_2}{1 - \eta_2}$. *Then*

$$\mathbb{P}(Y = n) = \begin{cases} \eta_1 + (1 - \eta_1)\eta_2, & n = 0 \\ (1 - \eta_1)\eta_2(1 - \eta_2)^n, & n \geq 1 \end{cases} .$$

Proof. We compute the probability generating function of Y. For $z \in \mathbb{R}$,

$$\mathbb{E}(z^Y) = \mathbb{P}(M = 0) + \mathbb{P}(M \geq 1)\, \mathbb{E}(z^{1+X_1})\, \mathbb{E}(z^Y).$$

Using

$$\mathbb{E}(z^{1+X_1}) = \textstyle\sum_{k=0}^{\infty} \mathbb{P}(X_1 = k) z^{k+1} = \frac{\theta_X z}{1-(1-\theta_X)z},$$

we obtain

$$\mathbb{E}(z^Y) = \eta_1 + (1 - \eta_1) \frac{\eta_2}{1-(1-\eta_2)z}.$$

Now, $\mathbb{P}(Y = n)$ is the coefficient at z^n in the above expression. □

Proposition A.13. *There exist constants* $0 < \theta < 1$ *and* $C_1, C_2 > 0$ *such that*

$$\mathbb{P}(r = n) = \begin{cases} C_1, & n = 0 \\ C_2 \theta^{n/N}, & n \in N\mathbb{N} \\ 0, & else. \end{cases}$$

Proof. Recall that $\mathbb{W}_k$ is the subset of $\mathbb{W}$ consisting of words of length k. Then $\mathbb{P}_{\mathbb{W}}(\mathbb{W}_k) = (1 - p_{-1})^k p_{-1}$. Elements of $\mathbb{W}_k$ have the form $w_1 \cdots w_k$ where $w_1, \ldots, w_k$ can be regarded as independent identically distributed random variables, drawn from $\mathbb{N}_0$ with distribution

$$\mathbb{P}(w_1 = n) = p_n/(1 - p_{-1}) = \begin{cases} \theta_1 (1 - \theta_1)^{n/N}, & n \in N\mathbb{N}_0 \\ 0, & \text{else} \end{cases},$$

where $\theta_1 = \frac{(1-\xi)\varepsilon}{1-\xi\varepsilon}$. In other words, $w_1/N \sim \text{Geom}(\theta_1)$.

Then the random variable r/N on $\mathbb{W}$ has the same distribution as Y in Proposition A.12 with $\theta_M = p_{-1}$ and $\theta_X = \theta_1$. The result follows. □

We are ready to complete the proof of Lemma 4.3. Let $\psi = dm/dm_\Delta = \bar{\tau} 1_{\Delta_0}$ and $\psi = \sum_{w \in \mathbb{W}} \psi_w$ be the decomposition from Proposition A.10.

Then $\psi = \sum_{w \in \mathbb{W}} \sum_{a \in A(w)} \psi_w 1_a$, where $A(w) = \{a \in \varkappa^{r(w)} : a \subset \Delta_0 \text{ and } f^{r(w)} a = \Delta_0\}$. To every $w \in \mathbb{W}$ and $a \in A(w)$ there corresponds $u \in \mathcal{A}$ such that $a = Y_u$ (modulo zero m measure) and $r(w) = h_{\mathcal{A}}(u)$. Thus we can write

$$m = \sum_{u \in \mathcal{A}} \mathbb{P}_{\mathcal{A}}(u) m_u,$$

where m_u are probability measures supported on Y_u and $\mathbb{P}_{\mathcal{A}}$ is a probability measure on $\mathcal{A}$ such that $\mathbb{P}_{\mathcal{A}}(h_{\mathcal{A}} = n) = \mathbb{P}_{\mathbb{W}}(r = n)$ for all n.

The result of Lemma 4.3 follows from Proposition A.13.

References

1. Berkes, I., Liu, W., Wu, W.: Komlós–Major–Tusnády approximation under dependence. Ann. Probab. **42**, 794–817 (2014)
2. Berkes, I., Philipp, W.: Approximation theorems for independent and weakly dependent random vectors. Ann. Probab. **7**, 29–54 (1979)
3. Bruin, H., Luzzatto, S., Strien, S. van: Decay of correlations in one-dimensional dynamics. Ann. Sci. École Norm. Sup. **36**, 621–646 (2003)
4. Chernov, N., Haskell, C.: Nonuniformly hyperbolic K-systems are Bernoulli. Ergod. Theory Dyn. Syst. **16**, 19–44 (1996)
5. Chernov, N., Markarian, R.: Chaotic Billiards. Mathematical Surveys and Monographs, vol. 127. American Mathematical Society, Providence (2006)
6. Conze, J.-P., Le Borgne, S.: Méthode de martingales et flow géodésique sur une surface de courbure constante négative. Ergod. Theory Dyn. Syst. **21**, 421–441 (2001)
7. Cuny C., Merlevède, F.: Strong invariance principles with rate for "reverse" martingales and applications. J. Theor. Probab. **28**, 137–183 (2015)
8. Denker, M., Philipp, W.: Approximation by Brownian motion for Gibbs measures and flows under a function. Ergod. Theory Dyn. Syst. **4**, 541–552 (1984)
9. Field, M.J., Melbourne, I., Török, A.: Decay of correlations, central limit theorems and approximation by Brownian motion for compact Lie group extensions. Ergod. Theory Dyn. Syst. **23**, 87–110 (2003)
10. Gouëzel, S.: Almost sure invariance principle for dynamical systems by spectral methods. Ann. Probab. **38**, 1639–1671 (2010)
11. Hofbauer, F., Keller, G.: Ergodic properties of invariant measures for piecewise monotonic transformations. Math. Z. **180**, 119–140 (1982)
12. Kiefer, J.: On the deviations in the Skorokhod-Strassen approximation scheme. Z. Wahrscheinlichkeitstheor. verw. Geb. **32**, 111–131 (1975); **13**, 321–332 (1969)
13. Komlós, J., Major, P., Tusnády, G.: An approximation of partial sums of independent RV'-s and the sample DF. I; II, Z. Wahrscheinlichkeitstheor. verw. Geb. **32**, 111–131 (1975); **34**, 34–58 (1976)
14. Korepanov, A.: Equidistribution for nonuniformly expanding dynamical systems, and application to the almost sure invariance principle. Commun. Math. Phys. **359**, 1123–1138 (2018)
15. Korepanov, A., Kosloff, Z., Melbourne, I.: Explicit coupling argument for nonuniformly hyperbolic transformations. Proc. R. Soc. Edinb. Sect. A. (to appear)
16. Korepanov, A., Kosloff, Z., Melbourne, I.: Martingale-coboundary decomposition for families of dynamical systems. Ann. Inst. H. Poincaré Anal. Non Linéaire (to appear) (2016)
17. Kuelbs, J., Philipp, W.: Almost sure invariance principles for partial sums of mixing B-valued random variables. Ann. Probab. **8**, 1003–1036 (1980)
18. Melbourne, I., Nicol, M.: Almost sure invariance principle for nonuniformly hyperbolic systems. Commun. Math. Phys. **260**, 131–146 (2005)
19. Melbourne, I., Nicol, M.: A vector-valued almost sure invariance principle for hyperbolic dynamical systems. Ann. Probab. **37**, 478–505 (2009)
20. Merlevède, F., Rio, E.: Strong approximation of partial sums under dependence conditions with application to dynamical systems. Stoch. Process. Appl. **122**, 386–417 (2012)
21. Merlevède, F., Rio, E.: Strong approximation for additive functionals of geometrically ergodic Markov chains. Electron. J. Probab. **20**, 1–27 (2015)
22. Philipp, W., Stout, W.: Almost Sure Invariance Principles for Partial Sums of Weakly Dependent Random Variables. American Mathematical Society Memoirs, vol. 161, American Mathematical Society, Providence (1975)
23. Sakhanenko, S.I.: Estimates in the invariance principle in terms of truncated power moments. Sib. Math. J. **47**, 1113
24. Strassen, V.: An invariance principle for the law of iterated logarithm. Z. Wahrscheinlichkeitstheor. Verw. Geb. **3**, 211–226 (1964)
25. Strassen, V.: Almost sure behavior of sums of independent random variables and martingales. Proc. Fifth Berkeley Symp. Math. Statist. Prob. **2**, 315–343 (1967)
26. Young, L.-S.: Statistical properties of dynamical systems with some hyperbolicity. Ann. Math. **147**, 585–650 (1998)
27. Young, L.-S.: Recurrence times and rates of mixing. Israel J. Math. **110**, 153–188 (1999)
28. Zaitsev, A.: The accuracy of strong Gaussian approximation for sums of independent random vectors. Russ. Math. Surv. **68**, 721 (2013)

29. Zweimüller, R.: Measure preserving transformations similar to Markov shifts. Israel J. Math. **173**, 421–443 (2009)

Communicated by C. Liverani