Pointwise convergence of Birkhoff averages for global observables

Marco Lenci, and Sara Munday

Citation: Chaos **28**, 083111 (2018); doi: 10.1063/1.5036652
View online: https://doi.org/10.1063/1.5036652
View Table of Contents: http://aip.scitation.org/toc/cha/28/8
Published by the American Institute of Physics

CHAOS **28**, 083111 (2018)

Pointwise convergence of Birkhoff averages for global observables

Marco Lenci[1,2,a)] and Sara Munday[3,b)]
[1]*Dipartimento di Matematica, Università di Bologna, Piazza di Porta San Donato 5, 40126 Bologna, Italy*
[2]*Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Via Irnerio 46, 40126 Bologna, Italy*
[3]*Dipartimento di Matematica, Università di Pisa, Largo Bruno Pontecorvo 5, 56127 Pisa, Italy*

(Received 17 April 2018; accepted 23 July 2018; published online 16 August 2018)

It is well-known that a strict analogue of the Birkhoff Ergodic Theorem in infinite ergodic theory is trivial; it states that for any infinite-measure-preserving ergodic system, the Birkhoff average of every integrable function is almost everywhere zero. Nor does a different rescaling of the Birkhoff sum that leads to a non-degenerate pointwise limit exist. In this paper, we give a version of Birkhoff's theorem for conservative, ergodic, infinite-measure-preserving dynamical systems where instead of integrable functions we use certain elements of L^∞, which we generically call global observables. Our main theorem applies to general systems but requires a hypothesis of "approximate partial averaging" on the observables. The idea behind the result, however, applies to more general situations, as we show with an example. Finally, by means of counterexamples and numerical simulations, we discuss the question of finding the optimal class of observables for which a Birkhoff theorem holds for infinite-measure-preserving systems. *Published by AIP Publishing.* https://doi.org/10.1063/1.5036652

Birkhoff's Ergodic Theorem is a cornerstone of the theory of dynamical systems. It states that for a dynamical system endowed with a finite invariant measure, the time, or Birkhoff, average of an integrable function exists almost everywhere. For an ergodic system, this is equivalent to the Strong Law of Large Numbers for the evolution of any integrable function. When the invariant measure is infinite, which is the case, for example, for most unbounded or extended Hamiltonian systems, Birkhoff's theorem is no longer significant, in the sense that at least for ergodic systems, the Birkhoff average of any integrable observable is almost everywhere zero. However, for a dynamical system preserving an infinite measure, the integrable functions are not the only observables of interest. For example, for an extended Hamiltonian system, the kinetic and potential energies and many other "delocalized" observables are not integrable. In this article, we make steps towards a formulation of Birkhoff's Ergodic Theorem for global observables in infinite-measure-preserving systems. A global observable is, in essence, a bounded function that is significantly different from zero throughout the space.

I. INTRODUCTION

Birkhoff's Ergodic Theorem is one of the cornerstones of probability theory and the theory of dynamical systems. It states that if T is a measure-preserving transformation of a probability space (X, μ) and $f \in L^1(X, \mu)$, the *Birkhoff average*

$$\mathcal{A}f(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} f \circ T^k(x) \tag{1.1}$$

exists for μ-a.e. $x \in X$. If T is also ergodic, the theorem states in addition that

$$\mathcal{A}f(x) = \int_X f \, d\mu, \tag{1.2}$$

for a.e. x. Here, and in the rest of the paper, we use the most general definition of ergodicity, which is valid for both finite and infinite ergodic theory: The map T is said to be ergodic if every invariant set B (this means that $T^{-1}B = B \bmod \mu$, where mod μ indicates that these sets are equal up to a μ-null set of points) has zero measure or full measure (so either $\mu(B) = 0$ or $\mu(X \setminus B) = 0$). Thus, for a probability-preserving system, ergodicity corresponds to the Strong Law of Large Numbers for the variables $f \circ T^n$, for all $f \in L^1$.

Let us now consider the case where (X, μ) is an infinite measure space. More precisely, let us assume that μ is a σ-finite infinite measure, which means that $\mu(X) = \infty$ and X can be written as $X = \bigcup_{j\in\mathbb{N}} X_j$, with each $\mu(X_j) < \infty$. For an ergodic T, the strict analogue of Birkhoff's theorem is trivial: For all $f \in L^1$, $\mathcal{A}f(x) = 0$ almost everywhere. (This is an easy consequence of Hopf's Ergodic Theorem, which we will recall momentarily.) One is thus led to ask what the growth rate is for the *Birkhoff sum*

$$\mathcal{S}_n f := \sum_{k=0}^{n-1} f \circ T^k, \tag{1.3}$$

of an integrable function f. Aaronson discovered that there is no universal growth rate. More precisely, if T is conservative (in other words, Poincaré recurrence holds (Ref. 1, Sec. 1.1)) and ergodic, given any sequence $(a_n)_{n\in\mathbb{N}}$ of positive numbers, one of the following two cases occurs (Ref. 1, Thm. 2.4.2):

1. For all $f \in L^1$ with $f > 0$, $\liminf_{n\to\infty} \frac{\mathcal{S}_n f(x)}{a_n} = 0$, for a.e. $x \in X$.

a)E-mail: marco.lenci@unibo.it
b)E-mail: sara.munday@dm.unipi.it

2. There exists a strictly increasing sequence $(n_k)_{k\in\mathbb{N}}$ of the natural numbers such that for all $f \in L^1$ with $f > 0$, $\lim_{k\to\infty} \frac{\mathcal{S}_{n_k} f(x)}{a_{n_k}} = \infty$, for a.e. $x \in X$.

Notice in the second case that the sequence (n_k) is the same for all f and for all x, and thus the assertion is stronger than the statement: For every $f \in L^1$ with $f > 0$, $\limsup_{n\to\infty} \mathcal{S}_n f / a_n = \infty$ almost everywhere.

The lack of a universal growth rate for $(\mathcal{S}_n f(x))_{n\in\mathbb{N}}$ is not due to its dependence on f but on x. In fact, for an ergodic T, Hopf's Ergodic Theorem[10,20] states that, for all $f, g \in L^1$ with $g > 0$,

$$\lim_{n\to\infty} \frac{\mathcal{S}_n f(x)}{\mathcal{S}_n g(x)} = \frac{\int_X f \, d\mu}{\int_X g \, d\mu}, \tag{1.4}$$

for a.e. x. Thus, if we choose a function $g \in L^1$ with $g > 0$ and $\int g \, d\mu = 1$, and set $a_n(x) := \mathcal{S}_n g(x)$, we indeed have that $\mathcal{S}_n f(x)/a_n(x) \to \int f d\mu$ almost everywhere, for all $f \in L^1$. But the variability of $x \mapsto (a_n(x))_{n\in\mathbb{N}}$ is so strong that only a zero-measure set of points produces the same rate.

In this paper, functions $f : X \longrightarrow \mathbb{C}$ are supposed to represent observations about the state of the system $x \in X$. Accordingly, they will be called *observables*. In particular, all functions $f \in L^1(X, \mu)$ will be called *local observables*. The name is due to the fact that they are well approximated by functions with a finite-measure support within an infinite-measure ambient space.

The results presented above lead one to think that local observables are not the right ones to average along the orbits of T. Even when a scaling sequence exists such that $\mathcal{S}_n f / a_n$ converges to a non-degenerate limit, cf. the Darling-Kac Theorem (Ref. 1, Sec. 3.6), this convergence is in distribution (more precisely, *strongly in distribution*) and the limit is a non-constant random variable. In any case, the limit cannot reasonably be called the average of f along the orbit s of T. A natural concept of average presupposes that if, for example, $f \equiv c$, its average is c. In other words, we are interested in bona fide Birkhoff averages, as in (1.1).

So, we need to change the class of observables. The simplest class beyond $L^1(X, \mu)$ that one might think to consider is $L^\infty(X, \mu)$, which does include the constant functions. However, the whole of L^∞ is, vaguely speaking, "too big" for us to expect constant Birkhoff averages for all of its elements. An interesting class of counterexamples is given by the indicator functions of infinite-measure sets with the property that orbits spend long stretches of time there before leaving; for example, neighborhoods of strongly neutral indifferent fixed points. The Birkhoff averages of these observables converge, strongly in distribution, to non-constant random variables. A classical example of this phenomenon is the arcsine law for the Boole transformation.[24] We will return to this example, along with others, in Sec. IV.

In this paper, we call *global observables* all essentially bounded functions for which, in principle, a Birkhoff Theorem could hold. Of course, depending on the system at hand, the Birkhoff Theorem will hold as well for many non-integrable, non-essentially bounded observables. Nonetheless, here we limit ourselves to subspaces of L^∞, for two reasons. First, as already discussed, L^∞ already contains "too many" observables. Second, we want to follow the approach of Lenci on the question of mixing for infinite-measure-preserving dynamical systems,[15,16] whereby global observables are taken from subspaces of L^∞. (This is an important assumption there because the theory exploits the duality between L^1 and L^∞.) Another observation to make is that with the vague "definition" given above, it is impossible to pre-determine the space of global observables. We do in fact expect it to depend significantly on the given system. Nevertheless, the common underlying concept can be expressed like this: a global observable is a function which is supported more or less all over the infinite-measure space and which measures a quantity that is roughly homogeneous in space. For example, if the reference space is $(\mathbb{R}^d, m)$, where m is the Lebesgue measure, all periodic or quasi-periodic bounded functions are in principle global observables. Another example is the case where $T : [0, 1] \longrightarrow [0, 1]$ is an expanding map with an indifferent fixed point at 0 and preserves an absolutely continuous measure that is non-integrable around 0. Then, all bounded functions which have a limit at 0 or oscillate in a controlled way in its neighborhood are candidates for global observables.

The main result of this paper, which we present in Sec. II, is an analogue of the Birkhoff Theorem for certain global observables relative to a conservative, ergodic, infinite-measure-preserving dynamical system (X, μ, T). The hypotheses of the theorem are formulated in terms of the partition of X determined by the hitting times to a set L_0. This partition is a very natural construction; for systems isomorphic to a Kakutani tower, which includes all invertible maps (Ref. 1, Sec. 1.5), it corresponds to the levels of the tower. In the Appendix, we recall the definition and basic properties of Kakutani towers. Returning to Sec. II, we also describe several concrete examples of systems and observables for which our results hold.

As our main theorem is certainly not optimal, we further discuss its core ideas and limitations. First, in Sec. III, we give an example of a family of dynamical systems—which happen to be conjugates of α-Farey maps[12]—and a family of global observables which do not satisfy the hypotheses of the theorem, but for which we are nevertheless able to prove that the Birkhoff average is almost everywhere constant. This is done using the same ideas as in the proof of the theorem, but the techniques are rather more complicated and specific to that case. Finally, in Sec. IV, we briefly recall the known examples mentioned above of L^∞ functions whose Birkhoff average does not converge pointwise, and we construct other examples of a similar nature which are interesting because they are representations of Lévy walks (see Refs. 26, 9, and 18 and references therein), thus highlighting the connections between infinite ergodic theory and anomalous stochastic processes. In light of the vague definition given above, these functions cannot really be considered counterexamples to our theorem. So, we also present numerical simulations of the Birkhoff averages for the observables and the systems discussed in Secs. II and III.

Let us also remark here that there is a related strand of research in which finite-measure spaces with non-L^1 observables are investigated. For instance, in the 1990s, first Major[19] and then Buczolich[6] constructed specific examples where two different finite-measure systems assign almost everywhere a different constant value to the limit of the Birkhoff averages for the same non-L^1 observable. That is, the limit $\mathcal{A}f$ exists almost everywhere and is constant for each of the two systems, but it is not equal to the integral of f (see also the survey article[7] for related references). More recently, Carney and Nicol[8] investigate growth rates of Birkhoff sums of non-integrable observables. Also, the effect on the strong law of large numbers of "trimming" the largest value(s) from the sums has been studied first by Aaronson and Nakada[2] and then by Kesseböhmer and Schindler.[13]

II. SETUP AND MAIN THEOREM

For the rest of this paper, we will indicate a dynamical system by means of a triple (X, μ, T), where X is a measurable space, μ a measure on it, and $T : X \longrightarrow X$ a measurable map. We shall always assume that (X, μ) is a σ-finite measure space. Strictly speaking, we should also mention the σ-algebra $\mathscr{A}$ of all measurable sets of X, but, as we only deal with one σ-algebra, we shall take it as understood. Unless otherwise stated, we shall always assume that T preserves μ, meaning that, for all measurable $A \subseteq X$, $\mu(T^{-1}A) = \mu(A)$. The measure μ can be infinite, that is, $\mu(X) = \infty$, or finite, in which case we assume it to be normalized, that is, $\mu(X) = 1$. Here, we are particularly interested in the first case.

So, given a dynamical system (X, μ, T), an observable $f : X \longrightarrow \mathbb{C}$, a positive integer n, and a point $x \in X$, we denote

$$\mathcal{S}_n f(x) := \sum_{k=0}^{n-1} f \circ T^k(x), \tag{2.1}$$

$$\mathcal{A}_n f(x) := \frac{\mathcal{S}_n f(x)}{n}, \tag{2.2}$$

$$\mathcal{A}f(x) := \lim_{n\to\infty} \mathcal{A}_n f(x), \tag{2.3}$$

whenever the limit exists.

Our goal is to find conditions under which $\mathcal{A}f(x)$ exists and is constant almost everywhere. The easiest such condition is perhaps that f is a *coboundary*, as in the next proposition, whose proof is trivial.

Proposition 2.1. *For a dynamical system as described above, let $f = g - g \circ T^k$, with $g \in L^\infty(X, \mu)$ and $k \in \mathbb{Z}^+$. Then, $\mathcal{A}f(x) = 0$ μ-almost everywhere.*

Another simple condition was already mentioned in the introduction. We repeat it here for completeness.

Proposition 2.2. *Suppose that (X, μ, T) is an infinite-measure-preserving ergodic dynamical system and $f \in L^1(X, \mu)$. Then, $\mathcal{A}f(x) = 0$ almost everywhere.*

Corollary 2.3. *If the observable f is such that $f - f^* \in L^1$ for some $f^* \in \mathbb{C}$, then $\mathcal{A}f(x) = f^*$ almost everywhere.*

Corollary 2.3 applies to a large number of observables which converge to a constant "at infinity." For this phrase to make sense, a topology and a notion of infinity must be defined on X. However, this fact is very general and can be stated in a purely measure-theoretic fashion, as in the following proposition.

Proposition 2.4. *For an infinite-measure-preserving ergodic system (X, μ, T), suppose that $f \in L^\infty(X, \mu)$ admits $f^* \in \mathbb{C}$ with the following property: For all $\varepsilon > 0$, there exists a finite-measure set A_ε such that $|f(x) - f^*| \le \varepsilon$ for every $x \in X \setminus A_\varepsilon$. Then, for μ-a.e. $x \in X$, $\mathcal{A}f(x) = f^*$.*

Proof. Without loss of generality, suppose that $f^* = 0$ (since if not, we can always consider the function $g := f - f^*$).

For $\varepsilon > 0$, define the observable $f_\varepsilon := f\, 1_{X\setminus A_\varepsilon}$, where 1_A denotes the indicator function of the set A. By hypothesis, $\|f_\varepsilon\|_\infty \le \varepsilon$, so it follows that $\|\mathcal{A}_n f_\varepsilon\|_\infty \le \varepsilon$ for all $n \in \mathbb{N}$. Consider now the function $f - f_\varepsilon = f\, 1_{A_\varepsilon} \in L^1(X, \mu)$. By Proposition 2.2, there exists a full-measure set B_ε such that, for all $x \in B_\varepsilon$,

$$\lim_{n\to\infty} (\mathcal{A}_n f(x) - \mathcal{A}_n f_\varepsilon(x)) = 0, \tag{2.4}$$

whence

$$\limsup_{n\to\infty} |\mathcal{A}_n f(x)| \le \varepsilon. \tag{2.5}$$

If we choose a sequence $\varepsilon_i \to 0$, we conclude that, for every $x \in \bigcap_i B_{\varepsilon_i}$, $\mathcal{A}f(x) = 0$. □

Proposition 2.4 is extremely general and, for that reason, also rather weak, because it works with observables that are almost constant on the overwhelming largest part of the space X. Our main theorem, which we state after giving an *ad hoc* construction, is a stronger result that effectively uses the dynamics of T.

Assume that T is conservative and ergodic. Given a set L_0 with $0 < \mu(L_0) < \infty$, we have that

$$\bigcup_{k\ge 0} T^{-k} L_0 = X \bmod \mu, \tag{2.6}$$

that is, L_0 is a sweep-out set. If we recursively define, for $k \ge 1$,

$$L_k := \left(T^{-1} L_{k-1}\right) \setminus L_0, \tag{2.7}$$

we see that $\{L_k\}_{k\in\mathbb{N}}$ forms a partition of X mod μ. (We use the convention whereby $0 \in \mathbb{N}$.) By construction, L_k is the set of points whose orbit intersects L_0 for the first time at the kth iteration. In other words, $\{L_k\}$ is the partition which consists of the level sets of the hitting time to L_0. By (2.7), $\mu(L_k) \le \mu(L_{k-1})$. Also, by conservativity, $\mu(L_k) \to 0$ as $k \to \infty$. Finally, $\sum_k \mu(L_k) = \mu(X) = \infty$.

Observe that if (X, μ, T) is isomorphic to a Kakutani tower and L_0 corresponds, via the isomorphism, to the base of the tower, then L_k corresponds to the kth level of the tower for all k. In particular, the above construction has a clear interpretation in the case of an invertible T. We refer to the Appendix for the definition of a Kakutani tower and results linking Kakutani towers to Theorem 2.5.

Theorem 2.5. *Let (X, μ, T) be an infinite-measure-preserving, conservative, ergodic dynamical system, endowed with the partition $\{L_k\}_{k\in\mathbb{N}}$, as described above. Let $f \in L^\infty(X, \mu)$ admit $f^* \in \mathbb{C}$ with the following property: $\forall\, \varepsilon > 0$, $\exists N, K \in \mathbb{N}$ such that $\forall\, x \in \bigcup_{k\ge K} L_k$,*

$$\left|\mathcal{A}_N f(x) - f^*\right| \le \varepsilon.$$

Then, for μ-a.e. $x \in X$, $\mathcal{A}f(x) = f^*$.

Proof. Once again, it is enough to prove the theorem in the case $f^* = 0$. Also, without loss of generality, we may assume that

$$\frac{N\|f\|_\infty}{K} \le \varepsilon, \tag{2.8}$$

otherwise, we can always take a larger K in the main hypothesis of the theorem. Define the sets

$$X_K := \bigcup_{k\ge K} L_k, \tag{2.9}$$

$$X_K^c := X \setminus X_K, \tag{2.10}$$

and the observable $f_\varepsilon := f\, 1_{X_K}$. The hypotheses on T imply that the orbit of μ-a.e. $x \in X$ must enter X_K infinitely often. Let us fix one such x and split its orbit into blocks which are subsets, alternately, of X_K and X_K^c. So, let $m_0 = m_0(x)$ denote the first time where $T^{m_0}(x) \in X_K$, and note that m_0 can be equal to zero (in which case x is already in X_K). In other words, $T^{m_0}(x) \in X_K$ and $T^k(x) \in X_K^c$ for all $k < m_0$. Denote $k_1 = k_1(x) \ge K$ the unique integer such that $T^{m_0}(x) \in L_{k_1}$. Set $n_1 := k_1 - K$ so that $T^{m_0+n_1-1}(x) \in L_K$. Now, let $m_1 = m_1(x)$ denote the length of the following excursion in X_K^c so that $T^{m_0+n_1+m_1-1}(x) \in L_0$ and the next orbit point jumps back to the set X_K, say, to the set L_{k_2}, for some $k_2 = k_2(x) \ge K$. Again, set $n_2 := k_2 - K$. Continuing in this way, we construct two sequences $(m_j)_{j\ge0}$ and $(n_j)_{j\ge1}$, where setting $M_j := m_0 + m_1 + \cdots + m_j$ and $N_j := n_1 + n_2 + \cdots + n_j$, we have that

- $m_0 \ge 0$ and $m_j \ge K$ for all $j \ge 1$.
- $T^{M_j+N_j-1}(x) \in L_0$ and $T^{M_j+N_j}(x) \in L_{k_{j+1}} \subset X_K$ for all $j \ge 1$.
- $f_\varepsilon(T^{M_j+N_{j+1}+i}(x)) = 0$ for all $j \ge 0$ and $0 \le i < m_{j+1}$.

Now fix $n \ge m_0 + n_1 + m_1$ for which there exists $j \in \mathbb{N}$ such that $n = M_j + N_j + i$, for $0 \le i < n_{j+1}$. In other words, we consider all sufficiently large $n \in \mathbb{N}$ which correspond to stopping the orbit of x during an excursion in X_K. We will treat the other values of n later. The only parts of the orbit that contribute in a non-zero way to the Birkhoff sum $\mathcal{S}_n f_\varepsilon(x)$ are the excursions in X_K, that is, the blocks of lengths $n_1, \ldots, n_j$ and i. Each of these $j+1$ blocks can be further decomposed into p_i sub-blocks of length N and a remainder sub-block of length $0 \le r_i < N$. By hypothesis, the Birkhoff sum corresponding to each sub-block, except for the remainder sub-blocks, is bounded in modulus by ε. The contribution of each remainder sub-block is instead bounded by $N\|f_\varepsilon\|_\infty \le N\|f\|_\infty$. Putting all these observations together, we have that

$$\begin{aligned} |\mathcal{A}_n f_\varepsilon(x)| &< \frac{1}{M_j+N_j+i}\left(\varepsilon \sum_{i=1}^{j+1} p_i + (j+1)N\|f\|_\infty\right) \\ &< \varepsilon \frac{\sum_{i=1}^{j+1} p_i}{M_j+N_j+i} + \frac{(j+1)N\|f\|_\infty}{jK} \\ &< \varepsilon + 2\varepsilon = 3\varepsilon, \end{aligned} \tag{2.11}$$

having used, among other arguments, (2.8) and the fact that $j \ge 1$.

Finally, consider those $n \ge m_0 + n_1 + m_1$ which correspond to stopping the orbit of x during an excursion in X_K^c. Say that $n = M_j + N_{j+1} + i$ for some $j \ge 1$ and $0 \le i < m_{j+1}$. The contribution to the Birkhoff sum of the last excursion in X_K^c is null so, in light of (2.11),

$$|\mathcal{A}_n f_\varepsilon(x)| = |\mathcal{A}_{M_j+N_{j+1}+i} f_\varepsilon(x)| \le |\mathcal{A}_{M_j+N_{j+1}} f_\varepsilon(x)| < 3\varepsilon. \tag{2.12}$$

In conclusion, $|\mathcal{A}_n f_\varepsilon(x)| < 3\varepsilon$, for *all* sufficiently large n, depending on x. Since $f - f_\varepsilon = f 1_{X_K^c} \in L^1$, the proof of Theorem 2.5 is completed in the same way as the proof of Proposition 2.4, cf. (2.5) *et seq*. □

Remark 2.6. Let us observe here that Proposition 2.4 can be thought of as a sub-case of Theorem 2.5 with $N = 1$. Indeed, bearing in mind the definition of the partition $\{L_k\}$, it is easy to see that the condition in Proposition 2.4 can be reformulated as follows: For all $\varepsilon > 0$, there exists $K \in \mathbb{N}$ such that for all $x \in \bigcup_{k\ge K} L_k$, $|f(x) - f^*| \le \varepsilon$. Here, $\bigcup_{k\ge K} L_k$ plays the role of the set A_ε and this condition is exactly that of Theorem 2.5 with $N = 1$.

Let us now see some examples of dynamical systems and observables to which our results can be applied. As already mentioned, Proposition 2.4 is quite general. Consider, for instance, a map $T : \mathbb{R} \longrightarrow \mathbb{R}$ which preserves an ergodic infinite, locally finite measure μ. A nice example of such a map is Boole's transformation $T(x) := x - 1/x$, which was shown by Boole in 1857[5] to preserve the Lebesgue measure on $\mathbb{R}$ and by Adler and Weiss more than a century later[3] to be ergodic. Other interesting examples are the quasi-lifts and finite modifications thereof studied in Ref. 17. For all the systems we have mentioned, every bounded $f : \mathbb{R} \longrightarrow \mathbb{C}$ such that

$$f^* := \lim_{|x|\to\infty} f(x), \tag{2.13}$$

exists verifies the hypotheses of Proposition 2.4, and therefore $\mathcal{A}f = f^*$ almost everywhere.

Consider now a piecewise-smooth, full-branched, expanding map $T : [0,1] \longrightarrow [0,1]$ of the type shown in Fig. 1. If 0 is a *strongly neutral* fixed point (which means that T'' is regular in a neighborhood of 0), it is known[22] that under general conditions T preserves an absolutely continuous infinite measure μ such that $\mu([a,1]) < \infty$, for all $0 < a \le 1$. Also, T is ergodic w.r.t. μ.[23] It is easy to verify that the sets L_k are those marked in Fig. 1. Therefore, Proposition 2.4 applies to all bounded $f : [0,1] \longrightarrow \mathbb{C}$ such that

$$f^* := \lim_{x\to0^+} f(x), \tag{2.14}$$

exists.

Remaining in the case of the map $T : [0,1] \longrightarrow [0,1]$ described above, let us now introduce a class of non-trivial global observables which verify the hypothesis of Theorem 2.5. Given $N \in \mathbb{Z}^+$ and $c_0, c_1, \ldots, c_{N-1} \in \mathbb{C}$, let $f : [0,1] \longrightarrow \mathbb{C}$ be defined by

$$f(x) = c_j \quad \Longleftrightarrow \quad x \in L_k \quad \text{with } k \cong j \ (\mathrm{mod}\ N). \tag{2.15}$$

In other words, f is a step function on the partition $\{L_k\}$, which is N-periodic in the index k. The stochastic properties of these observables have been studied in Ref. 4 (Sec. 3.1). It is easy to see that for all $x \in L_k$, with $k \ge N-1$,

$$\mathcal{A}_N f(x) = \frac{1}{N}\sum_{k=0}^{N-1} c_k =: f^*, \tag{2.16}$$

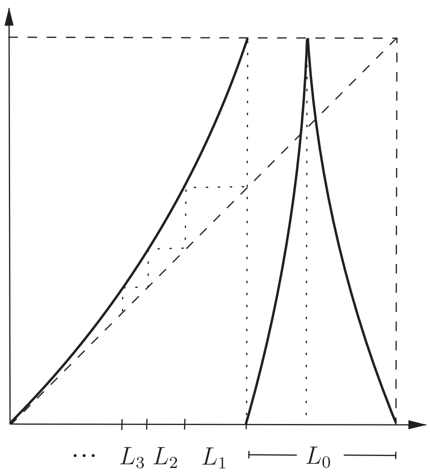

FIG. 1. A piecewise-smooth, full-branched expanding map of the unit interval.

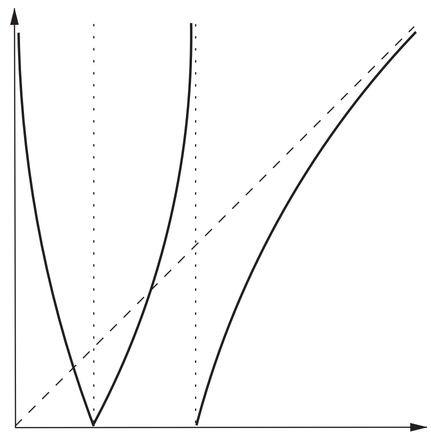

FIG. 2. A piecewise-smooth, full-branched expanding map of the half-line. This example corresponds to the example of Fig. 1 via the conjugation procedure explained in the body of the paper.

Therefore, Theorem 2.5 applies with N and $K \geq N-1$, independent of ε. Thus, $\mathcal{A}f = f^*$ almost everywhere.

It is easy to extend the above idea to a class of step functions on $\{L_k\}$ which are not periodic in k. For example, take a sequence $(c_k)_{k\in\mathbb{N}}$ of complex numbers and a number f^* such that, for every $\varepsilon > 0$, there exists $N \in \mathbb{Z}^+$ with the property that

$$\left| \frac{1}{N} \sum_{k=j}^{j+N-1} c_k - f^* \right| \leq \varepsilon, \tag{2.17}$$

for every $j \in \mathbb{N}$. Examples include quasi-periodic sequences $c_k := e^{2\pi i\alpha k}$ and many others. Then, let f be defined by $f|_{L_j} \equiv c_k$. This observable satisfies the hypothesis of Theorem 2.5 (again with $K \geq N-1$) by construction.

Looking for more general examples, let us consider a Kakutani tower $\mathcal{T} : Y \longrightarrow Y$, as defined in (A1)–(A4). We also choose L_0 to be the base of the tower $\Sigma \times \{0\}$. As explained in the Appendix, this implies that $L_k = \{(x,k)\varphi(x) \geq k\}$, that is, L_k is the kth level of the tower. It is not difficult to find global observables f which have the approximate averaging property required by Theorem 2.5 but with different values of f for different excursions outside of L_0. Take, for instance,

$$f(x,n) := e^{2\pi i(\omega(x)n+\gamma(x))}, \tag{2.18}$$

defined for $(x,n) \in Y$, where ω and γ are measurable real-valued functions of Σ. Assume that there exists $\delta \in (0,1)$ such that for all $x \in \Sigma$, $\delta \leq \omega(x) \leq 1-\delta$. This implies, for some $c = c(\delta) > 0$ and for all $x \in \Sigma$, that $|1 - e^{-2\pi i\omega(x)}| \geq c$.

This observable also satisfies the hypothesis of Theorem 2.5. In fact, for any $\varepsilon > 0$, select $N \geq 2/c\varepsilon$ and $K \geq N-1$. A point $(x,n) \in \bigcup_{k\geq K} L_k$ is one for which $n \geq K$. We have

$$|\mathcal{A}_N f(x,n)| = \frac{1}{N} \left| \sum_{k=0}^{N-1} e^{-2\pi i\omega(x)k} \right| \leq \frac{2}{Nc} \leq \varepsilon. \tag{2.19}$$

Thus, $\mathcal{A}f = 0$ almost everywhere.

As already alluded to in the Introduction, it is hard to determine *a priori* the maximal class of global observables for a given dynamical system. However, for certain systems, it is rather easy to agree on functions which ought to be considered global observables; for example, systems defined on Euclidean spaces (or large portions thereof) which preserve the Lebesgue measure. In this case, the translation-invariance of the reference measure suggests that at least all periodic and quasi-periodic bounded functions should be global observables.

Let us therefore consider an interesting class of Lebesgue-measure-preserving dynamical systems in Euclidean space: piecewise-smooth, expanding maps $T : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$, with full branches and an indifferent fixed point at $+\infty$, as in Fig. 2. These maps are of the same nature as the interval maps discussed earlier, cf. Ref. 4. Indeed, if T_o denotes a piecewise-smooth, full-branched, expanding map of $[0,1]$ onto itself, and μ denotes its infinite absolutely continuous invariant measure, then $\Phi(x) := \mu([x,1])$ defines a bijection $(0,1) \longrightarrow \mathbb{R}^+$ such that $T := \Phi \circ T_o \circ \Phi^{-1}$ is a piecewise-continuous, full-branched map $\mathbb{R}^+ \longrightarrow \mathbb{R}^+$ preserving the Lebesgue measure on $\mathbb{R}^+$. In many cases, T is also piecewise-smooth and expanding.

One notable example, which we will return to in Sec. IV, is the Farey map. This map is usually defined as a map on the unit interval, as follows:

$$F(x) := \begin{cases} \dfrac{x}{1-x} & \text{for } x \in [0, 1/2] \\ \dfrac{1-x}{x} & \text{for } x \in (1/2, 1]. \end{cases} \tag{2.20}$$

Up to factors, F has a unique Lebesgue-absolutely continuous invariant measure μ, which is given by the density $d\mu/dm(x) = 1/x$. Thus, here the function Φ is given by

$$\Phi(x) = \mu([x,1]) = \int_x^1 \frac{1}{\xi}\, d\xi = -\ln x. \tag{2.21}$$

The version of the Farey map transported to the positive real line is then given by

$$T_F(x) := -\ln(F(e^{-x})) = |\ln(e^x - 1)|. \tag{2.22}$$

Let us return to the general case of a map $T : \mathbb{R}^+ \longrightarrow \mathbb{R}^+$ preserving the Lebesgue measure. The considerations made above suggest that the first examples of global observables one should study are the functions $f(x) := e^{2\pi i\omega x}$, with $\omega \in$

$\mathbb{R} \setminus \{0\}$. Any reasonable notion of average[4,15] for these functions would suggest that $f^* = 0$. So, the problem is to show that, for a.e. $x \in \mathbb{R}^+$, $\mathcal{A}f(x) = 0$.

III. PERIODIC OBSERVABLES AND THE α-FAREY MAPS

In this section, we present non-trivial examples of piecewise-smooth, Lebesgue-measure-preserving, expanding maps on $\mathbb{R}^+$, with an indifferent fixed point at $+\infty$, for which the Birkhoff average of $f(x) := e^{2\pi i\omega x}$, $\omega \neq 0$, is almost everywhere zero. This will require more work than a simple application of Theorem 2.5, but the underlying ideas are the same.

Our maps will be conjugates, over the space $\mathbb{R}^+$, of the well-known α-Farey maps on $[0, 1]$ and will be obtained by means of the construction explained at the end of Sec. II.

Let us recall the definition of an α-Farey map, as introduced in Ref. 12 (Sec. 1.4). Start with a decreasing sequence $(t_k)_{k\in\mathbb{Z}^+}$ of real numbers such that $t_1 = 1$ and $\lim_{k\to\infty} t_n = 0$. This sequence allows us to define a partition

$$\alpha := \{A_k := (t_{k+1}, t_k] | k \geq 1\}, \tag{3.1}$$

of $(0, 1]$. We will write $a_k := m(A_k) = t_k - t_{k+1}$ for the Lebesgue measure of the kth partition element. Then, the map $F_\alpha : [0, 1] \longrightarrow [0, 1]$ is defined by setting

$$F_\alpha(x) := \begin{cases} (1-x)/a_1 & \text{for } x \in A_1 \\ a_{k-1}(x - t_{k+1})/a_k + t_k & \text{for } x \in A_k, \text{ for } k \geq 2 \\ 0 & \text{for } x = 0. \end{cases} \tag{3.2}$$

The map F_α preserves the (unique up to factors) Lebesgue-absolutely continuous measure μ_α given by the density

$$h_\alpha := \frac{d\mu_\alpha}{dm} = \sum_{k=1}^{\infty} \frac{t_k}{a_k} 1_{A_k}, \tag{3.3}$$

and the measure is infinite if and only if $\sum_k t_k = \infty$.

For later use, let us also recall the definition of the related α-Lüroth expansion (for more details, we refer again to Ref. 12 (Sec. 1.4)). Each partition α generates a series expansion of the numbers in the unit interval, in that we can associate to each x a sequence of positive integers $(\ell_i)_{i\geq 1}$ for which

$$\begin{aligned} x &= t_{\ell_1} + \sum_{k=2}^{\infty} (-1)^{k-1} \left(\prod_{i<k} a_{\ell_i} \right) t_{\ell_k} \\ &= t_{\ell_1} - a_{\ell_1} t_{\ell_2} + a_{\ell_1} a_{\ell_2} t_{\ell_3} - \cdots . \end{aligned} \tag{3.4}$$

To lighten the notation, we will write $x = [\ell_1, \ell_2, \ell_3, \ldots]_\alpha$. Observe that the map F_α acts on this expansion in the following way:

$$F_\alpha([\ell_1, \ell_2, \ell_3, \ldots]_\alpha) = \begin{cases} [\ell_1 - 1, \ell_2, \ell_3, \ldots]_\alpha & \text{for } \ell_1 \geq 2, \\ [\ell_2, \ell_3, \ldots]_\alpha & \text{for } \ell_1 = 1. \end{cases} \tag{3.5}$$

Throughout this section, we will restrict ourselves to the particular case $t_k := k^{-\beta}$, with $0 < \beta < 1/2$. The partition generated by this sequence will be denoted by $\alpha(\beta)$. In Ref. 12, it is referred to as an expansive partition with exponent β. Set

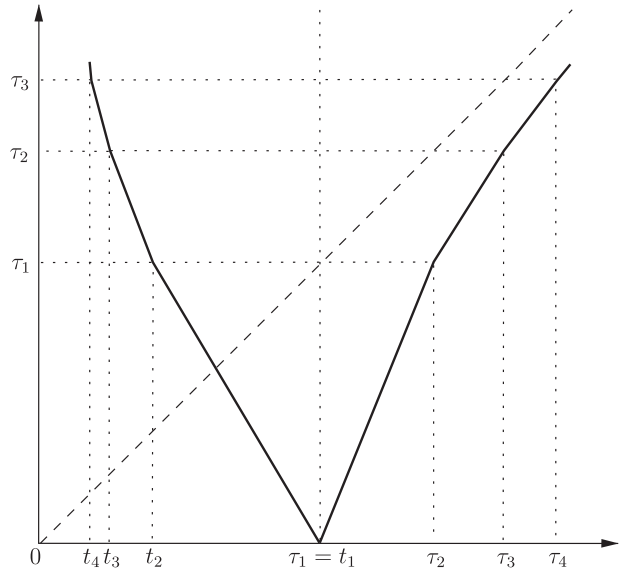

FIG. 3. The α-Farey map T_β on $\mathbb{R}^+$.

$$\tau_k := \sum_{j=1}^{k} t_j \sim \frac{k^{1-\beta}}{1-\beta}, \tag{3.6}$$

where we write $x_k \sim y_k$ to mean that $\lim_{k\to\infty}(x_k/y_k) = 1$.

As anticipated, we want to consider the map $T_\beta := \Phi \circ F_{\alpha(\beta)} \circ \Phi^{-1}$, where $\Phi(x) := \mu_{\alpha(\beta)}([x, 1])$ is defined for $x \in (0, 1)$. Then, T_β is a piecewise-continuous, full-branched, Lebesgue-preserving map on $\mathbb{R}^+$. In this case, Φ is a piecewise-linear function that maps, for each $k \geq 0$, the partition element A_{k+1} onto the interval $L_k := [\tau_k, \tau_{k+1})$. Setting $L_0 := [0, \tau_1)$, a series of straightforward calculations show that

$$T_\beta(x) = \begin{cases} \Phi(x) & \text{for } x \in L_0 \\ \dfrac{t_k}{t_{k+1}}(x - \tau_k) + \tau_{k-1} & \text{for } x \in L_k, \text{ with } k \geq 1. \end{cases} \tag{3.7}$$

See Fig. 3 for a picture of T_β. Note here that, by construction, the partition $\{L_k\}_{k\geq 0}$ matches for T_β the definition of the corresponding sequence of sets given for a general map T in Sec. II.

We are now in a position to state our main result of this section.

Proposition 3.1. *Let $f(x) := e^{2\pi i\omega x}$, with $\omega \in \mathbb{R} \setminus \{0\}$, and fix $\beta \in (0, 1/2)$. Then, for m-a.e. $x \in \mathbb{R}^+$,*

$$\mathcal{A}f(x) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} f \circ T_\beta^i(x) = 0.$$

Proof. First of all, let us assume without loss of generality that $\omega > 0$. Then, let us define a new observable g which is constructed from f in the following way: For each $j \geq 0$, denote by k_j the natural number such that $j/\omega \in L_{k_j}$. Then let

$$I_j := \bigcup_{i=k_j}^{k_{j+1}-1} L_i = \left[\tau_{k_j}, \tau_{k_{j+1}}\right), \tag{3.8}$$

$$\omega_j := \frac{1}{m(I_j)} = \frac{1}{\tau_{k_{j+1}} - \tau_{k_j}}. \tag{3.9}$$

Finally, for all $x \in I_j$, set $g(x) := e^{2\pi i \omega_j (x - \tau_{k_j})}$. □

Let us fix one more piece of notation that we shall use throughout the proof. For any observable ϕ, we shall call any interval $[a, b]$ with the property that $\phi(x) = e^{2\pi i (x-a)/(b-a)}$, for $a \le x \le b$, a *wavelength* for ϕ. Therefore, g is a modification of our original observable f so that the wavelengths of g are unions of intervals L_k. Note that, since $m(L_k) \to 0$ when $k \to \infty$, the modification is smaller and smaller for larger and larger values of the argument x. In other words, $\omega_k \sim \omega$.

For an arbitrary $\varepsilon > 0$, choose $K = K(\varepsilon) \in \mathbb{N}$ sufficiently large that:

- $|f(x) - g(x)| \le \varepsilon$ for all $x \in X_K := \bigcup_{k \ge K} L_k$;
- $m(L_K) \le \varepsilon$, whence $m(L_k) \le \varepsilon$ for all $k \ge K$;
- $K^{\beta-1} \le \varepsilon$.

To simplify the argument below, let us also suppose that $K = k_{j_o}$ for some $j_o \in \mathbb{N}$. Moreover, K will satisfy another condition which we will state later, when it is needed.

Define the observable $g_\varepsilon := g\, 1_{X_K}$ and consider the portion of g_ε defined on I_j with $k_j \ge K$, cf. (3.8). Denote by $r_j := k_{j+1} - k_j$ the number of intervals L_k that make up the interval I_j (recall that g_ε is defined so as to have its wavelengths start and end exactly at the endpoints of these partition elements).

Consider now a point $x \in \mathbb{R}^+$ whose forward orbit intersects some L_k with $k \ge k_{j+1} - 1$. This is equivalent to asking that, at some time s, $T_\beta^s(x) \in L_{k_{j+1}-1}$. Therefore, $T_\beta^{s+1}(x) \in L_{k_{j+1}-2}$ and so on, until $T_\beta^{s+r_j-1}(x) \in L_{k_j}$. In other words, the intervals $\{L_{k_j+i}\}_{i=0}^{r_j-1}$ that partition I_j each contain exactly one orbit point of x, from time s to time $s + r_j - 1$. We want to compare these intervals to the intervals $\{B_{k_j+i}\}_{i=0}^{r_j-1}$, which are defined to be a partition of I_j into intervals of the same size, labeled from left to right.

We claim that, for $0 \le i < r_j$,

$$L_{k_j+i} \cap B_{k_j+i} \neq \varnothing. \tag{3.10}$$

Indeed, observe first that the common size of the intervals B_{k_j+i} is the average of the sizes of the intervals L_{k_j+i}; therefore, the "relative discrepancy" between the sizes of corresponding sets can be estimated as follows:

$$\frac{|m(L_{k_j+i}) - m(B_{k_j+i})|}{m(B_{k_j+i})} = \left| \frac{m(L_{k_j+i})}{m(B_{k_j+i})} - 1 \right| \le \frac{m(L_{k_j})}{m(L_{k_{j+1}})} - 1, \tag{3.11}$$

because $m(L_k)$ is a decreasing function of k. Clearly, the relative discrepancy between the sizes of $\bigcup_{i=0}^q L_{k_j+i}$ and $\bigcup_{i=0}^q B_{k_j+i}$, for $0 \le q < r_j$, does not exceed the sum of the individual discrepancies (3.11). Condition (3.10) will be satisfied if the former is always less than or equal to 1. A sufficient condition for this is

$$r_j \left(\frac{m(L_{k_j})}{m(L_{k_{j+1}})} - 1 \right) \le 1. \tag{3.12}$$

Towards the proof of (3.12), we make several observations. First, by construction, $m(L_k) = t_{k+1} \sim k^{-\beta}$. On the other hand, the definition of k_j and (3.6) imply that, as $j \to \infty$,

$$\frac{j}{\omega} \sim \tau_{k_j} \sim \frac{k_j^{1-\beta}}{1-\beta}, \tag{3.13}$$

whence

$$k_j \sim \left(\frac{1-\beta}{\omega} j \right)^{1/(1-\beta)}, \tag{3.14}$$

and

$$\frac{m(L_{k_j})}{m(L_{k_{j+1}})} \sim \left(1 + \frac{1}{j} \right)^{\beta/(1-\beta)} \sim 1 + \frac{\beta}{1-\beta} \frac{1}{j}. \tag{3.15}$$

Also, by (3.14),

$$r_j := k_{j+1} - k_j \sim c_1 j^{\beta/(1-\beta)}, \tag{3.16}$$

for some $c_1 = c_1(\beta, \omega) > 0$.

Putting these observations together, we obtain that, for some positive constant c_2,

$$r_j \left(\frac{m(L_{k_j})}{m(L_{k_{j+1}})} - 1 \right) \sim c_2 j^{(2\beta-1)/(1-\beta)}. \tag{3.17}$$

Since $0 < \beta < 1/2$, the above term vanishes as $j \to \infty$. If we choose j_o sufficiently large, that is, if we choose $K = k_{j_o}$ sufficiently large (which is the condition we anticipated earlier we would state precisely), we can guarantee that (3.12) holds for all $j \ge j_o$. This proves the claim (3.10).

Now, for $k \in \mathbb{N}$, denote by b_k the midpoint of the interval B_k. Recalling that $m(L_k) < \varepsilon$ for all $k \ge K$ and that $T_\beta^{s+i}(x) \in L_{k_{j+1}-1-i}$, for all $0 \le i < r_j$, it follows from (3.10) that, for the same values of i,

$$\left| T_\beta^{s+i}(x) - b_{k_{j+1}-1-i} \right| < 2\varepsilon. \tag{3.18}$$

Then, since g_ε is Lipschitz continuous on its wavelength I_j, with constant $2\pi\omega_j$, and we can find an upper bound $c_3 > 0$ such that $2\pi\omega_j \le c_3$ for all $j \ge j_o$, we have

$$\left| g_\varepsilon(T_\beta^{s+i}(x)) - g_\varepsilon(b_{k_{j+1}-1-i}) \right| < 2c_3\varepsilon. \tag{3.19}$$

On the other hand,

$$\sum_{i=0}^{r_j-1} g_\varepsilon(b_{k_{j+1}-1-i}) = 0, \tag{3.20}$$

because, for $k_j \le k < k_{j+1}$, the b_k are the midpoints of the uniform partition of I_j. Therefore,

$$\left| \frac{1}{r_j} \sum_{i=0}^{r_j-1} g_\varepsilon(T_\beta^{s+i}(x)) \right| < 2c_3\varepsilon. \tag{3.21}$$

In summary, if we have a section of orbit of a point x under T_β that travels through an entire wavelength of the function g_ε, then the partial Birkhoff average through this excursion can be at most $2c_3\varepsilon$.

Now, for $x \in \mathbb{R}^+$, let $(n_i)_{i\in\mathbb{Z}^+}$ denote the sequence of hitting times to L_0. More precisely, $x \in L_{n_1}$, $T_\beta^{n_1}(x) \in L_0$, $T_\beta^{n_1+1}(x) \in L_{n_2}$, and so on. (Note that these digits are closely related to the $\alpha(\beta)$-Lüroth digits of $\Phi^{-1}(x) \in [0,1]$, which are $[n_1+1, n_2+1, n_3+1, \ldots]_{\alpha(\beta)}$.)

We want first to consider $\mathcal{A}_n g_\varepsilon$, with $n = N_q := \sum_{i=1}^q n_i$. In other words, we want to estimate the Birkhoff average of g_ε along an entire number of excursions back to L_0. Let us suppose first of all that $n_i \geq K$ for all $i \geq 1$, since otherwise we would be adding only zeros for certain portions of the orbit. Consider the portion of the orbit that lies in the sets $L_{n_i}, L_{n_i-1}, \ldots, L_K$. This can be split into p_i+1 blocks, where the initial block runs through a portion (in general) of a wavelength of g_ε and the other p_i blocks run through complete wavelengths. Using the index $0 \leq u \leq p_i$, to denote these blocks, where $u=0$ refers to the first block, which corresponds to the partial wavelength, let $\rho_u^{(i)}$ be the number of intervals L_k in the uth block. In other words, if the uth block corresponds to the wavelength I_j, then $\rho_u^{(i)} = r_j$. Notice that, by (3.13) and (3.16), r_j is asymptotic to $k_j^\beta \leq n_i^\beta$. This shows, in particular, that $\rho_0^{(i)} \leq c_4 n_i^\beta$, for some constant $c_4 = c_4(\beta,\omega)$.

So, in the special case $n = N_q = \sum_{i=1}^q (\sum_{u=0}^{p_i} \rho_u^{(i)} + K)$, we obtain

$$\begin{aligned} |\mathcal{A}_n g_\varepsilon(x)| &< \frac{1}{N_q} \sum_{i=1}^q \left(c_4 n_i^\beta + 2c_3\varepsilon \sum_{u=1}^{p_i} \rho_u^{(i)} \right) \\ &< c_4 K^{\beta-1} \frac{1}{N_q} \sum_{i=1}^q n_i + 2c_3\varepsilon \\ &\leq (c_4 + 2c_3)\,\varepsilon =: c_5\varepsilon. \end{aligned} \tag{3.22}$$

Notice that in the second inequality, we have used the fact that $n_i^\beta = n_i^{\beta-1} n_i \leq K^{\beta-1} n_i$ and estimate (3.21). In the third inequality, we have used the assumption $K^{\beta-1} \leq \varepsilon$.

Let us now consider the case $N_q < n < N_{q+1}$, i.e., we consider the Birkhoff average of g_ε up to a point which is in the middle of an excursion back to L_0. In the portion of orbit between time N_q and time n, there could be up to two blocks (an initial and a final block) that are neither contained in $\bigcup_{k=0}^{K-1} L_k$ or form a full wavelength. The contribution to the Birkhoff sum from these blocks, which is of order at most n_{q+1}^β, cannot be compensated for as in (3.22), because the denominator $n > N_q$ might be much smaller than $N_q + n_{q+1} = N_{q+1}$. This is a phenomenon that occurs because the numbers (n_i) are distributed (in a sense better specified in the proof of Lemma 3.2) as the outcomes of a non-integrable random variable (more precisely, a random variable in the domain of attraction of a β-stable distribution). Hence, for some q, the number n_{q+1} might be very large compared to N_q.

We appeal instead to the following lemma, which we prove at the end of this section.

Lemma 3.2. *Let $x = [\ell_1, \ell_2, \ldots]_{\alpha(\beta)} \in [0,1]$ denote the $\alpha(\beta)$-Lüroth expansion of the point x, where $0 < \beta < 1$. Then, for Lebesgue-almost every $x \in [0,1]$,*

$$\lim_{n\to\infty} \frac{\ell_n^\beta}{\ell_1 + \cdots + \ell_{n-1}} = 0.$$

Since the conjugation $\Phi : (0,1) \longrightarrow \mathbb{R}^+$ is non-singular and since, for a general n, $|\mathcal{A}_n g_\varepsilon(x)|$ does not exceed constant times

$$\frac{\sum_{i=1}^{q+1} n_i^\beta}{\sum_{i=1}^q n_i} + 2c_3\varepsilon \frac{\sum_{i=1}^{q+1} \sum_{u=1}^{p_i} \rho_u^{(i)}}{\sum_{i=1}^{q+1} \sum_{u=0}^{p_i} \rho_u^{(i)}}, \tag{3.23}$$

we conclude that, in view of Lemma 3.2 and estimate (3.22),

$$\limsup_{n\to\infty} |\mathcal{A}_n g_\varepsilon(x)| \leq c_5\varepsilon, \tag{3.24}$$

for a.e. $x \in \mathbb{R}^+$. Defining $f_\varepsilon := f\, 1_{X_K}$ and recalling that, by the assumption on K, $\|f_\varepsilon - g_\varepsilon\|_\infty \leq \varepsilon$, we deduce that, for all $\varepsilon > 0$, there exists a Lebesgue-full-measure set $B_\varepsilon \subseteq \mathbb{R}^+$ such that, for every $x \in B_\varepsilon$, $\limsup_{n\to\infty} |\mathcal{A}_n f_\varepsilon(x)| \leq (c_5 + 1)\varepsilon$. Since $f - f_\varepsilon \in L^1(\mathbb{R}^+, m)$, the conclusion of Proposition 3.1 is achieved in the same way as that of Proposition 2.4. $\square$

Proof of Lemma 3.2. We first claim that to obtain the statement of the lemma it is enough to show the following: For all $\varepsilon > 0$,

$$\sum_{n=2}^\infty m\left(\ell_n^\beta \geq \varepsilon \sum_{i=1}^{n-1} \ell_i \right) < \infty. \tag{3.25}$$

Indeed, if this relation holds, by the Borel-Cantelli lemma we infer that

$$m\left(\left\{ x = [\ell_1, \ell_2, \ldots]_{\alpha(\beta)} \in [0,1] \,\middle|\, \ell_n^\beta \geq \varepsilon \sum_{i=1}^{n-1} \ell_i \text{ for infinitely many } n \in \mathbb{Z}^+ \right\} \right) = 0. \tag{3.26}$$

In other words, there exists $B_\varepsilon \subseteq [0,1]$, with $m(B_\varepsilon) = 1$, such that

$$\frac{\ell_n^\beta}{\sum_{i=1}^{n-1} \ell_i} \leq \varepsilon, \tag{3.27}$$

for all n larger than some $N = N(x,\varepsilon)$. Fix a vanishing sequence $(\varepsilon_i)_{i\in\mathbb{N}}$ and define $B := \bigcap_i B_{\varepsilon_i}$. Clearly, $m(B) = 1$ and, for all $x \in [\ell_1, \ell_2, \ldots]_{\alpha(\beta)} \in B$, the limit in the statement of the lemma holds true.

It thus remains to prove (3.25). We start by observing that for every α-Lüroth map, the digits (ℓ_i) are independent identically distributed random variables w.r.t. m. We then make the following sequence of observations:

$$\begin{aligned} &\sum_{n=2}^\infty m\left(\ell_n^\beta \geq \varepsilon \sum_{i=1}^{n-1} \ell_i \right) \\ &\quad = \sum_{n=2}^\infty \sum_{k=n-1}^\infty m\left(\left(\ell_n^\beta \geq \varepsilon k\right) \cap \left(\sum_{i=1}^{n-1} \ell_i = k \right) \right) \\ &\quad = \sum_{n=2}^\infty \sum_{k=n-1}^\infty m\left(\ell_n \geq (\varepsilon k)^{1/\beta}\right) m\left(\sum_{i=1}^{n-1} \ell_i = k \right) \\ &\quad = \sum_{k=1}^\infty \sum_{n=2}^{k+1} m\left(\ell_n \geq (\varepsilon k)^{1/\beta}\right) m\left(\sum_{i=1}^{n-1} \ell_i = k \right). \end{aligned} \tag{3.28}$$

Now, for $k \in \mathbb{Z}^+$, let us define

$$\mathcal{C}_k^{(\alpha(\beta))} := \left\{ x = [\ell_1, \ell_2, \ldots]_\alpha \middle| \exists n \text{ with } \sum_{i=1}^{n} \ell_i = k \right\}. \quad (3.29)$$

In the language of Ref. 12, this is a *sum-level set* for the partition $\alpha(\beta)$, or an $\alpha(\beta)$-sum-level set. Observe that, when such an n exists, clearly $n \le k$. So, in light of the fact that (ℓ_n) are i.i.d., we can then rewrite (3.28) as

$$\sum_{n=2}^{\infty} m\left(\ell_n^\beta \ge \varepsilon \sum_{i=1}^{n-1} \ell_i\right) = \sum_{k=1}^{\infty} m\left(\ell_1 \ge (\varepsilon k)^{1/\beta}\right) m\left(\mathcal{C}_k^{(\alpha(\beta))}\right). \quad (3.30)$$

We have already mentioned that $\alpha(\beta)$ is an expansive partition with exponent β, therefore, by Theorem 1(2)(ii) of Ref. 12,

$$m\left(\mathcal{C}_k^{(\alpha(\beta))}\right) \sim \frac{1}{\Gamma(2-\beta)\Gamma(\beta)} \left(\sum_{j=1}^{k} t_j\right)^{-1} \sim \frac{c}{k^{1-\beta}}, \quad (3.31)$$

where Γ denotes Euler's Gamma function and c is a positive constant (recall that $t_k := k^{-\beta}$). Observing also that

$$m\left(\ell_1 \ge (\varepsilon k)^{1/\beta}\right) = t_{\lceil (\varepsilon k)^{1/\beta} \rceil} = \left\lceil (\varepsilon k)^{1/\beta} \right\rceil^{-\beta} \sim (\varepsilon k)^{-1}, \quad (3.32)$$

we finally obtain that

$$\sum_{n=2}^{\infty} m\left(\ell_n^\beta \ge \varepsilon \sum_{i=1}^{n-1} \ell_i\right) \sim \frac{c}{\varepsilon} \sum_{k=1}^{\infty} \frac{1}{k^{2-\beta}} < \infty. \quad (3.33)$$

This gives (3.25) and concludes the proof of Lemma 3.2. □

Remark 3.3. Proposition 3.1 can be extended to include the case $\beta = 1/2$. In that case, in fact, the r.h.s. of (3.17) does not vanish as $j \to \infty$, but it is still bounded above. In other words, the inequality (3.12) holds with a bound possibly larger than 1 on the r.h.s. The pivotal relation (3.10) may not hold anymore, but one can still claim that the distance between the intervals L_{k_j+i} and B_{k_j+i} is a bounded multiple of $m(B_{k_j+i})$, which tends to zero as $j \to \infty$. So, it suffices to select a large enough $K = k_{j_o}$ to guarantee that the l.h.s. of (3.18) does not exceed $c_6\varepsilon$, for some $c_6 > 0$. The rest of the proof holds, with possibly different constants.

Remark 3.4. Another way to generalize Proposition 3.1 is by proving its statement for any periodic continuous global observable f. In such a case, one calls "wavelength" any interval of the type $[jT, (j+1)T]$, where T is the period of f. Without loss of generality, as we have seen many times so far, one can assume that $f^* := \int_{jT}^{(j+1)T} f\, dm = 0$. As in the proof of the proposition, one constructs the observable g by means of a piecewise affine transformation that adapts the wavelengths of f to the intervals I_j defined by (3.8). Once again $|f(x) - g(x)| \le \varepsilon$, for all large enough x. The proof flows as before, except that:

1. The restrictions $g_\varepsilon|_{I_j} = g|_{I_j}$ are not Lipschitz continuous but only continuous (thus uniformly continuous), so (3.19) might not hold. On the other hand, since the graphs of $g|_{I_j}$ get closer and closer, upon suitable translation, as $j \to \infty$, it is easy to see that one can find a function $\delta \mapsto \Omega(\delta)$, with $\lim_{\delta\to 0^+} \Omega(\delta) = 0$, which is an upper bound for the moduli of continuity of all $g|_{I_j}$, for $j \ge j_o$. Therefore, (3.19) can be replaced by
$$\left| g_\varepsilon(T_\beta^{s+i}(x)) - g_\varepsilon(b_{k_{j+1}-1-i}) \right| < \Omega(2\varepsilon). \quad (3.34)$$
2. Equation (3.20) may not be true: the Riemann sum of g over the midpoints of the uniform partition of I_j is not *exactly* zero in general, but it is nonetheless close to zero if the partition is dense enough, which happens for j large enough. In other words, for every $j \ge j_o$ (with a possible redefinition of j_o),
$$\left| \frac{1}{r_j} \sum_{i=0}^{r_j-1} g_\varepsilon(b_{k_{j+1}-1-i}) \right| \le \varepsilon. \quad (3.35)$$

So, replacing (3.19) and (3.20) with (3.34) and (3.35), one rewrites (3.21) as

$$\left| \frac{1}{r_j} \sum_{i=0}^{r_j-1} g_\varepsilon(T_\beta^{s+i}(x)) \right| < \Omega(2\varepsilon) + \varepsilon. \quad (3.36)$$

What matters is that the above r.h.s. vanishes for $\varepsilon \to 0^+$. The rest of the proof is the same, except that one uses $\Omega(2\varepsilon) + \varepsilon$ instead of $2c_3\varepsilon$.

IV. COUNTEREXAMPLES AND DISCUSSION

The results of Sec. II are obviously not as strong as the original Birkhoff Theorem, which covers a wide functional space of observables (that is, L^1). In particular, one may point out that the main hypothesis of Theorem 2.5—namely, in overwhelmingly large portions of the space, the partial Birkhoff averages of f over long (though fixed!) times are well approximated by a constant—somehow contains the assertion of the theorem. On the other hand, the known examples of L^∞ observables whose Birkhoff average does not converge almost everywhere to a constant are precisely functions that are constant on regions of the space where the moving point takes longer and longer excursions.

The most famous such example concerns Boole's transformation $T(x) = x - 1/x$ on $\mathbb{R}$, cf. Sec. II. It is known that the *frequency* of visits to the positive half-line follows a non-trivial law.[24] More precisely, if ν is a Lebesgue-absolutely continuous probability measure on $\mathbb{R}$ (remember that T preserves the Lebesgue measure), then, for $0 \le t \le 1$,

$$\lim_{n\to\infty} \nu(\{\mathcal{A}_n(1_{\mathbb{R}^+}) \le t\}) = \frac{2}{\pi} \arcsin\sqrt{t}. \quad (4.1)$$

Thus, $\mathcal{A}_n$ cannot converge to a constant almost everywhere. As a matter of fact, it can be proved that, for m-a.e. $x \in \mathbb{R}$,

$$\liminf_{n\to\infty} \mathcal{A}_n(1_{\mathbb{R}^+}) = 0$$
$$\limsup_{n\to\infty} \mathcal{A}_n(1_{\mathbb{R}^+}) = 1. \quad (4.2)$$

The limit in (4.1) is usually referred to as the *arcsine law* for the occupation times of half-lines for Boole's transformation. In fact, since $1_{[a,+\infty)} - 1_{\mathbb{R}^+} \in L^1(\mathbb{R}, m)$ for all $a \in \mathbb{R}$, it follows from Proposition 2.2 that the indicator function of $\mathbb{R}^+$ in (4.1) can be replaced by that of any other right half-line. The law for $\mathcal{A}_n(1_{(-\infty,a]})$ follows straightforwardly.

This example has been generalized in a number of ways, cf. Refs. 25 and 21, and references therein. A strong recent result is that of Sera and Yano (Ref. 21, Thm. 2.7) about the joint distribution of the frequencies of visits to many infinite-measure sets. We describe it in loose terms: For an infinite-measure-preserving, conservative, ergodic dynamical system (X, μ, T), suppose that X is the disjoint union of $X_0, R_1, R_2, \ldots, R_d$, with $0 < \mu(X_0) < \infty$ and $\mu(R_i) = \infty$ for all i. Suppose also that, for $i \neq j$, an orbit point cannot pass from R_i to R_j without visiting X_0. One says that the *rays* $\{R_i\}$ are *dynamically separated* by the *junction* X_0. Under a couple of important technical assumptions (one concerning the so-called asymptotic entrance densities from the rays into the junction and the other concerning the regular variation of certain normalizing rates (Ref. 21, Ass's. 2.3 and 2.5)), the random vector $(\mathcal{A}_n(1_{R_1}), \ldots, \mathcal{A}_n(1_{R_d}))$ converges in distribution, w.r.t. any probability $\nu \ll \mu$, to the vector

$$\frac{(\xi_1, \ldots, \xi_d)}{\sum_{i=1}^d \xi_i}, \tag{4.3}$$

where $\xi_1, \ldots, \xi_d$ are positive i.i.d. random variables with one-sided β-stable distributions (save for degenerate cases). Therefore, typically, the Birkhoff average,

$$f = \sum_{i-1}^{d} \gamma_i 1_{R_i}, \tag{4.4}$$

converges in distribution of a non-constant variable.

The above discussion shows that some "dynamical averaging" hypothesis is needed for a bounded observable to fulfill the assertion of the Birkhoff Theorem. So, the question is, how slowly is a function allowed to vary over the orbit s of T (say, how close to a constant on each ray must it be) in order for it to still have a constant overall Birkhoff average?

The dynamical system and observables of Sec. III are a good case study. Proposition 3.1 and Remark 3.3 guarantee that, for $0 < \beta \leq 1/2$, the Birkhoff average of the "wave" $f(x) = e^{2\pi i \omega x}$, for the $\alpha(\beta)$-Farey map T_β, vanishes almost everywhere. Recall that the lengths of the cylinders L_k of T_β decrease like $k^{-\beta}$. This means that the smaller the β, the closer the partition $\{L_k\}$ is to the uniform partition, when restricted to a wavelength of f, implying that the orbit segments that traverse a wavelength of f contribute with an almost null partial average for f. By contrast, for β close to 1, there will be many more cylinders in the right half of the wavelength than in the left half, making the variation of f along an orbit segment in the right half much slower than the corresponding variation on the left half.

So, the arguments in the proof of Proposition 3.1 do not work for $\beta > 1/2$. We do not know whether $\mathcal{A}_n f$ vanishes for $\beta > 1/2$ as well, but numerical simulations do show a different behavior than the case $\beta \leq 1/2$; see Figs. 4–8.

We also point the reader to Fig. 9, which shows the erratic behavior of $\mathrm{Re}(\mathcal{A}_n f)$ for large β. Also, compare this figure to Fig. 10, which displays the same plot as in Fig. 9 but for the Farey map T_F as in (2.22). The Farey map is akin to T_β with $\beta = 1$. In fact, as can be calculated easily, the partition $\{L_k\}$ for T_F is given by $L_0 = [0, \ln 2)$ and $L_k = [\ln(k+1), \ln(k+2))$ for $k \geq 1$. Thus, $m(L_k) \sim k^{-1}$.

We can produce more counterexamples to a general Birkhoff Theorem for L^∞ observables than mentioned above. The ones that we present momentarily are interesting not only because they do not follow directly from the results of Ref. 21 but also because the Birkhoff sums that we write are representations of *Lévy walks*. Lévy walks are well-studied stochastic processes, often used in nonlinear and statistical physics as models for anomalous diffusion and transport.[26] In fact, we will use this representation to derive a very fine limit theorem for our observables, thus adding to the connections between the field of anomalous stochastic processes and infinite ergodic theory; cf. Ref. 14 and references therein.

Our dynamical system is a Kakutani tower, for which we employ the notation $(Y, \nu, \mathcal{T})$ of the Appendix. We start by defining the base map $S : \Sigma \longrightarrow \Sigma$, where $\Sigma := [0,1) \times [0,1)$ and $x = (x_1, x_2)$ is a generic element of Σ. Let $\mathcal{B} = \{B_i\}_{i \in \mathbb{Z}^+}$ be a partition of $[0,1)$ made up of right-open intervals, which are ordered from left to right. Assume also that $m(B_i) \sim c\, i^{-\beta-1}$, for some $c > 0$ and $\beta \in (0,1)$. Let us define $S_\mathcal{B} : [0,1) \longrightarrow [0,1)$ to be the full-branched, piecewise-linear and increasing Markov map relative to $\mathcal{B}$. In other words, $S_\mathcal{B}|_{B_i}$ maps B_i onto $[0,1)$ with derivative $1/m(B_i)$. It is clear that $S_\mathcal{B}$ preserves the Lebesgue measure m and that the partitions $\mathcal{B}, S_\mathcal{B}^{-1}\mathcal{B}, \ldots, S_\mathcal{B}^{-n}\mathcal{B}, \ldots$ are independent w.r.t. m. Then, let $\mathcal{C} = \{C_j\}_{j \in \mathbb{J}}$ be another partition of $[0,1)$ given by right-open intervals. Here, $\mathbb{J}$ can be either $\{1, 2, \ldots, N\}$, for some positive integer N, or $\mathbb{Z}^+$. Again let us assume that the intervals C_j are ordered from left to right. In analogy with the previous case, we denote $S_\mathcal{C}$ the full-branched, piecewise-linear and increasing Markov map of $[0,1)$ relative to $\mathcal{C}$. This map has the same properties as $S_\mathcal{B}$. Define $S := S_\mathcal{B} \times S_\mathcal{C}$, i.e., $S(x_1, x_2) := (S_\mathcal{B}(x_1), S_\mathcal{C}(x_2))$. Thus, S is a two-dimensional uniformly expanding map which preserves the Lebesgue measure of Σ; in accordance with the notation of the Appendix, this measure will be called ρ. Also, the partition $\mathcal{B} \otimes \mathcal{C} := \{B_i \times C_j\}$ of Σ has the property that all its back-iterates $S^{-n}(\mathcal{B} \otimes \mathcal{C})$ are mutually independent.

For $i \in \mathbb{Z}^+$ denote $A_i := B_i \times [0,1)$. The height function $\varphi : \Sigma \longrightarrow \mathbb{N}$ is defined by the identities

$$\varphi|_{A_i} \equiv i - 1. \tag{4.5}$$

Thus, $\rho(\{\varphi \geq k\}) = \sum_{i>k} \rho(A_i) = \sum_{i>k} m(B_i) \sim c\beta k^{-\beta}$. It follows that the invariant measure ν of $\mathcal{T}$, which is the Lebesgue measure on each level of the tower

$$L_k := \{x \in \Sigma \varphi(x) \geq k\} \times \{k\} = \bigcup_{i \geq k+1} A_i \times \{k\}, \tag{4.6}$$

cf. (A3), is infinite.

Lastly, we introduce the observable $f : Y \longrightarrow \mathbb{C}$. Let $\{\gamma_j\}_{j \in \mathbb{J}}$ be a set of complex numbers with $|\gamma_j| = 1$ and define f so that

$$f|_{[0,1) \times C_j} \equiv \gamma_j. \tag{4.7}$$

The easiest example of such an observable is when $\mathbb{J} = \{1, 2\}$ and $\gamma_1 = -1$, $\gamma_2 = 1$.

Proposition 4.1. *For the dynamical system $(Y, \nu, \mathcal{T})$ introduced above and the function f defined by (4.7), let us interpret the Birkhoff sum $\mathcal{S}_n f$ as a random variable for the probability measure $\nu_0 := \nu(\cdot \mid L_0)$, where this means that*

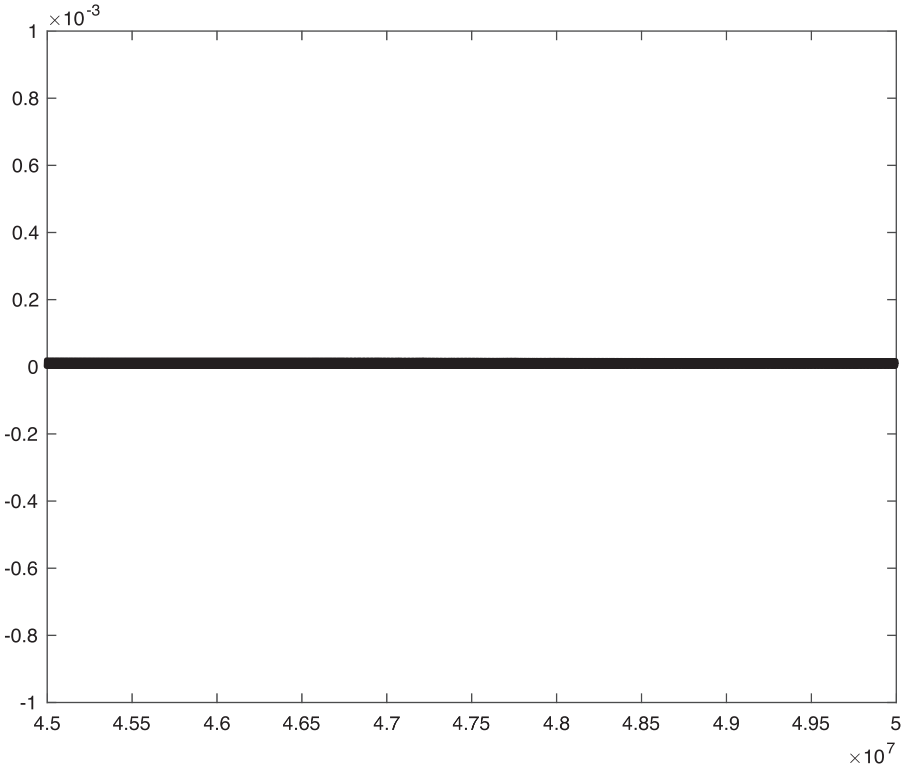

FIG. 4. For the map T_β with $\beta = 0.35$, the figure shows a plot of $\mathcal{A}_n g(x_0)$, with $g(x) = \cos(2\pi\omega x)$, $\omega = 0.2$ and $x_0 = 0.65$. Here, $4.5 \times 10^7 \leq n \leq 5 \times 10^7$ and the vertical scale is in units 10^{-3}.

$\mathcal{S}_n f(y)$ depends on the initial condition $y = (x, 0)$, where x is chosen randomly in Σ according to the Lebesgue measure. Then the process

$$(L_n(t))_{t\in\mathbb{R}_0^+} := \left(\frac{\mathcal{S}_{\lfloor nt\rfloor} f}{n}\right)_{t\in\mathbb{R}_0^+},$$

converges in distribution, w.r.t. the topology of the uniform convergence on all intervals $[0, T]$, to a continuous $\mathbb{C}$-valued process $(L(t))_{t\in\mathbb{R}_0^+}$. If not all γ_j are equal (assuming that $m(C_j) > 0$ for all $j \in \mathbb{J}$), then, for every $t \geq 0$, $L(t)$ is almost surely non-constant. In particular,

$$\mathcal{A}_n f = \frac{\mathcal{S}_n f}{n},$$

converges in distribution to a non-constant random variable.

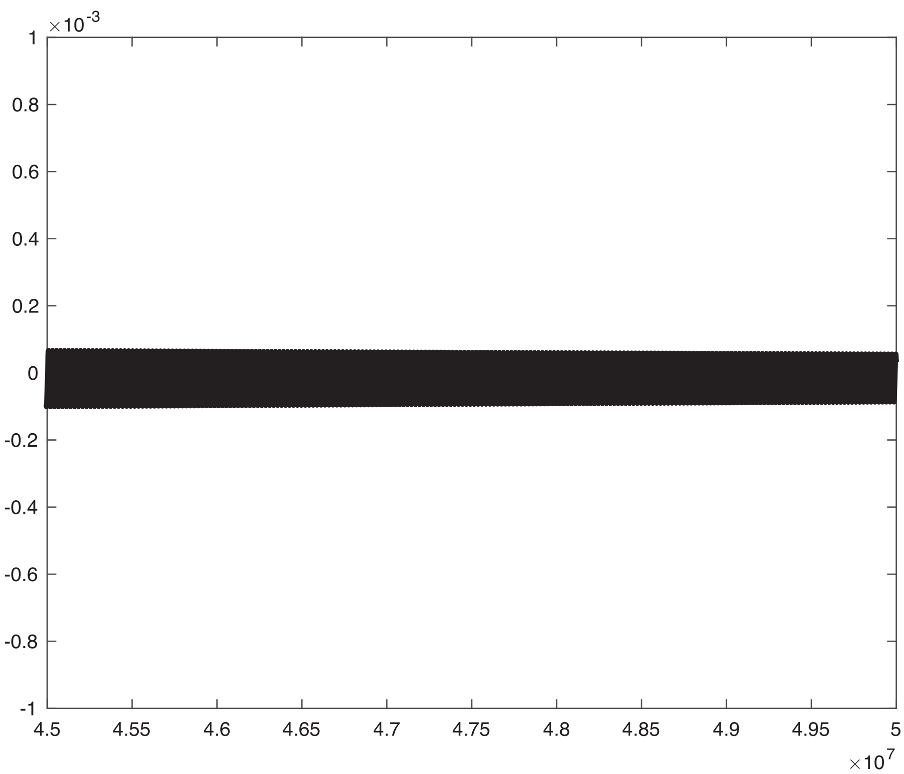

FIG. 5. Same plot as in Fig. 4, for the case $\beta = 0.48$.

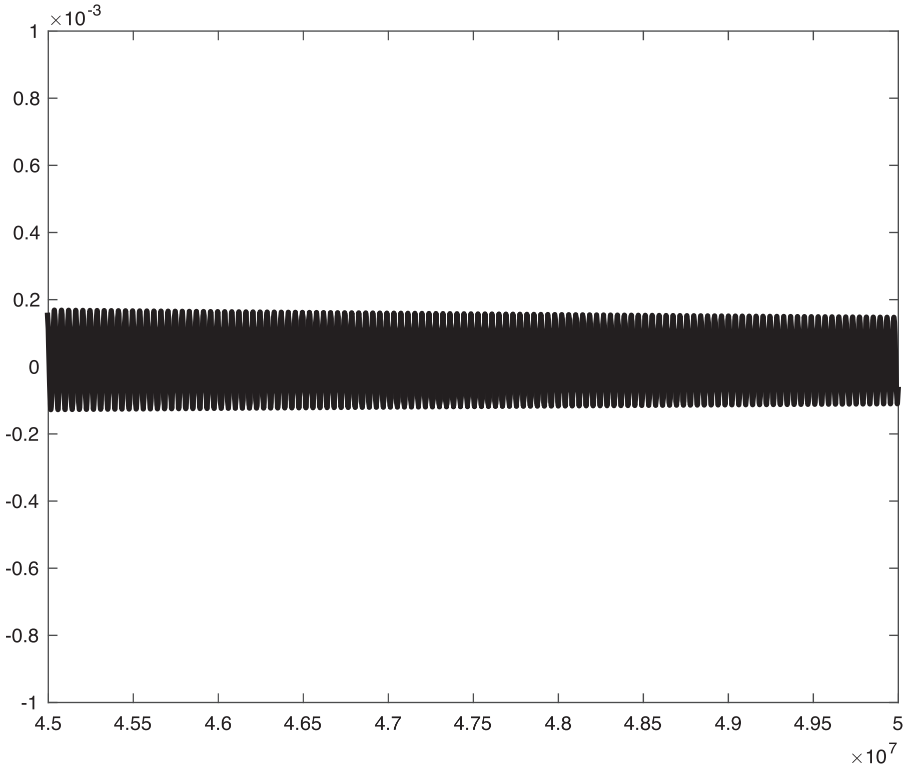

FIG. 6. Same plot as in Fig. 4, for the case $\beta = 0.50$.

Proof. We claim that

$$\mathcal{S}_n(f \circ \mathcal{T}) = \sum_{k=1}^{n} f \circ \mathcal{T}^k, \tag{4.8}$$

is the *Lévy walk* on $\mathbb{C}$ thus defined: A walker stands at the origin of $\mathbb{C}$ when she reads the value of a random integer I_1 and a random complex number Γ_1, with the following probabilities:

$$\forall\, i \in \mathbb{Z}^+, \quad j \in \mathbb{J}, \quad \mathrm{Prob}\{(I_1, \Gamma_1) = (i, \gamma_j)\} = m(B_i)m(C_j). \tag{4.9}$$

All other values of (I_1, Γ_1) occur with probability zero. Observe that I_1 and Γ_1 are independent by definition; remember also that $|\Gamma_1| = 1$. The walker then takes I_1 unit steps in the direction Γ_1 *one step at a time*—which is why we speak

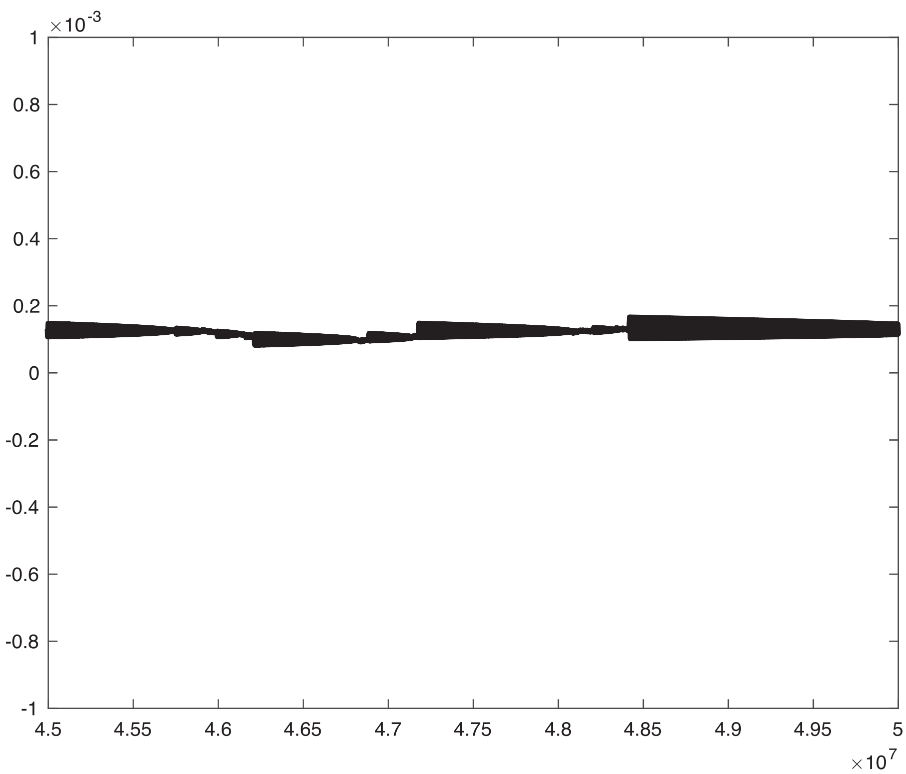

FIG. 7. Same plot as in Fig. 4, for the case $\beta = 0.52$.

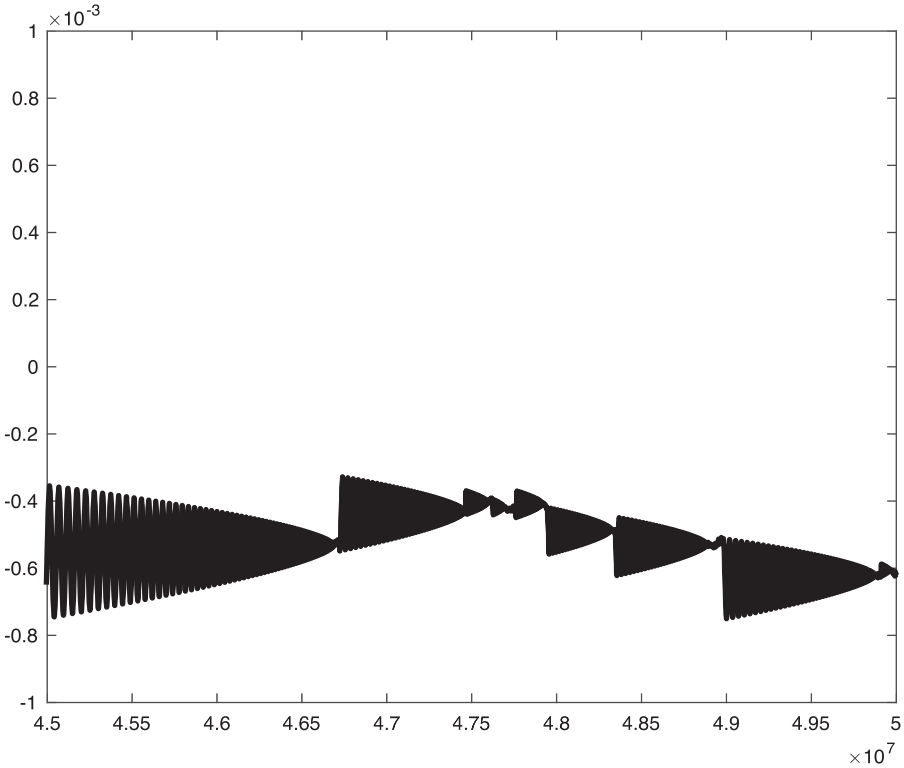

FIG. 8. Same plot as in Fig. 4, for the case $\beta = 0.65$.

of "walk" instead of "jump." At this point, the walker reads the value of another random pair (I_2, Γ_2), with the same probabilities as the previous pair and independent of it. This will determine, in the same way as before, the motion of the walker during the next I_2 time units. And so on.

In other words, we have described a persistent random walk on $\mathbb{C}$, equivalently, a random walk with an internal state,[9] with long-tailed inertial segments, since $\mathrm{Prob}(I_1 = i) \sim c\, i^{-\beta-1}$, with $0 < \beta < 1$. (The simple case $\mathbb{J} = \{1, 2\}$, $\gamma_1 = -1$, $\gamma_2 = 1$ corresponds to a *simple symmetric Lévy walk* on the real line.)

The claim is not hard to show. Let us for the moment suppose that the reference, or *initial*, measure is not ν_0 but $\nu(\cdot \mid B_{i_0} \times C_{j_0} \times \{0\})$, for some choice of $i_0 \in \mathbb{Z}^+$ and $j_0 \in \mathbb{J}$. Using the fact that the base map S sends each $B_i \times C_j$ affinely onto Σ and recalling the definition (A4) of the tower map $\mathcal{T}$,

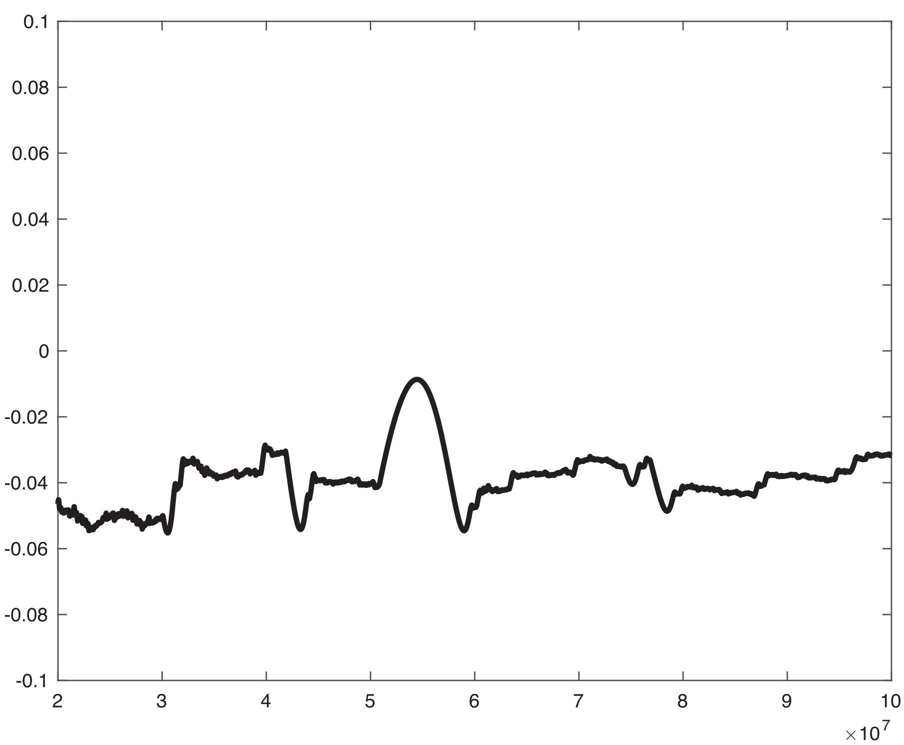

FIG. 9. Plot of $\mathcal{A}_n g(x_0)$, for the same g and x_0 as in Fig. 4, relative to T_β with $\beta = 0.98$. Here, $0.2 \times 10^8 \le n \le 10^8$ and the vertical scale is in absolute units.

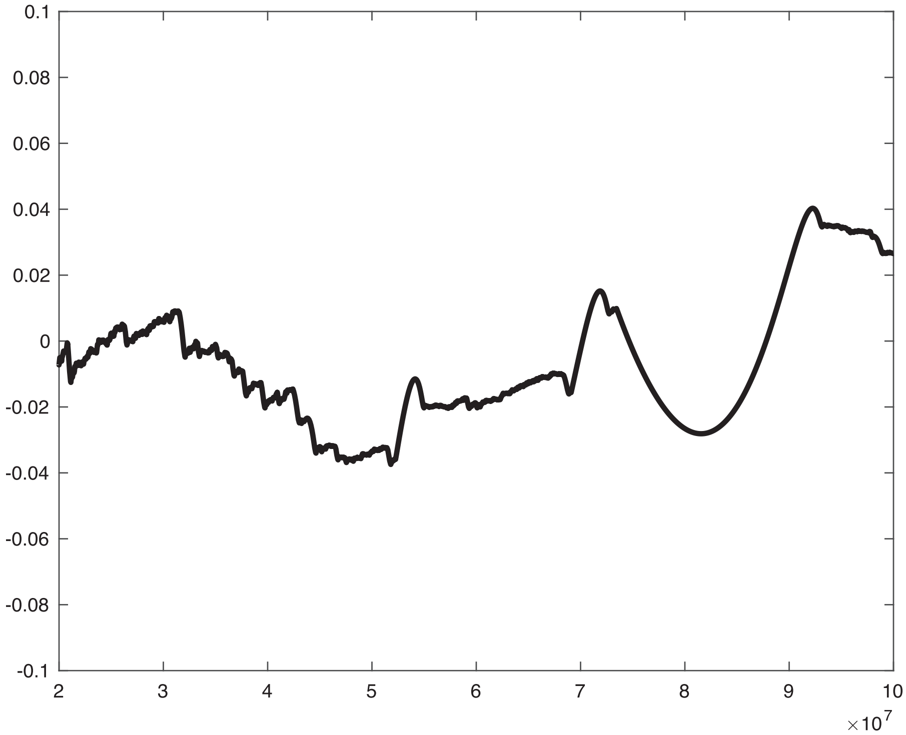

FIG. 10. Plot of $\mathcal{A}_n g(x_0)$, for the same g and x_0 as in Fig. 4, relative to the Farey map T_F. The horizontal range and vertical scale are the same as in Fig. 9.

we see that the push-forward of the initial measure is given by

$$\mathcal{T}_* \nu(\cdot \mid B_{i_0} \times C_{j_0} \times \{0\}) = \sum_{i_1, j_1} m(B_{i_1}) m(C_{j_1}) \nu(\cdot \mid B_{i_1} \times C_{j_1} \times \{i_1 - 1\}), \tag{4.10}$$

where the sum is over $(i_1, j_1) \in \mathbb{Z}^+ \times \mathbb{J}$. If we we fix one such pair (i_1, j_1) and condition the above to $B_{i_1} \times C_{j_1} \times \{i_1 - 1\}$—more precisely, if we condition the initial measure to the event $\{\mathcal{T}(x, 0) \in B_{i_1} \times C_{j_1} \times \{i_1 - 1\}\} = \{S(x) \in B_{i_1} \times C_{j_1}\}$—we can push-forward the resulting measure down the levels of the tower. In formula, for all $1 \le k \le i_1$,

$$\mathcal{T}_*^k \nu \left(\cdot \mid (B_{i_0} \times C_{j_0} \times \{0\}) \cap \mathcal{T}^{-1}(B_{i_1} \times C_{j_1} \times \{i_1 - 1\})\right) = \nu \left(\cdot \mid B_{i_1} \times C_{j_1} \times \{i_1 - k\}\right). \tag{4.11}$$

Therefore, for any y as specified by the conditioning in the above l.h.s., $f(\mathcal{T}^k(y)) = \gamma_{j_1}$, whence $\mathcal{S}_k(f \circ \mathcal{T})(y) = k\gamma_{j_1}$. At time $k = i_1$, the r.h.s. of (4.11) is the Lebesgue measure on $B_{i_1} \times C_{j_1} \times \{0\}$, that is, it has the same form as the initial measure. In other words, the process has *renewed*, losing all memory of the initial measure.

In more detail, this implies that if we fix $q \in \mathbb{Z}^+$, $(i_1, j_1), \ldots, (i_q, j_q) \in \mathbb{Z}^+ \times \mathbb{J}$ and consider all the initial conditions y such that the first excursion down the tower starts in $B_{i_1} \times C_{j_1} \times \{i_1 - 1\}$, the second excursion starts in $B_{i_2} \times C_{j_2} \times \{i_2 - 1\}$ and so on up to the qth excursion, then for all $n := i_1 + \cdots + i_{q-1} + k$, with $1 \le k \le i_q$, we have

$$\mathcal{S}_n(f \circ \mathcal{T})(y) = i_1 \gamma_{j_1} + \cdots + i_{q-1} \gamma_{j_{q-1}} + k \gamma_{j_q}. \tag{4.12}$$

Conditioning to the set of all such y and recalling that the pairs $(i_1, j_1), \ldots, (i_q, j_q)$ are i.i.d. for the initial measure $\nu(\cdot \mid B_{i_0} \times C_{j_0} \times \{0\})$ proves our claim, at least for such choice of the initial measure.

Extending the proof to the case where the initial measure is ν_0, as defined in the statement of the proposition, is immediate. Indeed, the arguments described above depend in no way on i_0, j_0, and ν_0 is a convex linear combination of the probability measures $\nu(\cdot \mid B_{i_0} \times C_{j_0} \times \{0\})$, for $(i_0, j_0) \in \mathbb{Z}^+ \times \mathbb{J}$.

Having established the claim, the assertion of Proposition 4.1 follows from Corollary 4.14 of Ref. 18 and the fact that $\mathcal{S}_n f = f + \mathcal{S}_{n-1}(f \circ \mathcal{T})$. The process $[L(t)]$ is a combination of certain Lévy processes whose marginals at any fixed time t are non-constant with probability 1 (Ref. 18, Eqs. (4.13) and (3.10)). (In truth, the results of Ref. 18 are stated for the case where $(L_n(t))$ is a continuous-time process, that is, the walker moves continuously with unit speed from one "renewal point" to the next. Extending such results to our case is trivial.) □

Remark 4.2. One might wonder why, in Proposition 4.1, the scaling rate of $\mathcal{S}f_n$ (that is, n, a.k.a. *ballistic scaling*) does not depend on β, the exponent of the tail of the distribution of the inertial segments, when $0 < \beta < 1$. This is a fact about Lévy walks, a rigorous proof of which can be found in Ref. 18. Here, we give a simple, heuristic, explanation. If $\{\mathcal{X}_n\}_{n \in \mathbb{N}}$ denotes the Lévy walk in the proof of Proposition 4.1, let $\{\mathcal{Y}_k\}_{k \in \mathbb{N}}$ denote its associated *Lévy flight*, defined by $\mathcal{Y}_0 \equiv 0$ and $\mathcal{Y}_k := \sum_{q=1}^k I_q \Gamma_q$. (Recall that $\{I_q\}_{q \in \mathbb{Z}^+}$ and $\{\Gamma_q\}_{q \in \mathbb{Z}^+}$ are two independent i.i.d. processes such that I_1 takes values in $\mathbb{Z}^+$ and is in the normal basin of attraction of a skewed β-stable distribution, and Γ_1 takes values in $\mathbb{S}^1 \subset \mathbb{C}$.) In other words, $\{\mathcal{Y}_k\}$ is the Lévy walk $\{\mathcal{X}_n\}$ seen at its renewal times. Furthermore, $\{\mathcal{X}_n\}$ is a unit-speed interpolation of $\{\mathcal{Y}_k\}$. Now, let $\tau_k := \sum_{q=1}^k I_q$ denote the sequence of renewal times, with $\tau_0 \equiv 0$. By the hypothesis on I_1, $\tau_k \approx k^{1/\beta}$, as $k \to \infty$.[11] The same hypothesis shows that the whole process $\{k^{-1/\beta} \mathcal{Y}_{\lfloor ks \rfloor}\}_{s \in \mathbb{R}_0^+}$ converges to a Lévy process $\{\overline{\mathcal{Y}}(s)\}_{s \in \mathbb{R}_0^+}$, in a sense that we do not specify here. Now, denote by $n \mapsto K_n$ the generalized inverse of $k \mapsto \tau_k$, i.e., the non-decreasing function $\mathbb{N} \longrightarrow \mathbb{N}$ such that $\tau_{K_n} \le n < \tau_{K_n + 1}$.

Clearly, $K_n \approx n^\beta$. By construction, $\mathcal{X}_n$ always lies between $\mathcal{Y}_{K_n}$ and $\mathcal{Y}_{K_n+1}$. These two processes are not the same—they are sometimes called the *lagging* and *leading* walks of $\mathcal{X}_n$, respectively—but it is easy to show that they scale in the same way. For the purposes of this explanation, we can approximate $\mathcal{X}_n$ with $\mathcal{Y}_{K_n}$; therefore, in non-rigorous notation, we can write that, for $n \to \infty$,

$$\mathcal{X}_{\lfloor nt \rfloor} \approx \mathcal{Y}_{K_{\lfloor nt \rfloor}} \approx \mathcal{Y}_{\lfloor n^\beta t^\beta \rfloor} \approx n\,\overline{\mathcal{Y}}(t^\beta). \tag{4.13}$$

We end this remark by observing that this explanation only holds for $\beta \in (0,1)$. In all other cases, since the first moment of I_1 is finite (or barely infinite), the scaling of $\mathcal{Y}_k$ is generally different from that of τ_k.

The example that we have presented in Proposition 4.1 is in the same spirit as the occupation times of dynamically separated sets. In fact, if we denote

$$R_j := \bigcup_{k\geq 1} \bigcup_{i\geq k+1} B_i \times C_j \times \{k\}, \tag{4.14}$$

cf. (4.6), we realize that, for $j \in \mathbb{J}$, the infinite-measure sets R_j are dynamically separated by the juncture L_0. Observe that, in the case where the γ_j are all different, $R_j = (Y \setminus L_0) \cap \{f = \gamma_j\}$. In any case, f can be expressed as

$$f = \sum_{j\in\mathbb{J}} \gamma_j 1_{R_j} + \sum_{j\in\mathbb{J}} \gamma_j 1_{[0,1)\times C_j \times\{0\}}, \tag{4.15}$$

where the second sum above amounts to an integrable function, cf. (4.4).

Nevertheless, the statistical properties of $\mathcal{A}_n f$ cannot be derived from the main theorem of Ref. 21, and not only because here we have infinitely many rays. The most important difference is that the assumption on the asymptotic entrance densities (Ref. 21, Ass. 2.3) is not satisfied. This follows from the triviality of the dynamics on the non-zero levels of the tower.

Moreover, our system can be generalized to the case of uncountably many rays. It suffices to replace the base map with $S := S_\mathcal{B} \times \sigma$, where $S_\mathcal{B}$ is the map defined earlier and σ is the left shift on the space $([0,1)^\mathbb{N}, m^\mathbb{N})$ of sequences of i.i.d. numbers uniformly distributed in $[0,1)$. If we define

$$f(y) = f(x_1, (\theta_q)_{q\in\mathbb{N}}, k) := e^{2\pi i\theta_0}, \tag{4.16}$$

and use the reference measure $\nu_0 = \nu(\cdot \mid L_0)$ (here ν_0 is isomorphic to $m \times m^\mathbb{N}$ on $\Sigma := [0,1) \times [0,1)^\mathbb{N}$), we see that during the qth excursion in the tower, the value of f is $e^{2\pi i\theta_q}$ and it is independent of the values taken during the previous excursions. Therefore, the process $\mathcal{S}_n f$ is a radially symmetric Lévy walk on $\mathbb{C}$. The assertions of Proposition 4.1 still hold. Finally, writing $\mathbb{N} = \{0\} \times \mathbb{Z}^+$, it is clear that the sets

$$R_{\theta_0} := (Y\setminus L_0) \cap \{f = e^{2\pi i\theta_0}\} = \bigcup_{k\geq 1}\bigcup_{i\geq k+1} B_i \times \left(\{\theta_0\} \times [0,1)^{\mathbb{Z}^+}\right) \times \{k\}, \tag{4.17}$$

for $\theta_0 \in [0,1)$, are dynamically separated rays.

ACKNOWLEDGMENTS

This research is part of the authors' activity within the DinAmicI community, see www.dinamici.org and also part of M.L.'s activity within the *Gruppo Nazionale di Fisica Matematica* (INdAM, Italy).

APPENDIX: KAKUTANI TOWERS

Let us briefly recall the definition and basic properties of a Kakutani tower. For more details, we refer to Ref. 1 (Sec. 1.5). In this appendix, we restore the indication of the σ-algebra in the notation. So, let $(\Sigma, \mathscr{B}, \rho, S)$ be a conservative, non-singular dynamical system on a σ-finite measure space, and suppose that $\varphi : \Sigma \longrightarrow \mathbb{N}$ is a measurable function . Then, the tower over S with *height function* φ is the dynamical system $(Y, \mathscr{C}, \nu, \mathcal{T})$ defined as follows:

$$Y := \{(x,k) \in \Sigma \times \mathbb{N} 0 \leq k \leq \varphi(x)\}, \tag{A1}$$

$$\mathscr{C} := \sigma\left(\{A \times \{k\} | k \in \mathbb{N}, A \subseteq \{\varphi \geq k\}, A \in \mathscr{B})\right\}, \tag{A2}$$

$$\nu(A \times \{k\}) := \rho(A), \tag{A3}$$

$$\mathcal{T}(x,k) := \begin{cases} (x, k-1) & \text{if } k \geq 1 \\ \{S(x), \varphi(S(x))\} & \text{if } k = 0. \end{cases} \tag{A4}$$

(In (A2) the notation $\sigma(\cdot)$ denotes the σ-algebra generated by the sets between parentheses.) The *tower map* $\mathcal{T}$ is conservative and non-singular, and if $\rho \circ S^{-1} = \rho$, then $\nu \circ \mathcal{T}^{-1} = \nu$. Furthermore, if S is ergodic, then $\mathcal{T}$ is ergodic.

Now, suppose that $(X, \mathscr{A}, \mu, T)$ is an invertible measure-preserving system, and let $\Sigma \in \mathscr{A}$ be a sweep-out set with $\mu(\Sigma) > 0$. (Note that this implies that T is conservative, by Maharam's Recurrence Theorem, see Ref. 1 (Thm. 1.1.7).) Denote the induced map of T on Σ by $T_\Sigma : \Sigma \longrightarrow \Sigma$, and by $\mathscr{A}_\Sigma$ and μ_Σ, respectively, the restrictions of $\mathscr{A}$ and μ to Σ. Also, set

$$\varphi(x) := \min\{n \geq 0 T^{-n-1}(x) \in \Sigma\}. \tag{A5}$$

In other words, $\varphi : \Sigma \longrightarrow \mathbb{N}$ is the first-return function of T^{-1} to the set Σ, minus one unit.

Proposition A.1. *The tower constructed over the dynamical system $(\Sigma, \mathscr{A}_\Sigma, \mu_\Sigma, T_\Sigma)$ w.r.t. the height function φ is measure-theoretically isomorphic to $(X, \mathscr{A}, \mu, T)$.*

Proof. In the following, we shall always restrict ourselves to the full-measure set of points in X for which T^n is invertible for every $n \geq 1$.

Since Σ is a sweep-out set, for almost every $x \in X$, there exists a smallest $n \geq 0$ such that $z := T^n(x) \in \Sigma$. We define the map $\Phi : X \longrightarrow Y$, where Y is the reference space of the tower as described above, by setting

$$\Phi(x) := (z, n). \tag{A6}$$

This map is well-defined because, by construction, the first-return time of z to Σ, w.r.t. T^{-1} must be strictly larger than n, implying that $\varphi(z) \geq n$. Clearly, Φ is injective, since, for $x_1 \neq x_2 \in X$, either these two points have different landing points z_1, z_2 in Σ, or $z_1 = z_2$. In the latter case, however, the T-trajectories of x_1, x_2 cannot get to $z_1 = z_2$ after the same number n of iterations, otherwise the map T^n would not be invertible. Moreover, for a.e. $(z,n) \in Y$, if we set $x := T^{-n}(z)$, then $\Phi(x) = (z,n)$, showing that Φ is also surjective.

It remains to demonstrate that $T = \Phi^{-1} \circ \mathcal{T} \circ \Phi$. Let us consider two cases. First, suppose that $x \in X \setminus \Sigma$. Then, we find our $z = T^n(x) \in \Sigma$ with $n \geq 1$. So,

$$\Phi^{-1} \circ \mathcal{T}(\Phi(x)) = \Phi^{-1}(\mathcal{T}(z,n)) = \Phi^{-1}(z, n-1) = T(x). \tag{A7}$$

The second case is where $x \in \Sigma$. Here, we have that $\Phi(x) = (z, 0)$, with $z = x$. Thus,

$$\begin{aligned}\Phi^{-1} \circ \mathcal{T}(\Phi(x)) &= \Phi^{-1}(\mathcal{T}(z,0)) = \Phi^{-1}\{T_\Sigma(x), \varphi(T_\Sigma(x))\} \\ &= T(x).\end{aligned} \tag{A8}$$

Here, the final equality holds because $\varphi(T_\Sigma(x))$ is equal to $\rho(x) - 1$, where ρ denotes the return-time function to Σ with respect to T, that is, $T_\Sigma(x) := T^{\rho(x)}(x)$. □

Consider again the general set-up from Sec. II. So, $(X, \mathscr{A}, \mu, T)$ is a conservative, ergodic measure-preserving dynamical system on a σ-finite, infinite measure space where we choose a set $L_0 \in \mathscr{A}$ with $0 < \mu(L_0) < \infty$ (L_0 is then a sweep-out set). If it happens that this system is isomorphic to a tower with L_0 identified with the base level $\Sigma \times \{0\}$ (as would be the case for T invertible, as shown above), then each partition element L_k is identified with the kth level of the tower, i.e., $\{\varphi \geq k\} \times \{k\}$. Moreover, since the tower map sends level k injectively into level $k-1$, we gain in this case that T maps L_k injectively into L_{k-1}.

[1] J. Aaronson, *An Introduction to Infinite Ergodic Theory*, Mathematical Surveys and Monographs Vol. 50 (American Mathematical Society, Providence, RI, 1997).

[2] J. Aaronson and H. Nakada, "Trimmed sums for non-negative, mixing stationary processes," Stoch. Process. Appl. **104**(2), 173–192 (2003).

[3] R. L. Adler and B. Weiss, "The ergodic infinite measure preserving transformation of Boole," Israel J. Math. **16**, 263–278 (1973).

[4] C. Bonanno, P. Giulietti, and M. Lenci, "Infinite mixing for one-dimensional maps with an indifferent fixed point," preprint (2017), arXiv:1708.09369.

[5] G. Boole, "On the comparison of transcendents with certain applications to the theory of definite integrals," Philos. Trans. R. Soc. Lond. **147**, 748–803 (1857).

[6] Z. Buczolich, "Ergodic averages and free $\mathbb{Z}^2$ actions," Fund. Math. **160**(3), 247–254 (1999).

[7] Z. Buczolich, "Almost everywhere convergence of ergodic averages," Real Anal. Exch. **34**(1), 1–15 (2009).

[8] M. Carney and M. Nicol, "Dynamical Borel-Cantelli lemmas and rates of growth of Birkhoff sums of non-integrable observables on chaotic dynamical systems," Nonlinearity **30**(7), 2854–2870 (2017).

[9] G. Cristadoro, T. Gilbert, M. Lenci, and D. P. Sanders, "Transport properties of Lévy walks: An analysis in terms of multistate processes," Europhys. Lett. **108**(5), 50002 (2014).

[10] E. Hopf, *Ergodentheorie* (Springer-Verlag, Berlin, 1937).

[11] I. A. Ibragimov and Yu. V. Linnik, *Independent and Stationary Sequences of Random Variables*, With a supplementary chapter by I. A. Ibragimov and V. V. Petrov (Wolters-Noordhoff Publishing, Groningen, 1971).

[12] M. Kesseböhmer, S. Munday, and B. O. Stratmann, "Strong renewal theorems and Lyapunov spectra for α-Farey and α-Lüroth systems," Ergodic Theory Dynam. Syst. **32**(3), 989–1017 (2012).

[13] M. Kesseböhmer and T. Schindler, "Strong laws of large numbers for intermediately trimmed sums of i.i.d. random variables with infinite mean," J. Theor. Probab. (published online).

[14] R. Klages, "From deterministic chaos to anomalous diffusion," in *Reviews of Nonlinear Dynamics and Complexity* (Wiley-VCH Verlag, Weinheim, 2010), Vol. 3, pp. 169–227.

[15] M. Lenci, "On infinite-volume mixing," Commun. Math. Phys. **298**(2), 485–514 (2010).

[16] M. Lenci, "Exactness, K-property and infinite mixing," Publ. Mat. Urug. **14**, 159–170 (2013).

[17] M. Lenci, "Uniformly expanding Markov maps of the real line: Exactness and infinite mixing," Discrete Continuous Dyn. Syst. **37**(7), 3867–3903 (2017).

[18] M. Magdziarz, H. P. Scheffler, P. Straka, and P. Zebrowski, "Limit theorems and governing equations for Lévy walks," Stoch. Process. Appl. **125**(11), 4021–4038 (2015).

[19] P. Major, "A counterexample in ergodic theory," Acta Sci. Math. (Szeged) **62**(1–2), 247–258 (1996).

[20] W. Stepanoff, "Sur une extension du théorème ergodique," Compositio Math. **3**, 239–253 (1936).

[21] T. Sera and K. Yano, "Multiray generalization of the arcsine laws for occupation times of infinite ergodic transformations," preprint (2017), arXiv:1711.03260.

[22] M. Thaler, "Estimates of the invariant densities of endomorphisms with indifferent fixed points," Israel J. Math. **37**, 303–314 (1980).

[23] M. Thaler, "Transformations on [0,1] with infinite invariant measures," Israel J. Math. **46**(1–2), 67–96 (1983).

[24] M. Thaler, "A limit theorem for sojourns near indifferent fixed points of one-dimensional maps," Ergodic Theory Dyn. Syst. **22**(4), 1289–1312 (2002).

[25] M. Thaler and R. Zweimüller, "Distributional limit theorems in infinite ergodic theory," Probab. Theory Relat. Fields **135**(1), 15–52 (2006).

[26] V. Zaburdaev, S. Denisov, and J. Klafter, "Lévy walks," Rev. Mod. Phys. **87**, 483–530 (2015).