CONVERGENCE OF MOMENTS FOR DISPERSING BILLIARDS WITH CUSPS

P. BÁLINT, N. CHERNOV, D. DOLGOPYAT

ABSTRACT. Dispersing billiards with cusps are deterministic dynamical systems with a mild degree of chaos, exhibiting "intermittent" behavior that alternates between regular and chaotic patterns. They are characterized by decay of correlations of order $1/n$ and a central limit theorem with a non-classical scaling factor of $\sqrt{n \log n}$. As for the growth of the pth moments of the appropriately normalized Birkhoff sums, it follows from the results of [28] that these converge to the moments of the limit normal distribution only for $p < 2$ and diverge for $p > 2$. Here we focus on the critical case $p = 2$ and prove a doubling effect: the second moments converge, but their limit is twice the second moment of the limit normal distribution.

Keywords: Chaos, decay of correlations, Central Limit Theorem, billiards, dispersing billiards, cusps.

1. INTRODUCTION

Limit theorems (and the related issue of convergence of moments) play an important role in the studies of dynamical systems.

By a dynamical system we mean a transformation $F\colon M \to M$ of a measure space M with an invariant probability measure μ. Let $A\colon M \to \mathbb{R}$ be a function (observable). Then the sequence of observed values $A(F^n(X))$, where $X \in M$, makes a stationary process with respect to the invariant measure μ. The main object of studies is the behavior of its partial sums

$$S_n A\colon = A + A \circ F + \cdots + A \circ F^{n-1}. \tag{1.1}$$

If μ is ergodic and $A \in L^1_\mu(M)$, then $S_n A = n\mu(A) + o(n)$ for a.e. $X \in M$, according to the Birkhoff ergodic theorem; we use standard notation $\mu(A) = \int_M A\, d\mu$. It is common to consider centered sums $S_n A - n\mu(A) = S_n(A - \mu(A))$, so we will always assume that $\mu(A) = 0$; otherwise we just replace A with $A - \mu(A)$. Now we have $S_n A = o(n)$.

Limit theorems describe asymptotic distribution of $(S_n A)/b_n$, where $b_n > 0$ is an appropriate scaling factor. The latter is selected so that $(S_n A)/b_n = \mathcal{O}(1)$ for typical points $X \in M$. Then a limit theorem

usually states that $(S_nA)/b_n$ converges in distribution, i.e.,

$$\lim_{n\to\infty} \mu\big\{X\colon (S_nA)/b_n < x\big\} = G(x) \tag{1.2}$$

where $G(x)$ is a probability distribution function, and (1.2) holds for every x at which $G(x)$ is continuous. If $b_n = \sqrt{n}$ and $G(x)$ is a normal (Gaussian) distribution function, then we refer to (1.2) as a classical Central Limit Theorem (CLT).

While (1.2) describes the limit distribution of $(S_nA)/b_n$ it is also important to describe the asymptotics of its moments. We will say that the pth (absolute) moment of $(S_nA)/b_n$ *properly converges* if

$$\lim_{n\to\infty} \mu\big(|(S_nA)/b_n|^p\big) = \int |x|^p\, dG(x). \tag{1.3}$$

The pth moment of $(S_nA)/b_n$ may also converge to a value different from the right hand side of (1.3) or diverge altogether. The following standard fact (e.g., [24, Exercise 3.2.5, p. 87]) helps to clarify the picture:

Theorem 1. *Suppose* (1.2) *holds and* $\sup_n \mu\big(|(S_nA)/b_n|^p\big) < \infty$. *Then the qth moment of $(S_nA)/b_n$ properly converges for every $q < p$.*

Corollary 1.1. *There is a* critical moment $p_* \in [0,\infty]$ *such that*
(a) *the qth moment of $(S_nA)/b_n$ properly converges for all $q < p_*$*
(b) *the qth moment of $(S_nA)/b_n$ diverges for all $q > p_*$.*

In case $p_* = \infty$ we have proper convergence of all moments. In case $p_* = 0$ we have divergence of all moments. The p_*th moment itself may converge (properly or improperly) or diverge. We note, however, that convergence in distribution implies that

$$\liminf_{n\to\infty} \mu\big(|(S_nA)/b_n|^p\big) \ge \int |x|^p\, dG(x)$$

therefore the limit of the p_* moment can only be greater than the p_*th moment of the limit distribution.

While the limit theorems (mostly, versions of the classical CLT) have been proven for many types of dynamical systems, the convergence of moments has received less attention until recently. On the other hand, physicists prefer to use moments, see e.g., [19], because those can be easily estimated in numerical experiments.

Recently, in part due to connections to large deviation estimates ([32]) and concentration inequalities ([10]), there has been substantial progress related to the convergence of moments in the mathematical literature ([34], [28], [21]). Of special significance is the paper [28],

which, following up on [34], investigates the problem for the important class of dynamical systems modeled by a Young tower.

Theorem 2 ([34], [28])**.** *Let an ergodic system (M, F, μ) be modeled by a Young tower and A be a Hölder continuous function on M. Then*
(a) *If the tower has an exponential tail bound, then we have proper convergence of all moments, i.e., $p_* = \infty$.*
(b) *If the tower has polynomial tails of order $\beta > 0$, then we have proper convergence of moments of order $p < 2\beta$.*
(c) *Assume again that the tower has polynomial tails of order $\beta > 0$. Then there is a nonempty subset $\mathcal{U}$ in the space of Hölder functions on the tower, such that for $A \in \mathcal{U}$ and $p > 2\beta$ the moments of order p diverge. Hence, for $A \in \mathcal{U}$, we have $p_* = 2\beta$.*

Remark 1.2. *The class $\mathcal{U}$ is reasonably large, and for many important situations can be described rather explicitly. For details we refer to the literature, eg.* [28, 27, 3, 2]*; see also Remark 1.3 below.*

The tail of a Young tower refers to the rates of convergence of the respective return times. Precisely, let Δ denote the tower, μ_Δ the invariant measure on it, $\Delta_0 \subset \Delta$ the base of the tower, and $R : \Delta \to \mathbb{N}^+$ (defined on all Δ) the first return time to Δ_0. Then the rate of convergence of $\mu_\Delta(R > n)$ to zero, as $n \to \infty$, is the key characteristic of the tower.

We say that the tower has exponential tail bound if $\mu_\Delta(R > n) = \mathcal{O}(\lambda^n)$ for some $\lambda < 1$. This implies exponential decay of correlations, i.e., $\mu\big((A \circ F^n)A\big) = \mathcal{O}(\lambda_A^n)$ for some $\lambda_A < 1$ ([39]). All Axiom A diffeomorphisms, uniformly expanding interval maps, Hénon-like attractors [5], dispersing billiards (without cusps), etc., belong to this category.

We say that the tower has polynomial tails of order $\beta > 0$ if $\mu_\Delta(R > n) \sim Cn^{-\beta}$ for some $C > 0$. This implies polynomial decay of correlations of order β, i.e., $\mu\big((A \circ F^n)A\big) = \mathcal{O}(n^{-\beta})$. If $\beta > 1$, then the correlations are summable, i.e., $\sum_n |\mu\big((A \circ F^n)A\big)| < \infty$, and the classical CLT holds [40]. Many so called intermittent systems [35] belong to this category, including interval maps with neutral fixed point [29], Bunimovich flowers [7, 16], dispersing billiards with vanishing curvature [17], etc. On the other hand, nonstandard limit theorems hold if $\beta \le 1$ ([27]).

In fact, the authors of [28] do not only prove divergence in case (c) of Theorem 2, they also provide bounds on the growth of moments $\mu(|S_n A|^p)$, which are in accordance with observations in the physics literature, in particular [1]. For instance if $\beta = 1$ (which is the case of special interest to us, see below), $\mu(|S_n A|^p) \asymp n^{p-1}$ for $p > 2$ as opposed

to $\mu(|S_nA|^p) \asymp (n\log n)^{p/2}$ for $p<2$. The notation $f \asymp g$ means that $f=\mathcal{O}(g)$ and $g=\mathcal{O}(f)$. On the other hand what precisely happens at the critical moment $p_*(=2\beta)$ (i.e., whether we have a proper or improper convergence or divergence) remains unclear; perhaps it is system-dependent.

We study here a system modeled by a Young tower with polynomial tail bound of order $\beta=1$. In this case the correlations decay as n^{-1}, i.e.,

$$\zeta_n(A) \colon = \mu\big(A\cdot(A\circ F^n)\big) = \mathcal{O}(1/|n|) \tag{1.4}$$

The rate of growth of the second moment is then

$$\mu\big([S_nA]^2\big) = \sum_{k=-n+1}^{n-1} (n-|k|)\zeta_k(A) = \mathcal{O}(n\log n), \tag{1.5}$$

so the proper normalization factor for the limit theorem (1.2) must be $b_n=\sqrt{n\log n}$, rather than the classical $b_n=\sqrt{n}$.

The corresponding non-classical limit theorems were proved for several systems of that type, most notably for Bunimovich stadium [3] and for dispersing billiards with cusps [2].

Theorem 3 ([2, 3])**.** *Let $F\colon M\to M$ be the collision map for a Bunimovich stadium or for a planar dispersing billiard table with cusps. Let A be a Hölder continuous function on the collision space M with $\mu(A)=0$. Then we have a nonclassical limit theorem*

$$\lim_{n\to\infty} \mu\big\{X\colon (S_nA)/\sqrt{n\log n} < x\big\} = G(x) \tag{1.6}$$

where $G(x)$ is the distribution function of a normal law with mean zero and variance $\sigma_A^2\ge 0$.

Explicit formulas for the variance σ_A^2 exist for both the stadium (see [3]) and billiards with cusps (see [2] and (2.2) below). It may happen that $\sigma_A^2=0$, and in that case a version of the classical CLT applies, i.e., $(S_nA)/\sqrt{n}$ converges to a normal law [3, 2], but this is a degenerate case not covered by our present work.

We deal here with dispersing billiards with cusps, continuing our work [2]. Our main result is

Theorem 4. *Let $F\colon M\to M$ be the collision map for a planar dispersing billiard table with cusps. Let A be a Hölder continuous function on the collision space M with $\mu(A)=0$. Then*

$$\lim_{n\to\infty} \frac{\mu\big([S_nA]^2\big)}{n\log n} = 2\sigma_A^2. \tag{1.7}$$

More precisely,

$$\mu\big([S_n A]^2\big) = 2\sigma_A^2 n \log n + \mathcal{O}(n). \tag{1.8}$$

Note that the limit of the second moment is *not* equal to the second moment of the limit distribution, the former is *twice* the latter. This doubling effect will be explained in the end of Section 3.

Remark 1.3. *Let $\sigma_A^2 > 0$. Then the critical moment is $p_* = 2$, for which we have improper convergence. In particular, if $\sigma_A^2 > 0$, $A \in \mathcal{U}$ (cf. Remark 1.2): all moments of order $q < 2$ properly converge, while all moments of order $q > 2$ diverge.*

Remark. The limit law (1.6) holds true if we replace μ with any measure μ' that is absolutely continuous with respect to μ; see [2]. Similarly, our limit (1.7) remains valid if we replace μ with any $\mu' \ll \mu$, because the images $F^n\mu'$ weakly converge to μ.

We do not handle the Bunimovich stadium here. Despite its similarity to our billiards with cusps (in terms of the same rates of the decay of correlations and the same scaling factor in the non-classical limit theorem), the mechanism of nonuniform hyperbolicity is very different (see [18]), so our arguments will not apply to the stadium.

Finally it is worth mentioning another important model, the infinite horizon Lorentz gas, where the billiard flow is characterized by slow (polynomial) mixing rates [31]. The position of the moving particle $\mathbf{q}(t)$ at time t satisfies a non-classical limit theorem with a $\sqrt{t \log t}$ scaling factor (this is proved in [13]; see also [38]). Regarding the convergence of the second moment, a doubling effect analogous to our Theorem 4 has been observed and studied in [22]; for further discussion and numerical evidence see also [19, 23, 20]. However, this model is quite different from dispersing billiards with cusps and requires a different approach. Hence we plan to address the issue of the second moment in the infinite horizon Lorentz gas in a separate paper.

2. Billiards with cusps

Billiards are dynamical systems where a point particle moves in a planar domain $\mathcal{D}$ (the billiard table) and bounces off its boundary $\partial\mathcal{D}$ according to the classical rule "the angle of incidence is equal to the angle of reflection". The boundary $\partial\mathcal{D}$ is assumed to be a finite union of C^3 smooth compact curves that may have common endpoints.

Between collisions at $\partial\mathcal{D}$, the particle moves with a unit speed and its velocity vector remains constant. At every collision, the velocity

vector changes by

$$\mathbf{v}^+ = \mathbf{v}^- - 2\langle \mathbf{v}^-, \mathbf{n}\rangle \mathbf{n} \tag{2.1}$$

where $\mathbf{v}^-$ and $\mathbf{v}^+$ denote the velocities before and after collision, respectively, $\mathbf{n}$ stands for the inward unit normal vector to $\partial\mathcal{D}$, and $\langle \cdot, \cdot \rangle$ designates the scalar product.

If the boundary $\partial\mathcal{D}$ is concave inward and the curvature of $\partial\mathcal{D}$ does not vanish, the billiard is said to be *dispersing*. Such billiards were studied by Sinai [37] and Bunimovich [6] under the assumptions that the boundary components are smooth closed curves.

Sinai proved that the resulting billiard dynamics is strongly (uniformly) hyperbolic, ergodic, and K-mixing. Gallavotti and Ornstein [26] proved that dispersing billiards are Bernoulli. Young [39] proved that correlations decay exponentially fast; see also [11] for an infinite horizon situation. The classical CLT was derived in [8, 9].

All these results have been extended to dispersing billiards with piecewise smooth boundaries, i.e., to tables with corners, provided the boundary components intersect each other transversally, i.e., the angles made by the walls at corner points are positive; see [9, 11].

We deal with dispersing billiards where some boundary components converge tangentially at a corner, i.e., make a *cusp*. Such billiards were first studied by Machta [30] who found (based on heuristic arguments) that correlations for the collision map decay as $1/n$. The reason for such a slow decay is weak (non-uniform) hyperbolicity of the collision map. Whenever the moving particle gets deep into a cusp, it experiences a large number of rapid collisions that do not contribute much to the expansion or contraction of tangent vectors.

Reháček [36] proved that dispersing billiards with cusps are ergodic, K-mixing and Bernoulli. The rates of the decay of correlations (as predicted by Machta) were rigorously derived in [15, 18], and the non-classical limit theorem was proved in [2].

We note that correlations decay slowly only in discrete time, when each collision counts as a unit of time. In real (continuous) time, collisions inside a cusp occur in rapid succession and their effect is much less pronounced. As a result, the corresponding billiard flow is *rapid mixing* in the sense that correlations for smooth observables decay faster than any polynomial rate, and a classical CLT holds [4].

Our interest in billiards with cusps is motivated by [12] where the authors consider an interaction of a heavy particle of positive size with a light point particle moving in a dispersive domain. The case when the heavy particle is near the boundary reduces to a study of billiards with almost cusp which in turn requires a good understanding of billiards

with cusps. We observe that in the model of [12] the particles exchange energy only at the moment of collisions, so the observable studied in [12] can not be represented as a integral along the flow orbit and the discrete time CLT of [2] is relevant.

Next we introduce some notation. Just as in the previous work [2], we assume for simplicity that the table $\mathcal{D}$ has exactly one cusp; the generalization to several cusps is straightforward.

There are natural coordinates r and φ in the collision space M, where r denotes the arc length parameter on $\partial\mathcal{D}$ and φ the angle of reflection, i.e., the angle between $\mathbf{v}^+$ and $\mathbf{n}$ in the notation of (2.1). Note that $-\pi/2 \le \varphi \le \pi/2$. The billiard map F preserves the measure μ on M given by $d\mu = c_\mu \cos\varphi \, dr, d\varphi$, where $c_\mu = [2\,\mathrm{length}(\partial\mathcal{D})]^{-1}$ is the normalizing factor. In these coordinates, M is a union of rectangles $[r_i', r_i''] \times [-\pi/2, \pi/2]$, where the intervals $[r_i', r_i'']$ correspond to smooth components (arcs) of $\partial\mathcal{D}$.

The cusp is a common terminal point of two arcs, i_i and i_2, of $\partial\mathcal{D}$; thus the coordinate r takes two values at the cusp, $r' = r'_{i_1}$ and $r'' = r''_{i_2}$. Now the variance σ_A^2 in Theorem 3 is given by

$$\sigma_A^2 = \frac{c_\mu}{8\bar{a}} \left[\int_{-\pi/2}^{\pi/2} [A(r', \varphi) + A(r'', \varphi)] \sqrt{\cos\varphi} \, d\varphi \right]^2 \tag{2.2}$$

where $\bar{a} = (a_1 + a_2)/2$ and a_1, a_2 denote the curvatures of the two arcs making the cusp measured at the vertex of the cusp.

If the table $\mathcal{D}$ has more than one cusp, then σ_A^2 is the sum of expressions (2.2) corresponding to individual cusps.

It is common in the studies of nonuniformly hyperbolic maps, like our $F\colon M \to M$, to reduce the dynamics onto a subset $\mathcal{M} \subset M$ so that the induced map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ will be strongly hyperbolic and have exponential decay of correlations.

In the present case the hyperbolicity is slow only because of the cusp. So we cut out a small vicinity of the cusp. That is, we remove from M two rectangles, $R_1 = [r'_{i_1}, r'_{i_1} + \varepsilon_0] \times [-\pi/2, \pi/2]$ and $R_2 = [r''_{i_2} - \varepsilon_0, r''_{i_2}] \times [-\pi/2, \pi/2]$, with some small $\varepsilon_0 > 0$ and consider the induced map $\mathcal{F}$ on the remaining collision space $\mathcal{M} = M \setminus (R_1 \cup R_2)$. It preserves the conditional measure ν on $\mathcal{M}$, where $\nu(B) = \mu(B)/\mu(\mathcal{M})$ for any $B \subset \mathcal{M}$. The map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ is strongly hyperbolic and has exponential decay of correlations [2, 15].

Now let $\mathcal{R}(x) = \min\{m \ge 1\colon F^m x \in \mathcal{M}\}$ denote the return time function on $\mathcal{M}$. The domains

$$\mathcal{M}_m = \{x \in \mathcal{M}\colon \mathcal{R}(x) = m\}$$

for $m \geq 1$ are called *cells*; note that $\mathcal{M} = \cup_{m\geq 1}\mathcal{M}_m$. It is known [15] that $\nu(\mathcal{M}_m) = \mathcal{O}(m^{-3})$.

For $m \geq 1$ and $i = 0, 1, \ldots, m-1$ we denote

$$M_{m,i} = F^i(\mathcal{M}_m) \qquad \text{and} \qquad M_m = \cup_{i=0}^{m-1} M_{m,i}.$$

Then the sets $\{M_{m,i}\}$ constitute a partition of M. See Figure 1 for an illustration. Note that

$$\mu(M_{m,i}) = \mu(\mathcal{M}_m) = \mathcal{O}(m^{-3}) \qquad \text{and} \qquad \mu(M_m) = \mathcal{O}(m^{-2}).$$

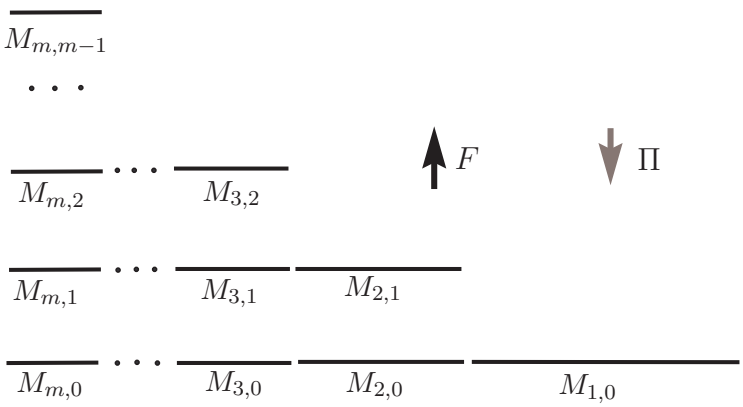

FIGURE 1. The structure of cells $M_{m,i}$ in the space M. The "bottom level" of M is the induced phase space $\mathcal{M} = \cup_{m=0}^{\infty}\mathcal{M}_m = \cup_{m=0}^{\infty} M_{m,0}$. The map F moves each cell one level up, but the top cell in each column is mapped back down to $\mathcal{M}$. The projection Π (Section 6) collapses each column $\cup_{i=0}^{m-1} M_{m,i}$ onto its bottom element $M_{m,0} = \mathcal{M}_m$.

For the given function A on M we construct the "induced" function on $\mathcal{M}$ as follows:

$$\mathcal{A}(x) = \sum_{m=0}^{\mathcal{R}(x)-1} A(F^m x). \tag{2.3}$$

Since we assume $\mu(A) = 0$, we also have $\nu(\mathcal{A}) = 0$. Note that $\mathcal{A}$ is of order m on $\mathcal{M}_m$, hence in generic situation $\nu(\mathcal{A}^2) = \infty$. In fact, $\nu(\mathcal{A}^2) < \infty$ if and only if $\sigma_A^2 = 0$, which is a degenerate case; cf. [2].

For any $p \geq 1$ we introduce the following notations:

- $\mathcal{M}_{1,p} = \cup_{m\leq p}\mathcal{M}_m$, and
- $\mathcal{A}_{1,p}$ the "truncated" version of $\mathcal{A}$, defined by $\mathcal{A}_{1,p}(X) = \mathcal{A}(X)$ for $X \in \mathcal{M}_{1,p}$ and $\mathcal{A}_{1,p}(X) = 0$ elsewhere.

The following estimate is proved in [2] by direct calculation:

Lemma 2.1 ([2]). *We have* $\mu(\mathcal{A}_{1,p}^2) = 2\sigma_A^2 \log p + \mathcal{O}(1)$.

The map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ is uniformly hyperbolic, i.e., it expands unstable curves and contracts stable curves at an exponential rate. More precisely, if u is an unstable tangent vector at any point $x \in \mathcal{M}$, then $\|D_x\mathcal{F}^n(u)\| \geq c\Lambda^n\|u\|$ for some constants $c > 0$ and $\Lambda > 1$ and all $n \geq 1$. Similarly, if v is a stable tangent vector, then $\|D_x\mathcal{F}^{-n}(v)\| \geq c\Lambda^n\|v\|$ for all $n \geq 1$.

The singularities of the original map $F\colon M \to M$ are made by trajectories hitting corner points (other than cusps) or experiencing grazing (tangential) collisions with $\partial\mathcal{D}$. The singularities of F lie on finitely many smooth compact curves. Those curves are stable in the sense that their tangent vectors belong to stable cones. Likewise, the singularities of F^{-1} are unstable curves.

The singularities of the induced map $\mathcal{F}$ are those of F plus the boundaries of the cells $\mathcal{M}_m$, $m \geq 1$. Those boundaries form a countable union of smooth compact stable curves that accumulate near the (unique) phase point whose trajectory runs directly into the cusp.

The structure of cells $\mathcal{M}_m$ and their boundaries are described in [15]. Each cell has length $\asymp m^{-7/3}$ in the unstable direction and length $\asymp m^{-2/3}$ in the stable direction. Its measure is

$$\mu(\mathcal{M}_m) \asymp m^{-7/3} \times m^{-2/3} = m^{-3}.$$

The map $\mathcal{F} = F^m$ expands the cell $\mathcal{M}_m$ in the unstable direction by a factor $\asymp m^{5/3}$ and contracts it in the stable direction by a factor $\asymp m^{5/3}$, too. So the image $\mathcal{F}(\mathcal{M}_m)$ has 'unstable size' $\asymp m^{-2/3}$ and 'stable size' $\asymp m^{-7/3}$. The images accumulate near the (unique) phase point whose trajectory emerges directly from the cusp.

A characteristic feature of hyperbolic dynamics with singularities is the competition between hyperbolicity and the cutting by singularities. The former causes expansion of unstable curves, it makes them longer. The latter breaks unstable curves into pieces and thus produces shorter curves. One of the main results of [15] is a so called one-step expansion estimate [15, Eq. (5.1)] for the induced map $\mathcal{F}$, which guarantees that the expansion prevails over the cutting by singularities, i.e., "on average" the unstable curves grow fast, at an exponential rate.

The one-step expansion estimate implies the entire spectrum of statistical facts: the growth lemmas, the coupling lemma for standard pairs and standard families, equidistribution estimates, exponential decay of correlations for bounded Hölder continuous functions, limit theorems for the same type of functions, etc. All these facts with detailed proofs are presented in [14, Chapter 7] for general dispersing

billiards (without cusps), but those proofs work for our map $\mathcal{F}$ almost verbatim (see [15, p. 749]).

The main tool in our analysis of the map $\mathcal{F}$ is standard pairs and standard families; see [14, Section 7.4] for the definition and basic properties. Given a standard family $\mathcal{G} = \{(W, \nu_W)\}$ of unstable curves $\{W\}$ with smooth probability measures $\{\nu_W\}$ on them, and a factor measure $\lambda_\mathcal{G}$ that defines a probability measure $\mu_\mathcal{G}$ on $\cup W$, its Z-function is defined by

$$Z(\mathcal{G}) \colon = \sup_{\varepsilon>0} \frac{\mu_\mathcal{G}(r_\mathcal{G} < \varepsilon)}{\varepsilon}$$

where $r_\mathcal{G}(x)$ denotes the distance from a point $x \in W \in \mathcal{G}$ to the nearer endpoint of W, i.e., $r_\mathcal{G}(x) = \operatorname{dist}(x, \partial W)$. If the curves $W \in \mathcal{G}$ have lengths of $\asymp L$, then $Z(\mathcal{G}) \asymp 1/L$ (see [14, p. 171]). The images $\mathcal{G}_n = \mathcal{F}^n(\mathcal{G})$ are also standard families, and their Z-function satisfies

$$Z(\mathcal{G}_n) \le c_1 \vartheta^n Z(\mathcal{G}) + c_2 \tag{2.4}$$

where $\vartheta \in (0,1)$ and $c_1, c_2 > 0$ are constants.

A standard family $\mathcal{G}$ is said to be proper if $Z(\mathcal{G}) \le C_\text{p}$ where C_p is a suitable large constant; see [14, p. 172]. The condition $Z(\mathcal{G}) \le C_\text{p}$ means that the family mostly consists of long unstable curves. If a family $\mathcal{G}$ consists of small curves (of length of order ε), then its Z-function is of order $1/\varepsilon$, and due to (2.4) it takes $C|\log\varepsilon|$ iterations of $\mathcal{F}$ (where $C > 0$ is a large constant) to transform $\mathcal{G}$ into a proper standard family (of mostly long curves).

3. Proof of Theorem 4

In this section we prove (1.8) modulo two technical lemmas that will be proved in subsequent sections. Expanding the square gives

$$\begin{aligned}\mu\big([S_n A]^2\big) &= n\mu(A^2) + 2\sum_{k=1}^{n-1}(n-k)\,\mu\big(A\cdot(A\circ F^k)\big)\\ &= n\sum_{k=-n}^{n}\mu\big(A\cdot(A\circ F^k)\big) + \mathcal{O}(n)\end{aligned}$$

in the last line we used the fact $\mu(A\cdot(A\circ F^k)) = \mathcal{O}(1/k)$ according to (1.4). Now (1.8) is equivalent to

$$\sum_{k=-n}^{n}\mu\big(A\cdot(A\circ F^k)\big) = 2\sigma_A^2\log n + \mathcal{O}(1) \tag{3.1}$$

or in a slightly different form

$$\mu\Big(A\cdot\Big(\sum_{k=-n}^{n}A\circ F^k\Big)\Big)=2\sigma_A^2\log n+\mathcal{O}(1). \tag{3.2}$$

Next we estimate the contribution to (3.1) from "high" cells M_m with $m\ge n/10$. We easily see that the contribution from points $X\in M$ such that either $X\in M_m$ or $F^n(X)\in M_m$ or $F^{-n}(X)\in M_m$ for some $m\ge n/10$ is bounded by

$$3\|A\|_\infty^2 n\sum_{m=n/10}^{\infty}\mu(M_m)=\mathcal{O}(1),$$

hence it can be neglected. Of course, 10 can be replaced here with any other fixed constant. So we assume that $X\in M_{m,i}$ for some $m<n/10$ and $0\le i\le m-1$.

Our next step is to express the formula (3.2) in terms of the induced map $\mathcal{F}\colon\mathcal{M}\to\mathcal{M}$.

Note that X either is in $\mathcal{M}$ (away from the cusp) or belongs to a series of collisions in the cusp that includes all the points $F^j(X)$ with $-i<j<m-i$. The corresponding values $A\circ F^j(X)$ for $-i<j<m-i$ make a part of the longer sequence of values $A\circ F^j(X)$ for $|j|\le n$ that appears in (3.2). We will see that it is this part that makes the crucial contribution to (3.2) and determines its asymptotic behavior. Indeed, if we only take into account the values $A\circ F^j(X)$ for $-i<j<m-i$, then after some obvious rearrangements

$$\begin{aligned}\sum_{m=1}^{n/10}\int_{\mathcal{M}_m}\Big(\sum_{i=0}^{m-1}A\circ F^i\Big)^2d\mu&=\sum_{m=1}^{n/10}\mu\big(\mathcal{A}^2|_{\mathcal{M}_m}\big)\\&=\mu(\mathcal{A}^2_{1,n/10})\\&=2\sigma_A^2\log n+\mathcal{O}(1)\end{aligned} \tag{3.3}$$

according to Lemma 2.1.

So it remains to show that the contribution from values $A\circ F^j(X)$ for $j\le -i$ and $j\ge m-i$ does not exceed $\mathcal{O}(1)$. In other words, the asymptotic of (3.2) is determined by series of collisions in the cusp that contains the present point X; the contribution from all future and past collisions is negligible, as we will show.

We note that the entire sequence of values $A\circ F^j(X)$ for $|j|\le n$ is naturally divided into subsequences corresponding to subsequent returns to $\mathcal{M}$. Between any two consecutive returns to $\mathcal{M}$ we have a series of collisions in the cusp. Of course, the lengths of these subsequences and their number depend on X. The very last subsequence

containing the point $F^n(X)$ and the very "first" one containing the point $F^{-n}(X)$ may be incomplete, as the corresponding series of collisions in the cusp may stretch beyond our time limits n and $-n$.

Our next step is to replace the sum in (3.2) involving the original map F with another sum in terms of the induced map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$. More precisely, we will replace (3.2) with

$$\mu\left(\mathcal{A}_{1,\frac{n}{10}} \cdot \left(\sum_{k=k^-(X)}^{k^+(X)} \mathcal{A}\circ\mathcal{F}^k\right)\right). \tag{3.4}$$

This function takes non-zero values only on $\mathcal{M}_{1,n/10}$. The variable summation limits $k^+(X)$ and $k^-(X)$ should be selected, roughly, so that

$$\sum_{k=0}^{k^+(X)} \mathcal{R}(\mathcal{F}^k X) \approx n, \qquad \sum_{k=k^-(X)}^{-1} \mathcal{R}(\mathcal{F}^k X) \approx n,$$

where $\mathcal{R}$ is the return function on $\mathcal{M}$, so that the new summation limits in (3.4) roughly correspond to the old ones in (3.2).

In order to construct the sum (3.4) we first replace (3.2) with

$$\mu\Big(A \cdot \Big(\sum_{k=-n^-(X)}^{n^+(X)} A\circ F^k\Big)\Big). \tag{3.5}$$

We define the function $n^+(X)$ as follows. Let $X \in M_{p,i}$ for some $p = p(X) < n/10$ and $i = i(X) \in [0, p-1]$. Let $F^{n-i}(X) \in M_{q',j'}$ for some $q' = q'(X) \geq 1$ and $j' = j'(X) \in [0, q'-1]$. Then we define

$$n^+(X) = n - i + q' - j' - 1. \tag{3.6}$$

The idea is that X is in the series of points $F^{-i}(X), \dots, F^{p-i-1}(X)$ (corresponding to a series of collisions in the cusp) that lie in the column of cells $M_{p,0}, \dots, M_{p,p-1}$ (which are the images of the cell $\mathcal{M}_p$; see Fig. 1). In this column the cell $M_{p,0} = \mathcal{M}_p$ plays the role of a "base", and the point $F^{-i}(X)$ plays the role of a "base point" (it belongs to $\mathcal{M}_p$). Its image under F^n (which is $F^{n-i}(X)$) falls into another column of cells $M_{q',0}, \dots, M_{q',q'-1}$, i.e., it is a part of the sequence of points $F^{n-i-j'}(X), \dots, F^{n-i+q'-j'-1}(X)$ lying in those cells. Then we want that whole sequence be included in (3.5) and nothing beyond it.

Now for all the points $F^{-i}(X), \dots, F^{p-i-1}(X)$ the summation in (3.5) will terminate at the end of the same column of cells. This will allow us to collect all the values of A in each column of cells and replace it with the corresponding value of the induced function $\mathcal{A}$. After that

(3.5) will be easily converted to (3.4). Precise formulas for $k^+(X)$ will be derived in Section 6, where we will need them.

The lower summation limit $n^-(X)$ can be defined similarly, we omit details.

In the case $n^+(X) > n$ we will need to add something to the sum in (3.2), and in the case $n^+(X) < n$ we will need to remove something from it, in order to convert it to the sum in (3.5). These adjustments are usually small (of order 1), but occasionally they may be large, up to order n.

In any case, these additions and removals will alter the value of (3.2), i.e., the values of (3.2) and (3.5) will differ, and we will need to estimate by how much. The following lemma will be proved in Section 5:

Lemma 3.1. *We have*

$$\mu\Big(A \cdot \Big(\sum_{k=n}^{n^+(X)} A \circ F^k\Big)\Big) = \mathcal{O}(1)$$

(if $n^+(X) < n$, then the index k runs from $n^+(X)$ to n). A similar estimate holds when k runs between $-n$ and $-n^-(X)$.

From now on we deal with (3.5), in fact with its equivalent version (3.4). The central term of (3.4), corresponding to $k = 0$, gives the entire desired asymptotic $2\sigma_A^2 \log n$, according to (3.3), so it remains to show that the rest of (3.4) is negligible, i.e.,

$$\Sigma_1^+ : = \mu\Big(\mathcal{A}_{1,\frac{n}{10}} \cdot \Big(\sum_{k=1}^{k^+(X)} \mathcal{A} \circ \mathcal{F}^k\Big)\Big) = \mathcal{O}(1), \tag{3.7}$$

and a similar estimate for the sum over $k = -1, \dots, k^-(X)$. Due to the time symmetry it is enough to prove (3.7).

It is known that the correlations for the map $\mathcal{F}$ and the function $\mathcal{A}$ and all its truncated versions decay exponentially fast:

Lemma 3.2 ([2])**.** *For each $k \geq 1$ and any $p \geq 1$ we have*

$$\big|\nu\big(\mathcal{A}_{1,p} \cdot (\mathcal{A} \circ \mathcal{F}^k)\big)\big| \leq C\theta^k \tag{3.8}$$

for some $C > 0$ and $\theta \in (0,1)$ that are determined by the function $\mathcal{A}$ but do not depend on p or k. (The condition $k \geq 1$ is crucial, cf. Lemma 2.1.)

Thus extending the summation in (3.7) to, say, $2n$, gives

$$\Sigma_2^+ : = \mu\Big(\mathcal{A}_{1,\frac{n}{10}} \cdot \Big(\sum_{k=1}^{2n} \mathcal{A} \circ \mathcal{F}^k\Big)\Big) = \sum_{k=1}^{2n} \mu\big(\mathcal{A}_{1,\frac{n}{10}} \cdot (\mathcal{A} \circ \mathcal{F}^k)\big) = \mathcal{O}(1). \tag{3.9}$$

We see that the "longer" sum Σ_2^+ is $\mathcal{O}(1)$, but this does not immediately imply that the "shorter" sum Σ_1^+ is of the same order. The limits $k^+(X)$ in (3.7) are point-dependent, and they may "conspire" to increase Σ_1^+ up to $\mathcal{O}(n)$. The following lemma will be proved in Section 6:

Lemma 3.3. *We have* $\Sigma_2^+ - \Sigma_1^+ = \mathcal{O}(1)$.

This implies that $\Sigma_1^+ = \mathcal{O}(1)$ and completes our proof of Theorem 4.

Lastly we explain the bizarre doubling effect mentioned in the Introduction, i.e., why the limit of the second moment of $(S_nA)/\sqrt{n\log n}$ is exactly twice the second moment of its limit distribution.

A closer look at (3.3) reveals that if we truncate the function $\mathcal{A}$ at the level n^b with $b \le 1$, instead of $n/10$, then we get

$$\sum_{m=1}^{n^b} \mu\big(\mathcal{A}^2|_{\mathcal{M}_m}\big) = \mu(\mathcal{A}^2_{1,n^b}) = 2b\sigma_A^2 \log n + \mathcal{O}(1).$$

Thus values of $\mathcal{A}$ in the range $[0, n^b]$ for $0 \le b \le 1$ account for a fraction of the total second moment proportional to b. However values of $\mathcal{A}$ larger than $n^{1/2}$ occur too rarely to affect the limit distribution of $(S_nA)/\sqrt{n\log n}$, as it was shown in [2]. Thus it is exactly half the range of relevant values of $\mathcal{A}$ that affect the limit distribution, while the entire range of values affect the second moment.

In the Appendix we describe a simple probabilistic model that exhibits a similar doubling effect in the second moment.

4. Young tower and equidistribution for F

The induced map $\mathcal{F}\colon \mathcal{M} \to \mathcal{M}$ has exponential mixing rates and can be modeled by a Young tower $\Delta_{\mathcal{F}}$ with exponential tail bounds [15]. The hyperbolic set $\Lambda_0 \subset \mathcal{M}$ used to define the base of that tower may be constructed pretty much anywhere in $\mathcal{M}$ and we can assume that $\Lambda_0 \subset \mathcal{M}_1$.

Now one can use the set Λ_0 to define the base for a Young tower Δ_F modeling the original map F; see again [15]. The tail bound on return times for the tower Δ_F will be polynomial of order $\beta = 1$, in terms of our Introduction.

Let us introduce some additional notation concerning the tower $\Delta = \Delta_F$. Unless otherwise stated, $T\colon \Delta \to \Delta$ will denote the tower map that models the original dynamics $F\colon M \to M$ (that is, the subscript F is typically omitted). Then $\Delta_0 (= \Lambda_0) \subset \Delta$ is the base and $\Delta_m \subset \Delta$ is the mth level of the tower. $\Delta_{m,m} \subset \Delta_m$ denotes the part of Δ_m that is mapped down to the base, i.e., $\Delta_{m,m} = \Delta_m \cap T^{-1}(\Delta_0)$. For $r > m$, let $\Delta_{m,r} = \Delta_m \cap T^{m-r}(\Delta_{r,r})$ be the part of Δ_m that goes up another

$r-m$ steps before coming down to Δ_0. We will call $\Delta^{(m)} = \cup_{i=0}^{m} \Delta_{i,m}$ the mth column of the tower. In our tower, $\mu_\Delta(\Delta_m) = \mathcal{O}(1/m^2)$, $\mu_\Delta(\Delta^{(m)}) = \mathcal{O}(1/m^2)$ and $\mu_\Delta(\Delta_{0,m}) = \mathcal{O}(1/m^3)$; see [18]. Altogether we obtain a picture similar to the one displayed on Figure 1, however, it is important to point out that these two towers describe the dynamics in two different ways. In particular, it is only the base Δ_0 of the Young tower that is in one-to-one correspondence with a subset of the phase space (notably with $\Lambda_0 \subset \mathcal{M}_1 \subset \mathcal{M} \subset M$); in general, there is a projection $\pi\colon \Delta \to M$ that semi-conjugates the tower map $T\colon \Delta \to \Delta$ with the billiard map $F\colon M \to M$. In our arguments (for example, in the proof of Lemma 4.1 below) it will play an important role where the π-preimages of different parts of the phase space M (that is, of the picture on Figure 1) appear on the Young tower Δ.

We use the tower to derive useful estimates on the equidisrtibution of the images of cells $M_{p,i}$ under the iterations of the original map F. Our first goal is to estimate the measure of the intersection $\mu\big(F^n(M_{p,i}) \cap M_{q,j}\big)$. We will always assume $p, q < n/10$.

Suppose for a moment that the images of our cells were completely independent (which is a highly idealized situation). Then we would get

$$\mu\big(F^n(M_{p,i}) \cap M_{q,j}\big) = \mu(M_{p,i})\,\mu(M_{q,j}) = \mathcal{O}(p^{-3}q^{-3})$$

and summing over j and q would get

$$\mu\big(F^n(M_{p,i}) \cap (\cup_{q>m} \cup_{j=0}^{q-1} M_{q,j})\big) = \mathcal{O}(p^{-3}m^{-1})$$

In fact, we will derive only a slightly weaker bound:

Lemma 4.1. *We have*

$$\mu\big(F^n(M_{p,i}) \cap (\cup_{q>m} \cup_{j=0}^{q-1} M_{q,j})\big) = \mathcal{O}(p^{-3}m^{-1}) + \chi_p \tag{4.1}$$

where $\chi_p = \mathcal{O}(p^{-3-a})$ *for some* $a > 0$.

Proof. We foliate the cell $M_{p,i}$ by unstable curves. Then the measure μ conditioned on $M_{p,i}$ becomes a standard family, $\mathcal{G}$. As the images of this family under the iterations of F moves between consecutive returns to $\mathcal{M}$ (i.e., during a series of collisions inside the cusp), the corresponding unstable curves grow slowly but they cannot be cut by singularities. At the time of the very first exit from the cusp, their lengths are of order $p^{-2/3}$, thus their Z-function will be $\asymp p^{2/3}$.

After exiting the cusp, our unstable curves may be cut into pieces by singularities, those pieces enter other cells $\mathcal{M}_k$ and they continue their motion under F. Our strategy will be based on the following two principles. First, pieces that enter cells $\mathcal{M}_k$ whose index k is "high" (see below) will be discarded, their union will have a negligibly small

measure. Second, pieces that remain in cells $\mathcal{M}_k$ with "low" indices k will return to $\mathcal{M}$ quite frequently and thus will grow sufficiently fast.

The first part of our strategy is based on a standard lemma:

Lemma 4.2 ([38, 18])**.** *There are constants $a_1, a_2 > 0$ such that for any large $B > 0$ and any $p > 0$ there is a subset $\mathcal{M}'_p \subset \mathcal{M}_p$ such that*

$$\mu(\mathcal{M}_p \setminus \mathcal{M}'_p) \le Cp^{-a_1}\mu(\mathcal{M}_p),$$

where $C = C(B) > 0$ is a constant, and for every $X \in \mathcal{M}'_p$ the images $\mathcal{F}^t(X)$ for $t = 1, \dots, B\log p$ never appear in cells $\mathcal{M}_r$ with $r > p^{1-a_2}$.

This lemma can be applied to $\mathcal{M}_p = F^{-i}(M_{p,i})$ and it gives a subset $M'_{p,i} \subset M_{p,i}$ of measure

$$\mu(M'_{p,i}) \ge \big(1 - \mathcal{O}(p^{-a_1})\big)\mu(M_{p,i})$$

such that the images of points $X \in M'_{p,i}$ will move through cells $M_{p',i'}$ for some $p' < p^{1-a_2}$ until they make $B\log p$ returns to $\mathcal{M}$. The set $M_{p,i} \setminus M'_{p,i}$ is then discarded and its measure is incorporated into $\chi_p = \mathcal{O}(p^{-3-a})$.

It is important to note that the returns of points $X \in M'_{p,i}$ to $\mathcal{M}$ are separated by series of iterations of F, and each of these series has length less than $p^{1-a_2} < n^{1-a_2}$. Hence, by the time when the images of points $X \in M_{p,i} \setminus M'_{p,i}$ make $B\log p$ returns to $\mathcal{M}$, only $< Bp^{1-a_2}\log p < Bn^{1-a_2}\log n$ iterations of the map F will have passed. This number of iterations is $o(n)$, so we will still have $n - o(n)$ iterations of F to go.

Now we consider those $B\log p$ returns to $\mathcal{M}$ assuming that B is large enough. During the first half of that sequence of returns, i.e., during the first $\frac{1}{2}B\log p$ iterations of $\mathcal{F}$, the images of our unstable curves will grow exponentially (per number of returns to $\mathcal{M}$), so that the corresponding Z-function will decrease exponentially, and in the end (i.e., after $\frac{1}{2}B\log p$ iterations of $\mathcal{F}$) the Z-function will be of order one, so we get a *proper* standard family at that time.

Then during the second half, i.e., during the next $\frac{1}{2}B\log p$ iterations of $\mathcal{F}$, images of unstable curves in our proper standard family will start making "full returns" to the hyperbolic set Λ_0 used to define the base of the Young tower, i.e., they will stretch completely across Λ_0. When a curve W stretches across Λ_0, we register a "return" for points $X \in W \cap \Lambda_0$, take them out of circulation, and continue iterating the rest of W, i.e., $W \setminus \Lambda_0$, under F. The relative measure of the remaining points (not yet stopped due to a "full return" to Λ_0) will decrease exponentially per number of returns to $\mathcal{M}$ (by a standard argument used in the proof of the Coupling Lemma [14, Chapter 7]).

Thus at the end of our series of $\frac{1}{2}B\log p$ iterations of $\mathcal{F}$ those remaining points constitute a subset of measure $\mathcal{O}(p^{-3-A})$ for some large $A>0$, so they can be discarded and their measure can be incorporated into $\chi_p=\mathcal{O}(p^{-3-a})$.

We denote by $M''_{p,i}\subset M'_{p,i}$ the set of points whose images do make a "full return" to Λ_0 during the above series of iterations; then

$$\mu(M''_{p,i})\ge\big(1-\mathcal{O}(p^{-a_1})\big)\mu(M_{p,i}).$$

We note that each point $X\in M''_{p,i}$ makes a "full return" to Δ_0 at a certain time $r(X)$, i.e., $F^{r(X)}(X)\in\Delta_0$. Accordingly, $M''_{p,i}=\cup_r M''_{p,i}(r)$, where $M''_{p,i}(r)=\{X\colon r(X)=r\}$.

The measure μ conditioned on $F^r\big(M''_{p,i}(r)\big)$ induces a probability measure, we call it $\mu_{p,i}(r)$, on the base Δ_0 of the Young tower $\Delta=\Delta_F$. Due to the regularity properties $d\mu_{p,i}(r)/d\mu_\Delta<C$ for some constant $C>0$, where μ_Δ denotes the invariant measure on the tower Δ. Hence further images of the measure $\mu_{p,i}(r)$ under the tower dynamics will be absolutely continuous with respect to μ_Δ with densities bounded by the same constant C. Since $r<n$, the measure μ conditioned on $F^n\big(M''_{p,i}(r)\big)$ induces a probability measure on the Young tower Δ with density $\le C$ with respect to μ_Δ. Averaging over r implies that the measure μ conditioned on $F^n\big(M''_{p,i}\big)$ induces a probability measure on Δ with density $\le C$.

Now recall that $\Lambda_0\subset\mathcal{M}_1$, so the base of the tower corresponds to points in $\mathcal{M}_1$. Points in $M_{q,j}$ have to make at least $j+1$ iterations in the past to get to $\mathcal{M}_1$ and at least $q-j$ iterations in the future to get to $\mathcal{M}_1$, i.e., it will take at least $q/2>m/2$ iterations to get to $\mathcal{M}_1$ either in the past or in the future. Thus the set $\cup_{q>m}\cup_{j=0}^{q-1}M_{q,j}$ corresponds to kth columns $\Delta^{(k)}$ of the tower with $k\ge m/2$, and the combined μ_Δ-measure of those columns is $\mathcal{O}(1/m)$. Therefore

$$\mu\big(F^n(M''_{p,i})\cap(\cup_{q>m}\cup_{j=0}^{q-1}M_{q,j})\big)=\mathcal{O}\big(\mu(M_{p,i})/m\big)$$

which completes the proof of Lemma 4.1. □

We note that in the above proof we only needed to transform the images of $M''_{p,i}$ to the set Λ_0 corresponding to the base of the Young tower Δ, it did not matter how long we iterated them further. Thus we can replace F^n with F^K with $K=Bn^{1-a_2}\log n$.

In particular, the measure μ conditioned on $F^K\big(M''_{p,i}(r)\big)$ induces a probability measure on the Young tower Δ with density $\le C$ with respect to the invariant measure μ_Δ. Now further images of this measure will converge to μ_Δ at a polynomial rate, i.e., after $N\ge1$ iterations

of F they will be $\mathcal{O}(1/N)$-close to μ_Δ. This follows from Young's coupling argument [40] in a non-invertible setting, and can be extended to the present (hyperbolic, invertible) case by an approximation argument ([33, Appendix B]). As a result, we obtain

Corollary 4.3. *We have*

$$\int_{M_{p,i}} A \circ F^{K+N}\, d\mu = \mu(M_{p,i})\big[\mu(A) + \mathcal{O}(1/N)\big] + \chi_p \tag{4.2}$$

where $K = Bn^{1-a_2}\log n$ *and* $\chi_p = \mathcal{O}(p^{-3-a})$, *as before.*

Lastly, all our estimates apply to the inverse map F^{-1}, too.

5. Proof of Lemma 3.1

First, we derive a crude estimate by "brute force", i.e., by using the absolute values of the function A. That estimate will be a little unsatisfactory, and we will then take extra steps to improve it.

Recall that in Lemma 3.1 we need to estimate additions and removals in the course of replacing the constant time limits $\pm n$ in (3.2) with the variable time limits $n^{\pm}(X)$ in (3.5).

Note that additions occur when $n^+(X) > n$, i.e., when $i + j' + 1 < q'$. In that case we add products $A(X)A\big(F^{n+k}(X)\big)$ for $k = 1, \dots, q' - (i + j' + 1)$. Now suppose $F^n(X) \in M_{q,j}$ for some $q = q(X) \geq 1$ and $j = j(X) \in [0, q-1]$. We see that $q = q'$ and $j = j' + i$. Then the points $F^{n+k}(X)$ from the above products are in the cells $M_{q,j+1}, \dots, M_{q,q-1}$. These are the last $q - j - 1$ cells in the series $M_{q,0}, \dots, M_{q,q-1}$.

We can phrase it differently. Suppose, as before, $X \in M_{p,i}$ and $F^n(X) \in M_{q,j}$ (we do not use q' and j' anymore). Then we will need to add terms if and only if $j \geq i$, and precisely we will add the products

$$A(X)A\big(F^{n+1}(X)\big), \dots, A(X)A\big(F^{n+q-j-1}(X)\big) \tag{5.1}$$

using the points $F^{n+1}(X)$, …, $F^{n+q-j-1}(X)$ from the cells $M_{q,j+1}$, …, $M_{q,q-1}$, respectively, i.e., until the current column of cells ends. Thus points from the "end" of every column of cells will be added more frequently than those from its "beginning"; more precisely points from $M_{q,j}$ may need to be added j times. Hence we get the total upper bound on the additions:

$$\|A\|_\infty^2 \sum_{q=1}^{n/10} \sum_{j=0}^{q-1} j\mu(\mathcal{M}_q) \leq \|A\|_\infty^2 \sum_{q=1}^{n/10} q^2 \mu(\mathcal{M}_q) = \mathcal{O}(\log n). \tag{5.2}$$

This is a little unsatisfactory, as we need $\mathcal{O}(1)$, and we will improve our estimate (5.2) next.

Note that we have not used any mixing properties of the map F in deriving (5.2). Due to mixing we expect the average values of the product (5.1) be much smaller than $\|A\|_\infty^2$.

Suppose for a moment that we will add the products (5.1) whenever $F^n(X)$ falls into $M_{q,j}$ (regardless of the condition $j \ge i$). Thus we will be adding a little more often than according to our actual rules expressed by (3.6).

Then effectively for each $q = 1, \dots, n/10$ and $j = 0, \dots, q-1$ we will add the terms

$$\mu\big(A|_{M_{q,j}} \cdot (A \circ F^{-n-s-1})\big) \qquad \text{for} \qquad s = 0, \dots, j \tag{5.3}$$

to the sum (3.2). These are correlations, which can be estimated by Corollary 4.3 applied to the map F^{-1}, and we get the desired bound

$$\mathcal{O}\Big(\sum_{q=1}^{n/10}\sum_{j=0}^{q-1}\sum_{s=0}^{j}\Big[\frac{\mu(M_{q,j})}{n} + \frac{1}{q^{3+a}}\Big]\Big) = \mathcal{O}(1).$$

Now we need to estimate the total contribution of the extra additions that were introduced above, contrary to the requirement $j \ge i$. Those extra additions correspond to the cases where $j < i$, i.e., where $X \in M_{p,i}$ with some $i > j$. These extra additions will be estimated by "brute force", i.e., by the absolute values of A. Their total contribution is bounded by

$$\begin{aligned}
&\|A\|_\infty^2 \sum_{q=1}^{n/10}\sum_{j=0}^{q-1}\sum_{p=j+2}^{n/10}\sum_{i=j+1}^{p-1}(q-j)\mu\big(F^n(M_{p,i}) \cap M_{q,j}\big) \\
&\le \|A\|_\infty^2 \sum_{q=1}^{n/10}\sum_{j=0}^{q-1}(q-j)\mu\Big(\big(\cup_{p=j+2}^{n/10}\cup_{i=0}^{p-1} M_{p,i}\big) \cap F^{-n}(M_{q,j})\Big)
\end{aligned}$$

Now the estimate (4.1) can be applied to the inverse map F^{-n}; it gives the bound

$$\text{const}\,\|A\|_\infty^2 \sum_{q=1}^{n/10}\sum_{j=0}^{q-1}(q-j)\big(q^{-3}j^{-1} + q^{-3-a}\big) = \mathcal{O}(1)$$

which is small enough to be incorporated in the error term of (3.2).

This completes our estimation of the additions that we have to make to convert (3.2) to (3.5).

The analysis of the removals is very similar, we only sketch it. Suppose, as before, $X \in M_{p,i}$ and $F^n(X) \in M_{q,j}$. Then we will need to remove terms if and only if $i > j$. More precisely, we will need to remove at least $j+1$ terms, down to the bottom of the column of cells

into which the point $F^n(X)$ falls. We may need to remove more, as we need to go down to the bottom of the column into which the point $F^{n-i}(X)$ falls. But in any case we will not remove more than i terms. So we will remove

$$A(X)A\big(F^{n-1}(X)\big), \dots, A(X)A\big(F^{n-i'}(X)\big) \tag{5.4}$$

for some $i' \in [j+1, i]$. Note that the condition $i > j$ implies $p > i > 0$, i.e., $p \ge 2$.

Again, suppose we remove a little more than our rules dictate, i.e., we will remove all the products

$$A(X)A\big(F^{n-1}(X)\big), \dots, A(X)A\big(F^{n-i}(X)\big)$$

whenever $X \in M_{p,i}$ with $p \ge 2$ (regardless of the condition $i > j$). Then effectively, for each $p = 2, \dots, n/10$ and $i = 1, \dots, p-1$ we will subtract the terms

$$\mu\big(A|_{M_{p,i}} \cdot (A \circ F^{n-s})\big) \qquad \text{for} \qquad s = 1, \dots, i \tag{5.5}$$

from the sum (3.2). These are correlations that can be readily estimated by Corollary 4.3, now applied to forward iterations of F.

It remains to estimate the extra removals that we had to make in order to form complete correlations (5.5). Those extra removals are made whenever $F^n(X) \in M_{q,j}$ with $j \ge i$. We will estimate them by "brute force", i.e., by the absolute value of A Their total contribution is bounded by

$$\begin{aligned} &\|A\|_\infty^2 \sum_{p=1}^{n/10} \sum_{i=0}^{p-1} \sum_{q=i+1}^{n/10} \sum_{j=i}^{q-1} i\mu\big(M_{p,i} \cap F^{-(n-1)}(M_{q,j})\big) \\ &\qquad \le \|A\|_\infty^2 \sum_{p=1}^{n/10} \sum_{i=0}^{p-1} i\mu\Big(F^{n-1}(M_{p,i}) \cap \big(\cup_{q=i+1}^{n/10} \cup_{j=0}^{q-1} M_{q,j}\big)\Big) \end{aligned}$$

Now the estimate (4.1) gives the bound

$$\text{const}\, \|A\|_\infty^2 \sum_{p=1}^{n/10} \sum_{i=0}^{p-1} i\big(p^{-3} i^{-1} + p^{-3-a}\big) = \mathcal{O}(1)$$

which is small enough to be incorporated in the error term of (3.2).

This completes our estimation of the removals that we have to make to convert (3.2) to (3.5).

6. Proof of Lemma 3.3

We need to compare the correlation sum (3.8) with a variable time limit, $k^+(X)$, to the correlation sum (3.9) with a fixed time limit, $2n$. Since $2n > k^+(X)$ for any point $X \in \mathcal{M}_{1,n/10}$, we need to show that

$$\mu\Big(\mathcal{A}_{1,\frac{n}{10}} \cdot \Big(\sum_{k=k^+(X)+1}^{2n} \mathcal{A}\circ\mathcal{F}^k\Big)\Big) = \mathcal{O}(1). \tag{6.1}$$

We can replace μ with ν, as these two measures are proportional on $\mathcal{M}$. Then we can rewrite the left hand side of (6.1) as follows:

$$\mu\Big(A|_{\tilde{M}} \cdot \Big(\sum_{k=k^+(X)+1}^{2n} \mathcal{A}\circ\mathcal{F}^k\Big)\Big), \tag{6.2}$$

where $\tilde{M} = \sum_{m=1}^{n/10} M_m$, and we assume that the map $\mathcal{F}$ and the function k^+ are naturally extended to M by the rules

$$\mathcal{F}(X) = \mathcal{F}\big(\Pi(X)\big) \qquad \text{and} \qquad k^+(X) = k^+\big(\Pi(X)\big) \tag{6.3}$$

where $\Pi(X) = F^{-i}X$ whenever $X \in M_{m,i}$ is, in a sense, a natural projection of M onto $\mathcal{M}$; see Fig. 1.

We will describe the expression (6.2) differently. First, recall that $k^+(X)$ for $X \in \mathcal{M}_{1,n/10}$ was defined in Section 3 so that

$$\sum_{k=0}^{k^+(X)} \mathcal{R}(\mathcal{F}^k X) = n^+(X) = n + q' - j' - 1,$$

where we use the notation of Section 3, according to which $X \in M_{p,i}$ and $T^n(\Pi(X)) \in M_{q',j'}$. Therefore $k^+(X)$ can be defined by

$$\sum_{k=0}^{k^+(X)-1} \mathcal{R}(\mathcal{F}^k X) < n \le \sum_{k=0}^{k^+(X)} \mathcal{R}(\mathcal{F}^k X). \tag{6.4}$$

or, by using shorthand notation $\mathcal{S}_k = \sum_{i=0}^{k-1} \mathcal{A}\circ\mathcal{F}^i$, we have

$$\mathcal{S}_{k^+(X)}\mathcal{R}(X) < n \le \mathcal{S}_{k^+(X)+1}\mathcal{R}(X).$$

With the extension (6.3), and a similar extension of $\mathcal{R}$ defined by $\mathcal{R}(X) = \mathcal{R}\big(\Pi(X)\big)$, all the above formulas apply to every point $X \in \tilde{M}$.

By changing variable $Y = F^n(X)$ we can rewrite (6.2) as

$$\mu\Big(\big(A\circ F^{-n}(Y)\big) \cdot \Big(\sum_{k=0}^{n^*(Y)} \mathcal{A}\circ\mathcal{F}^k\big(F^{r^*(Y)}(Y)\big)\Big)\Big) \tag{6.5}$$

or equivalently

$$\mu\Big(\big(A\circ F^{-n}(Y)\big)\cdot\Big(\sum_{k=r^*(Y)}^{K^*(Y)} A\circ F^k(Y)\Big)\Big). \tag{6.6}$$

The number of summands in (6.2) and (6.5) must be the same, so $n^*(Y)=2n-k^+(F^{-n}Y)-1$, though $n^*(Y)$ and $K^*(Y)$ will not be so important to us. We will determine (and modify) $r^*(Y)$ next.

As in Section 3, let $X\in M_{p,i}$ and $F^n(\Pi(X))=F^{n-i}(X)\in M_{q',j'}$. Then $r^*(Y)=q'-j'-i$, in accordance with (6.4). In other words, the summation in (6.6) begins when the future trajectory of the point $F^{n-i}(X)$ first returns to the base $\mathcal{M}$. We want to modify $r^*(Y)$ so that the summation will begin when the future trajectory of the point $F^n(X)$ first returns to the base $\mathcal{M}$. Let $Y=F^n(X)\in M_{q,j}$, as in Section 3. According to the above remarks, we define $r^*_{\text{new}}=q-j$. There are two cases:

Case 1: $i\le j$. Then $F^{n-i}(X)$ is in the same column of cells as $F^n(X)$, hence $q'=q$, $j'=j-i$ and $r^*_{\text{new}}=r^*$, so no modification is needed.

Case 2: $i>j$. Then $n-i<n-j$ and moreover $n-i+q'-j'\le n-j$. Thus we need to remove from (6.6) the terms

$$A(X)A\big(F^{n-i+q'-j'}(X)\big),\dots,A(X)A\big(F^{n+q-j-1}(X)\big) \tag{6.7}$$

which form (one or several) complete columns.

The procedure for addition and removal of similar terms was well developed in Section 5; we only sketch our main steps here.

First, we remove all the terms

$$A(X)A\big(F^{n-i+1}(X)\big),\dots,A(X)A\big(F^{n}(X)\big)$$

in both cases 1 and 2 (i.e., regardless of the condition $i>j$). The averages of the above products are correlations whose contribution to (6.6) is $\mathcal{O}(1)$ based on Corollary 4.3.

Second, in Case 2 (when $i>j$) we remove the products

$$A(X)A\big(F^{n+1}(X)\big),\dots,A(X)A\big(F^{n+q-j-1}(X)\big).$$

Their contribution is estimated by "brute force" as

$$\begin{aligned}&\|A\|_\infty^2\sum_{q=1}^{n/10}\sum_{j=0}^{q-1}\sum_{p=j+2}^{n/10}\sum_{i=j+1}^{p-1}(q-j)\mu\big(F^n(M_{p,i})\cap M_{q,j}\big)\\ &\le\|A\|_\infty^2\sum_{q=1}^{n/10}\sum_{j=0}^{q-1}(q-j)\mu\Big(\big(\cup_{p=j+2}^{n/10}\cup_{i=0}^{p-1}M_{p,i}\big)\cap F^{-n}(M_{q,j})\Big)\end{aligned}$$

which by (4.1) applied to the inverse map F^{-n} gives

$$\text{const}\,\|A\|_\infty^2 \sum_{q=1}^{n/10}\sum_{j=0}^{q-1}(q-j)\big(q^{-3}j^{-1}+q^{-3-a}\big)=\mathcal{O}(1).$$

Third, in Case 1 (when $i\le j$) we add back the products

$$A(X)A\big(F^{n-i+1}(X)\big),\dots,A(X)A\big(F^n(X)\big)$$

Their contribution is estimated by "brute force" as

$$\|A\|_\infty^2 \sum_{p=1}^{n/10}\sum_{i=0}^{p-1}\sum_{q=i+1}^{n/10}\sum_{j=i}^{q-1} i\mu\big(F^n(M_{p,i})\cap M_{q,j}\big)$$
$$\le \|A\|_\infty^2 \sum_{p=1}^{n/10}\sum_{i=0}^{p-1} i\mu\Big(F^n(M_{p,i})\cap\big(\cup_{q=i+1}^{n/10}\cup_{j=0}^{q-1}M_{q,j}\big)\Big)$$

which by (4.1) gives an upper bound of

$$\text{const}\,\|A\|_\infty^2 \sum_{p=1}^{n/10}\sum_{i=0}^{p-1} i\big(p^{-3}i^{-1}+p^{-3-a}\big)=\mathcal{O}(1).$$

Lastly, in Case 2 (when $i>j$ and $i\ge q'-j'$) we add back the products

$$A(X)A\big(F^{n-i+1}(X)\big),\dots,A(X)A\big(F^{n-i+q'-j'-1}(X)\big).$$

Their contribution is estimated by "brute force" as

$$\|A\|_\infty^2 \sum_{q'=1}^{n/10}\sum_{j'=0}^{q'-1}\sum_{p=q'-j'}^{n/10}\sum_{i=q'-j'}^{p-1}(q'-j')\mu\big(F^n(M_{p,i})\cap M_{q',j'}\big)$$
$$\le \|A\|_\infty^2 \sum_{q'=1}^{n/10}\sum_{j'=0}^{q'-1}(q'-j')\mu\Big(\big(\cup_{p=q'-j'}^{n/10}\cup_{i=0}^{p-1}M_{p,i}\big)\cap F^{-n}(M_{q',j'})\Big)$$

which by (4.1) applied to the inverse map F^{-n} gives

$$\text{const}\,\|A\|_\infty^2 \sum_{q'=1}^{n/10}\sum_{j'=0}^{q'-1}(q'-j')\big([q']^{-3}[q'-j']^{-1}+[q']^{-3-a}\big)=\mathcal{O}(1).$$

As a result, we can replace r^* with r^*_{new} and rewrite (6.5) as

$$\mu\Big(\big(A\circ F^{-n}(Y)\big)\cdot\Big(\sum_{k=0}^{n^*(Y)}\mathcal{A}\circ\mathcal{F}^k\big(\mathcal{F}\circ\Pi(Y)\big)\Big)\Big)+\mathcal{O}(1). \tag{6.8}$$

We will denote $\Pi^+=\mathcal{F}\circ\Pi$. This is also a projection of M onto $\mathcal{M}$, but it takes every point $X\in M$ to its first *future* image in $\mathcal{M}$.

One may wonder why we wanted to transform one sum of correlations with a variable upper limit into another, i.e., (3.7) into (6.5). The reason is that $k^+(X)$ in (3.7) was determined by the *future* of the point X, so it was constant on stable manifolds. On the contrary, $n^*(Y)$ in (6.8) is determined by the *past* of the point Y, so it is constant on unstable manifolds; this will be essential for the next steps.

Let $W_\alpha \subset \tilde{M}$ denote the family of all unstable manifolds for the map F with conditional probability measures ρ_α on them; here $\alpha \in \mathfrak{A}$ is some index set. Let λ denote the respective factor measure on $\mathfrak{A}$. We denote by $n^*(W_\alpha)$ the common value of $n^*(Y)$ on W_α. Now (6.8) can be written as

$$\int_{\mathfrak{A}} \left[\sum_{k=0}^{n^*(W_\alpha)} \rho_\alpha\big((A \circ F^{-n})(\mathcal{A} \circ \mathcal{F}^k \circ \Pi^+)\big) \right] d\lambda + \mathcal{O}(1). \tag{6.9}$$

Recall that we need to show that this expression is $\mathcal{O}(1)$.

The function $A \circ F^{-n}$ is almost constant on each W_α, so its fluctuations can be incorporated into the density of ρ_α. In other words, it is enough to show that

$$\int_{\mathfrak{A}} \left[\sum_{k=0}^{n^*(W_\alpha)} \Pi^+ \rho'_\alpha\big(\mathcal{A} \circ \mathcal{F}^k\big) \right] d\lambda = \mathcal{O}(1) \tag{6.10}$$

for any regular probability measures ρ'_α on the unstable manifolds W_α with the above factor measure λ. The rest of this section is devoted to proving (6.10).

Note that $\Pi^+\rho'_\alpha$ is a smooth probability measure on the curve $\Pi^+ W_\alpha \subset \mathcal{M}$, which is an unstable manifold for the map $\mathcal{F}$. Thus the sum within the brackets in (6.10) is the sum of integrals of a fixed function, $\mathcal{A}$, with respect to the images, under the iterations of $\mathcal{F}$, of a standard pair, $(\Pi^+ W_\alpha, \Pi^+\rho'_\alpha)$. We consider *proper* standard pairs first.

We will need the following fact.

Lemma 6.1. *(Coupling Lemma; see* [14, Sect. 7]*).*

Let $\tilde{\nu}$ be a measure on $\mathcal{M}$ corresponding to a proper standard family and ν be the smooth invariant measure. Then there are constants $C > 0, \theta < 1$, a measure ζ (coupling) on $\mathcal{M} \times \mathcal{M}$ and a measurable function τ defined ζ almost everywhere such that

(a) Marginals of ζ are ν and $\tilde{\nu}$;

(b)$d(\mathcal{F}^n X, \mathcal{F}^n \tilde{X}) \le C\theta^{n-\tau(X,\tilde{X})}$;

(c) $\zeta(\tau > n) \le C\theta^{2n}$.

If $\tau(X, \tilde{X}) \le n$ we say that X and $\tilde{X}$ are coupled by time n. We refer to second marginal of ζ restricted to $\{\tau \le n\}$ as "the coupled part of $\tilde{\nu}$ at time n."

Lemma 6.2. *For any proper standard pair (W, ρ) in $\mathcal{M}$ we have*

$$\sum_{k=0}^{N} \rho\big(\mathcal{A} \circ \mathcal{F}^k\big) = \mathcal{O}(1) \tag{6.11}$$

uniformly in (W, ρ) and $N \ge 0$.

Proof. The Z-function of (W, ρ) is $\le C_{\mathrm{p}}$ (cf. Section 2), so $\rho(\mathcal{M}_m) \le Cm^{-7/3}$ for some absolute constant $C > 0$ because the width of $\mathcal{M}_m$ in the unstable direction is $\mathcal{O}(m^{-7/3})$. This implies

$$\rho(|\mathcal{A}|) \le C\|A\|_\infty \sum_{m \ge 1} m/m^{7/3} \le C_{\mathcal{A}},$$

where $C_{\mathcal{A}} > 0$ is independent of (W, ρ).

The images under forward iterations of $\mathcal{F}$ of a proper standard pair (W, ρ) are proper standard families (whose Z-function is $\le C_{\mathrm{p}}$), so the above argument gives $\rho(|\mathcal{A} \circ \mathcal{F}^k|) \le C_{\mathcal{A}}$ for all $k \ge 0$.

Next, after k iterations of $\mathcal{F}$, most of the measure $\mathcal{F}^k\rho$ will be coupled with the $\mathcal{F}$-invariant measure ν, due to the Coupling Lemma 6.1. More precisely, the fraction of $\mathcal{F}^k\rho$ which has not been coupled with ν within the first $k/2$ iterations of $\mathcal{F}$, has norm $\le C\theta^k$ for some absolute constants $C > 0$ and $\theta \in (0, 1)$. For brevity let us introduce the notations $\rho_k = \mathcal{F}^k\rho$, and $\hat{\rho}_k$ for the uncoupled part of ρ_k. Knowing that $\hat{\rho}_k(\mathcal{M}) \le C\theta^k$, we would like to estimate $\hat{\rho}_k(|\mathcal{A}|)$, where $\mathcal{A}$ is unbounded: its value is proportional to m on the cell $\mathcal{M}_m$. The worst case scenario is when ρ_k gives the largest possible weight to the "highest" cells $\cup_{m \ge m_{0,k}} \mathcal{M}_m$, and all this weight corresponds to the uncoupled part $\hat{\rho}_k$. However, as ρ_k is a proper standard family, by the above argument we have $\rho_k(\mathcal{M}_m) \le Cm^{-7/3}$ uniformly in k. Hence to estimate $\hat{\rho}_k(|\mathcal{A}|)$ we set

$$C\theta^k = \hat{\rho}_k(\mathcal{M}) = \hat{\rho}_k(\cup_{m \ge m_{0,k}} \mathcal{M}_m) = \sum_{m=m_{0,k}}^{\infty} Cm^{-7/3} = Cm_{0,k}^{-4/3}$$

so that $m_{0,k} = C\theta^{-3k/4}$, and we have

$$\hat{\rho}_k(|\mathcal{A}|) \le C \sum_{m=m_{0,k}}^{\infty} m \cdot m^{-7/3} = Cm_{0,k}^{-1/3} = C\theta^{k/4}.$$

Now due to the coupling we have

$$\big|\rho(\mathcal{A}\circ\mathcal{F}^k)-\nu(\mathcal{A})\big|\le\chi_0+\chi_1+\chi_2, \tag{6.12}$$

where χ_0 accounts for $\hat\rho_k$ – the "uncoupled" part of ρ_k – which we have estimated as $|\chi_0|\le C\theta^{k/4}$. Similarly, χ_1 accounts for the "uncoupled" part of ν, for which we have

$$|\chi_1|\le\sup_{B\colon\nu(B)<C\theta^k}\int_B|\mathcal{A}|\,d\nu\le C_{\mathcal{A}}\theta^{k/2}$$

by a similar argument, using that $\nu(\mathcal{M}_m)=\mathcal{O}(m^{-3})$. Finally, the term χ_2 in (6.12) accounts for the variation of $\mathcal{A}$ on stable manifolds containing pairs of points that have been coupled together. If two points have been coupled during the first $k/2$ iterations of $\mathcal{F}$, then during the next $k/2$ iterations of $\mathcal{F}$ their images get exponentially close, i.e., at the kth iteration of $\mathcal{F}$ for any pair of coupled points X,Y we have $\operatorname{dist}(X,Y)\le C\theta^k$ for some absolute constants $C>0$ and $\theta\in(0,1)$. If $X,Y\in\mathcal{M}_m$, then

$$|\mathcal{A}(X)-\mathcal{A}(Y)|\le\sum_{i=0}^{m-1}|A(F^i(X))-A(F^i(Y))|=\mathcal{O}(m\theta^{\gamma_A k}),$$

where $\gamma_A>0$ is the Hölder exponent of A. Therefore

$$|\chi_2|=\mathcal{O}\Big(\sum_{m=1}^{\infty}\frac{m\theta^{\gamma_A k}}{m^3}\Big)=\mathcal{O}(\theta^{\gamma_A k}).$$

Thus all the terms in (6.12) are $\mathcal{O}(\theta^k)$ for some $\theta\in(0,1)$, uniformly in (W,ρ). Summing up over k proves (6.11). □

If all the standard pairs $(\Pi^+W_\alpha,\Pi^+\rho'_\alpha)$ were proper, then integration with respect to λ would readily give us (6.10). But there are arbitrarily short unstable manifolds in $\mathcal{M}$, which cannot be proper; they will be handled next.

For any point $X\in\Pi^+W_\alpha$ let $n^\dagger(X)$ denote the first iteration of F such that $F^{n^\dagger(X)}(X)\in\mathcal{M}$ and the unstable manifold W' of the map $\mathcal{F}$ that contains the point $F^{n^\dagger(X)}(X)$ is long enough to make a proper standard pair with any regular probability measure on it. The image $F^{n^\dagger(X)}\circ\Pi^+\rho_\alpha$ restricted to W' and conditioned on W' will be a smooth probability measure ρ', and (W',ρ') will make a proper standard pair.

It is known that $n^\dagger(X)$ is finite ρ_α-a.e. (we will also see this in the proof of Lemma 6.3). Note that $n^\dagger(X)$ is constant on the subcurve $F^{-n^\dagger(X)}(W')\subset\Pi^+W_\alpha$. Thus Π^+W_α is divided into subcurves, each of which is transformed into a proper standard family in $\mathcal{M}$ under a certain iteration of F.

Now further images of the proper standard pair (W', ρ') can be handled by Lemma 6.2. It remains to account for the $n^\dagger(X)$ iterations of each point $X \in \Pi^+ W_\alpha$ before it falls into a proper standard pair. This will be done by "brute force", i.e., by the absolute value of A. It is clear from (6.10) that the contribution from those iterations are bounded by $\|A\|_\infty n^\dagger(X)$. Thus it remains to show that $\int_{\mathfrak{A}} \Pi^+ \rho'_\alpha(n^\dagger)\, d\lambda = \mathcal{O}(1)$. Since the density of ρ'_α with respect to ρ_α cannot exceed $\|A\|_\infty$, this is equivalent to

$$\int_{\mathfrak{A}} \Pi^+ \rho_\alpha(n^\dagger)\, d\lambda = \mathcal{O}(1). \tag{6.13}$$

It may happen that $n^\dagger(W) > n^*(W_\alpha)$ for some $X \in \Pi^+ W_\alpha$, so (6.13) might be an overestimation. But since the integrand in (6.13) is positive, it would not hurt. Also note that if $X \in \mathcal{M}$ already belongs to a proper standard pair, then $n^\dagger(X) = 0$, so (6.13) includes such points, too. The estimate (6.13) can be stated, equivalently, as follows:

Lemma 6.3. *We have* $\mu(n^\dagger \circ \Pi^+) = \Pi^+ \mu(n^\dagger) < \infty$.

In other words, the average number of iterations of F it takes for points $Y \in \mathcal{M}$ to get into long unstable manifolds is finite. Note that averaging is done with respect to the projected measure $\Pi^+ \mu$, rather than the invariant measure ν of the induced map $\mathcal{F}$. These measures are different: $\Pi^+ \mu(\mathcal{M}_m) \asymp 1/m^2$ while $\nu(\mathcal{M}_m) \asymp 1/m^3$.

Proof. Let $\Delta = \Delta_F$ again denote the Young tower modeling the map $F\colon M \to M$. Throughout the proof, we will us the notations introduced in Section 4 concerning the Young tower. That is, $T\colon \Delta_F \to \Delta_F$ will denote the tower map, μ_Δ the invariant measure on Δ, Δ_0 the base, Δ_m the mth level, and $\Delta^{(m)}$ the mth column of the tower. Let, furthermore, $\Delta_\mathcal{M} \subset \Delta$ denote the part of the tower corresponding to the subset $\mathcal{M} \subset M$.

For each point $Y \in \Delta$ denote

$$k_\mathcal{M} = \min\{k \geq 1 \colon T^k(Y) \in \Delta_\mathcal{M}\}$$

the first time the trajectory of Y visits $\Delta_\mathcal{M}$ and

$$k_0 = \min\{k \geq 1 \colon T^k(Y) \in \Delta_0\}$$

the first time the trajectory of Y returns to the base Δ_0.

Suppose $Y \in \Delta$ models a point $X \in M$. Then $T^{k_\mathcal{M}(Y)}(Y)$ models the point $\Pi^+(X)$. Next, whenever $T^n(Y) \in \Delta_0$, the corresponding point $F^n(X)$ belongs to the basic hyperbolic rectangle (which Young called a horseshoe with hyperbolic structure) on which the tower is

constructed. In particular, $F^n(X)$ belongs to a long enough unstable manifold which qualifies for a proper standard pair. As a result,

$$n^\dagger\big(\Pi(X)\big) \le k_0\big(T^{k_{\mathcal{M}}(Y)}(Y)\big) = k_0(Y) - k_{\mathcal{M}}(Y).$$

By the way, note that $k_0(Y) < \infty$ for every $Y \in \Delta$, hence $n^\dagger(X') < \infty$ for a.e. point $X' \in \mathcal{M}$, as we mentioned earlier.

Now to prove Lemma 6.3 it will be enough to show that

$$\mu_\Delta(k_0 - k_{\mathcal{M}}) < \infty. \tag{6.14}$$

Next we prove (6.14) for our Young tower. Note that $k_0 = m - i + 1$ on $\Delta_{i,m}$, hence

$$\mu_\Delta(k_0) \sim \sum_m m^2 \mu_\Delta(\Delta_{0,m}) = \infty, \tag{6.15}$$

therefore subtracting $k_{\mathcal{M}}$ in (6.14) is essential.

Subtraction of $k_{\mathcal{M}}$ requires a delicate procedure. We claim that each set $\Delta_{0,m}$ can be divided into a good part $\Delta^{\mathrm{g}}_{0,m}$ and a bad part $\Delta^{\mathrm{b}}_{0,m}$ with the following properties. First, the relative measure of the bad part is small, i.e., $\mu_\Delta(\Delta^{\mathrm{b}}_{0,m}) < Cm^{-3-a}$ for some constants $C, a > 0$, thus its contribution to (6.14) and (6.15) is finite, because $\sum_m m^2 \mu_\Delta(\Delta^{\mathrm{b}}_{0,m}) < \infty$, hence it can be neglected.

Second, for each point $Y \in \Delta^{\mathrm{g}}_{0,m}$ in the good part there are $0 \le p < q \le m$ such that $T^p(Y)$ models a phase point $X \in \mathcal{M}_{q-p}$ hence $\mathcal{F}(X) = F^{q-p}(X)$. In addition, we have $\max\{p, m-q\} < Cm^{1-a}$ for some constants $C, a > 0$. As a consequence, for each $Y \in \Delta^{\mathrm{g}}_{0,m}$ and $j \in [p, q]$ we have $k_{\mathcal{M}}(T^jY) = q - j$ and hence $k_0(T^jY) - k_{\mathcal{M}}(T^jY) = m - q < Cm^{1-a}$. Thus the contribution of the points $Y, TY, ..., T^{m-1}Y$ to (6.14) will be bounded by $mp + Cm^{1-a}(m-p) < 2Cm^{2-a}$. Hence

$$\mu_\Delta(k_0 - k_{\mathcal{M}}) \le 2C \sum_m m^{2-a} \mu_\Delta(\Delta_{0,m}) < \infty$$

as required. This completes the proof of (6.14).

It remains to construct the good and bad parts of $\Delta_{0,m}$. We follow the argument in [18, Sect. 5]. For each $Y \in \Delta_{0,m}$ denote by $R(Y) = \#\{0 \le i \le m \colon T^i(Y) \in \Delta_{\mathcal{M}}\}$ the number of times the trajectory of Y visits $\Delta_{\mathcal{M}}$ as it moves up the column. Points $Y \in \Delta_{0,m}$ for which $R(Y) > \bar{C} \log m$, where $\bar{C} > 0$ is a large constant, make a set of measure $< m^{-3-a}$, where $a = a(\bar{C}) > 0$ (see [18, p. 309]), so they are included into the bad part. For other points the largest interval $[p, q] \subset [0, m]$ between successive returns to $\Delta_{\mathcal{M}}$ has length $r = q - p \ge m/(\bar{C} \log m)$. Thus the point $T^p(Y)$ models a phase point $X \in \mathcal{M}_r$.

Next apply Lemma 4.2 with $B \gg \bar{C}$. If $X \in \mathcal{M}_r \setminus \mathcal{M}'_r$, then we include Y into the bad part, too. If $X \in \mathcal{M}'_r$, then the first $B \log r$ images of X under $\mathcal{F}$ can only visit cells $\mathcal{M}_s$ with $s < r^{1-a_2}$. Hence the next $B \log r \gg \bar{C} \log m$ intervals between successive returns to $\Delta_{\mathcal{M}}$ within our column are shorter than r^{1-a_2}. This implies $m - q < Cm^{1-a}$ for some constant $C, a > 0$. Similarly, we apply Lemma 4.2 to $\mathcal{F}^{-1}$ and handle the first $C \log r$ images of X under $\mathcal{F}^{-1}$, and this will ensure $p < Cm^{1-a}$. This completes the proof of (6.14) and Lemma 6.3. □

Now (6.10) is fully established and Lemma 3.3 is proved.

Acknowledgments. This work was done during our visit at the Fields Institute in Toronto in June 2011, and we acknowledge its hospitality. We are grateful to F. Bonetto, C. Dettmann, S. Gouëzel, P. Jung, J. Lebowitz, M. Lenci, and I. Melbourne for useful discussions. P. Bálint was partially supported by the Bolyai scholarship of the Hungarian Academy of Sciences and Hungarian National Fund for Scientific Research (NKFIH OTKA) grants F60206, K71693 and T104745. N. Chernov was partially supported by NSF grant DMS-0969187. D. Dolgopyat was partially supported by NSF grant DMS-1101635.

Appendix

The dynamics in dispersing billiards with cusps is characterized by intermittence: periods of chaotic bounces away from cusps intersperse with long series of collisions deep in a cusp; during the latter the observed values $A \circ F^n$ change slowly. Here we describe a simple stochastic process which exhibits similar features and show that the doubling effect takes place as well.

Our stochastic process $\xi(t)$ has continuous time $t > 0$ and takes values ± 1. Switching from one value to the other occurs at random moments $0 < T_0 < T_1 < \cdots$, and intervals between switching times, $L_k = T_k - T_{k-1}$, are independent identically distributed random variables with a polynomial tail bound $\mathbb{P}(L_k > x) \sim cx^{-2}$ for $x \to \infty$. We denote by $\mathbb{E}(L_k) = \mu$ their common mean value.

We note that T_0 can be chosen so that the sequence $\{T_k\}$ will be stationary in the following sense. For each $t > 0$, denote $m(t) = \min\{m \geq 0 \colon T_m > t\}$ and $H(t) = T_{m(t)} - t$. Then the stationarity means that

$$\mathbb{P}(H(t) > u) = \mathbb{P}(T_0 > u)$$

does not depend on t (for each $u > 0$). By [25], Chapter XI, Equation (4.6) we have

$$\mathbb{P}(T_0 > t) = \frac{1}{\mu} \int_t^\infty P(L_k > x)\, dx \sim \frac{c}{\mu t}.$$

Now we define our process $\xi(t)$. Let $\xi_0, \xi_1, \dots$ be i.i.d. random variables taking values ± 1, each with probability $1/2$. We set $\xi(t) = \xi_k$ if $t \in [T_{k-1}, T_k]$ and $\xi(t) = \xi_0$ if $t < T_0$.

Now consider $S(T) = \int_0^T \xi(t)\, dt$. Denote $\mathcal{L}_k = L_k \xi_k$ and $S_m = \sum_{k=0}^m \mathcal{L}_k$. Then obviously $S(T) \sim S_{m(T)}$ as $T \to \infty$. By [25], Section XVII.5, Theorem 2 the sequence $S_m/\sqrt{mV_m}$ converges in distribution to the standard normal law $\mathbf{N}(0,1)$, where

$$V_m = \int_1^{\sqrt{cm}} x^2\, d\mathbb{P}(L_k < x) \sim c \ln m. \tag{A.1}$$

By the Law of Large Numbers, $m(T) \sim T/\mu$, so that

$$\frac{S(T)}{\sqrt{T \ln T}} \Rightarrow \mathbf{N}\left(0, \frac{c}{\mu}\right)$$

as $T \to \infty$. On the other hand,

$$\mathbb{E}\big(S^2(T)\big) = 2 \iint_{0<s<t<T} \mathbb{E}(\xi(s)\xi(t))\, ds\, dt.$$

Since ξ_k's are independent, we have

$$\mathbb{E}(\xi(s)\xi(t)) = \mathbb{P}\big(H(s) > t - s\big) = \mathbb{P}(T_0 > t - s).$$

Accordingly

$$\mathbb{E}\big(S^2(T)\big) \sim 2 \int_{0<s<T} \frac{c}{\mu} \ln(T - s)\, ds \sim \frac{2c}{\mu} T \ln T \tag{A.2}$$

hence we observe the doubling effect again. It can be traced to the upper limit $\sqrt{cm}$ in the integration (A.1). If we change it to cm, then V_m would double and would match the asymptotics of the second moment (A.2).

The fact that the variance of the limit distribution is only affected by values of $L_k \le \sqrt{cm}$ is similar to the fact that values of our induced function $\mathcal{A}$ larger than $\sqrt{n}$ do not affect the limit distribution of $(S_n A)/\sqrt{n \log n}$; see the end of Section 3.

References

[1] Armstead D. N., Hunt B. R., and Ott E. Anomalous diffusion in infinite horizon billiards. *Physical Review E*, 67(2):021110, 2003.

[2] Bálint P., Chernov N. and Dolgopyat D., *Limit theorems for dispersing billiards with cusps*, Comm. Math. Phys. **308** (2011), 479–510.

[3] Bálint P. and Gouëzel S., *Limit theorems in the stadium billiard*, Comm. Math. Phys. **263** (2006), 461–512.

[4] Bálint P. and Melbourne I., *Decay of correlations and invariance principles for dispersing billiards with cusps, and related planar billiard flows*, J. Stat. Phys. **133** (2008), 435–447.

[5] Benedicks M. and Young L.-S., *Markov extensions and decay of correlations for certain Hénon maps*, Asterisque **261** (2000), 13–56.

[6] Bunimovich L. A. and Sinai Ya. G. *On a fundamental theorem in the theory of scattering billiards*, Math. USSR Sb. **90** (1973), 415–431.

[7] Bunimovich L. A., *On billiards close to dispersing*, Math. USSR Sbornik, **23** (1974), 45–67.

[8] Bunimovich L. A. and Sinai Ya. G. *Statistical properties of Lorentz gas with periodic configuration of scatterers,* Comm. Math. Phys. **78** (1980/81), 479–497.

[9] Bunimovich L. A., Sinai Ya. G., and Chernov N. I. *Statistical properties of two-dimensional hyperbolic billiards,* Russ. Math. Surv. **46** (1991), 47–106.

[10] Chazottes J.–R. and Gouëzel S., *Optimal concentration inequalities for dynamical systems,* Communications in Math. Physics, 316(3):843–889, 2012.

[11] Chernov N. *Decay of correlations and dispersing billiards*, J. Stat. Phys. **94** (1999), 513–556.

[12] Chernov N. and Dolgopyat D. *Brownian Brownian Motion-1,* Memoirs AMS **198** (2009) no. 927.

[13] Chernov N. and Dolgopyat D. *Anomalous current in periodic Lorentz gases with infinite horizon,* Russ. Math. Surv., **64** (2009), 73–124.

[14] Chernov N. and Markarian R., *Chaotic Billiards*, Mathematical Surveys and Monographs, **127**, AMS, Providence, RI, 2006. (316 pp.)

[15] Chernov N. and Markarian R., *Dispersing billiards with cusps: slow decay of correlations*, Comm. Math. Phys., **270** (2007), 727–758.

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

[17] Chernov N. and Zhang H.-K., *A family of chaotic billiards with variable mixing rates*, Stochast. Dynam. **5** (2005), 535–553.

[18] Chernov N. and Zhang H.-K., *Improved estimates for correlations in billiards.* Commun. Math. Phys., **277** (2008), 305–321.

[19] Courbage M., Edelman M., Saberi Fathi S. M., and Zaslavsky G. M., *Problem of transport in billiards with infinite horizon*, Phys. Rev. E **77** (2008), 036203.

[20] Cristadoro G., Gilbert T., Lenci M., and Sanders D. P., *Measuring logarithmic corrections to normal diffusion in infinite-horizon billiards*, Phys. Rev. E **90** (2014) 022106.

[21] Dedecker J. and Merlevède F., *Moment bounds for dependent sequences in smooth Banach spaces, Stochastic Processes and their Applications*, 2015.

[22] Dettmann, C., *New horizons in multidimensional diffusion: The Lorentz gas and the Riemann Hypothesis*, J. Stat. Phys. **146** (2012), 181–204

[23] Dettmann C. P. *Diffusion in the Lorentz gas,* Comm. Theor. Phys. **62** (2014) 521.
[24] Durrett R., *Probability: Theory and Examples*, 4th Ed., Cambridge U. Press, 2010.
[25] Feller W. *An introduction to probability theory and its applications,* Vol. II. 2nd edition John Wiley & Sons, Inc., New York-London-Sydney 1971 xxiv+669 pp.
[26] Gallavotti G. and Ornstein D., *Billiards and Bernoulli schemes*, Comm. Math. Phys. **38** (1974), 83–101.
[27] Gouëzel S., *Central limit theorem and stable laws for intermittent maps,* Probability Theory and Related Fields, **128** (2004) 82–122.
[28] Gouëzel S. and Melbourne I. *Moment bounds and concentration inequalities for slowly mixing dynamical systems,* Electron. J. Prob. **19** (2014) 1–30.
[29] Liverani C., Saussol B., Vaienti S., *A probabilistic approach to intermittency*, Ergod. Th. Dynam. Syst. **19** (1999), 671–685.
[30] Machta J., *Power law decay of correlations in a billiard problem*, J. Statist. Phys. **32** (1983), 555–564.
[31] Melbourne I., *Decay of correlations for slowly mixing flows*, Proc. London Math. Soc. **98** (2009) 163–190.
[32] Melbourne I. and Nicol M. *Large deviations for nonuniformly hyperbolic systems, Transactions of the AMS* **360** (2008) 661–676.
[33] Melbourne I. and Terhesiu D. *Decay of correlations for non-uniformly expanding systems with general return times,* Erg. Th. & Dyn. Sys. **34** (2014) 893–918.
[34] Melbourne I. and Török A. *Convergence of moments for axiom a and non-uniformly hyperbolic flows,* Erg. Th. & Dyn. Sys. **32** (2012) 1091–1100.
[35] Pomeau Y. and Manneville P., *Intermittent transition to turbulence in dissipative dynamical systems*, Comm. Math. Phys. **74** (1980), 189–197.
[36] Reháček J. *On the ergodicity of dispersing billiards*, Rand. Comput. Dynam. **3** (1995), 35–55.
[37] Sinai Ya. G., *Dynamical systems with elastic reflections. Ergodic properties of dispersing billiards*, Russ. Math. Surv. **25** (1970), 137–189.
[38] Szasz D. and Varju T. *Limit Laws and Recurrence for the Planar Lorentz Process with Infinite Horizon,* J. Statist. Phys. **129** (2007), 59–80.
[39] Young L.-S. *Statistical properties of dynamical systems with some hyperbolicity*, Ann. Math. **147** (1998) 585–650.
[40] Young L.-S. *Recurrence times and rates of mixing*, Israel J. Math. **110** (1999), 153–188.

P. Bálint: MTA–BME Stochastics Research Group, H-1111, Egry József u. 1, Budapest, Hungary; and Department of Stochastics, Budapest University of Technology and Economics, H-1111, Egry József u. 1, Budapest, Hungary;
D. Dolgopyat: Department of Mathematics, University of Maryland, College Park, MD 20742