A TEMPORAL CENTRAL LIMIT THEOREM FOR REAL-VALUED COCYCLES OVER ROTATIONS

MICHAEL BROMBERG AND CORINNA ULCIGRAI

Abstract. We consider deterministic random walks on the real line driven by irrational rotations, or equivalently, skew product extensions of a rotation by α where the skewing cocycle is a piecewise constant mean zero function with a jump by one at a point β. When α is badly approximable and β is badly approximable with respect to α, we prove a *Temporal Central Limit theorem* (in the terminology recently introduced by D.Dolgopyat and O.Sarig), namely we show that for any fixed initial point, the *occupancy random variables*, suitably rescaled, converge to a Gaussian random variable. This result generalizes and extends a theorem by J. Beck for the special case when α is quadratic irrational, β is rational and the initial point is the origin, recently reproved and then generalized to cover any initial point using geometric renormalization arguments by Avila-Dolgopyat-Duryev-Sarig (Israel J., 2015) and Dolgopyat-Sarig (J. Stat. Physics, 2016). We also use renormalization, but in order to treat irrational values of β, instead of geometric arguments, we use the renormalization associated to the continued fraction algorithm and dynamical Ostrowski expansions. This yields a suitable symbolic coding framework which allows us to reduce the main result to a CLT for non homogeneous Markov chains.

RESUME On considère des marches aléatoires sur la droite réelle, engendrés par des rotations irrationnelles, ou, de manière équivalente, des produits croisés d'une rotation par un nombre réel α, dont le cocycle est une fonction constante par morceaux de moyenne nulle admettant un saut de un à une singularité β. Si α est mal aapproché par des rationnels et β n'est pas bien approché par l'orbite de α, nous démontrons une version temporelle du Théorème de la Limite Centrale (ou un *Temporal Central Limit theorem* dans la terminologie qui a été introduite récemment par D. Dolgopyat et O. Sarig). Plus précisément, nous montrons que, pour chaque point initial fixé, les *variables aléatoires d'occupation*, proprement renormalisées, tendent vers une variable aléatoire de loi normale. Ce résultat généralise un théorème de J. Beck dans le cas sparticulier où α est un nombre irrationnel quadratique, β est un nombre rationnel et le point initial est l'origine. Ce résultat de Beck a été montré avec de nouvelles méthodes et étendu par Avila-Dolgopyat-Duryev-Sarig (Israel J. , 2015) et Dolgopyat-Sarig (J. Stat. Physics, 2016) à l'aide d'une renormalisation géométrique. Dans ce papier, nous utilisons aussi la renormalisation, mais, au lieu d'avoir recours à un argument géométrique, nous proposons d'utiliser l'algorithme de fraction continue avec une version dynamique de l'expansion de Ostrowski. Cela nous donne un codage symbolique qui nous permet de réduire le résultat principal à un théorème de la limite centrale pour de chanes de Markov non-homogènes.

1. INTRODUCTION AND RESULTS

The main result of this article is a temporal distributional limit theorem (see Section 1.1 below) for certain functions over an irrational rotation (Theorem 1.1 below). In order to introduce and motivate this result, in the first section, we first define two types of distributional limit theorems in the study of dynamical systems, namely spatial and temporal. Temporal limit theorems in dynamics are the focus of the recent survey [?] by D. Dolgopyat and O. Sarig; we refer the interested reader to [?] and the references therein for a comprehensive introduction to the subject, as well as for a list of examples of dynamical systems known up to date to satisfy temporal distributional limit theorems. In section 1.2 we then focus on irrational rotations, which are one of the most basic examples of low complexity dynamical systems, and recall previous results on temporal limit theorems for rotations, in particular Beck's temporal CLT. Our main result in stated in section 1.3, followed by a description of the structure of the rest of the paper in section 1.4.

1.1. Temporal and Spatial Limits in dynamics.

Distributional limit theorems appear often in the study of dynamical systems as follows. Let X be a complete separable metric space, m a Borel probability measure on X and denote by $\mathcal{B}$ is the Borel σ-algebra on X. Let $T : X \to X$ be a Borel measurable map. We call the quadruple $(X, \mathcal{B}, m, T)$ a *probability preserving dynamical system* and assume that T is ergodic with respect to m. Let $f : X \to \mathbb{R}$ be a Borel measurable function and set

$$S_n(T, f, x) := \sum_{k=0}^{n-1} f \circ T^k(x)$$

We will also use the notation $S_n(x)$, or $S_n(f, x)$ instead of $S_n(T, f, x)$, when it is clear from the context, what is the underlying transformation or function. The function $S_n(x)$ is called (the n^{th}) *Birkhoff sum* (or also ergodic sum) of the function f over the transformation T. The study of Birkhoff sums, their growth and their behavior is one of the central themes in ergodic theory. When the transformation T is *ergodic* with respect to m, by the *Birkhoff ergodic theorem*, for any $f \in L^1(X, m)$, for m-almost every $x \in X$, $S_n(f, x)/n$ converges to $\int f dm$ as n grows; equivalently, one can say that the random variables $X_n := f \circ T^n$ where x is chosen randomly according to the measure m, *satisfy the strong law of large numbers.* We will now introduce some limit theorems which allow to study the error term in the Birkhoff ergodic theorem.

The function f is said to satisfy a *spatial distributional limit theorem* (*spatial* DLT) if there exists a random variable Y which is non-deterministic (i.e. there there is no $a \in \mathbb{R}$ such that $Prob(Y = a) = 1$) and sequences of constants $A_n, B_n \in \mathbb{R}$, $B_n \to \infty$ such that the random variables $\frac{S_n(x)-A_n}{B_n}$, where x is chosen randomly according to the measure m (which, for short, will be denoted by $x \sim \mu$), converge in distribution to Y. In this case we write

$$\frac{S_n - A_n}{B_n} \xrightarrow{dist} Y \quad \text{for } x \sim \mu.$$

It is the case that many *hyperbolic* dynamical systems, under some regularity conditions on f, satisfy a spatial DLT with the limit being a Gaussian random variable. In the cases that we have in mind, the rate of mixing of the sequence of random variables $X_n := f \circ T^n$ is sufficiently fast, in order for them to satisfy the Central Limit Theorem (CLT). On the other hand, in many classical examples of dynamical systems with *zero entropy,* for which the random variables $X_n := f \circ T^n$ are highly correlated, the spatial DLT fails if f is sufficiently regular. For example, this is the case when T is an irrational rotation and f is of bounded variation.

Perhaps surprisingly, many examples of dynamical systems with zero entropy satisfy a CLT when instead of averaging over the space X, one considers the Birkhoff sums $S_n(x_0)$ over a *single orbit* of some fixed initial condition $x_0 \in X$. Fix an initial point $x_0 \in X$ and consider its orbit under T. One can define a sequence of *occupation measures* on $\mathbb{R}$ by

$$\nu_n(F) := \frac{1}{n} \# \{1 \le k \le n : \quad S_k(x_0) \in F\}$$

for every Borel measurable $F \subset \mathbb{R}$. One can interpret the quantity $\nu_n(F)$ as the fraction of time that the Birkhoff sums $S_k(x_0)$ spend in the set F, up to time n. Let Y_n be a sequence of random variables distributed according to ν_n. We say that the pair (T, f) satisfies a *temporal distributional limit theorem* (*temporal* DLT) along the orbit of x_0, if there exists a non-deterministic random variable Y, and two sequences $A_n \in \mathbb{R}$ and $B_n \to \infty$ such that $(Y_n - A_n)/B_n$ converges in distribution to Y. In other words, the pair (T, f) satisfies a temporal DLT along the orbit of x_0, if

$$\frac{1}{n} \# \left\{1 \le k \le n : \frac{S_k(x_0) - A_n}{B_n} < a\right\} \underset{n\to\infty}{\longrightarrow} Prob(Y < a)$$

for every $a \in \mathbb{R}$, such that $Prob(Y = a) = 0$. If the limit Y is a Gaussian random variable, we call this type of behavior a *temporal CLT* along the orbit of x_0. Note, that this type of result may be interpreted as convergence in distribution of a sequence of normalized random variables, obtained by considering the Birkhoff sums $S_k(x_0)$ for $k = 1, .., n$ and choosing k randomly uniformly.

1.2. **Beck's temporal CLT and its generalizations.** One example of occurrence of a temporal CLT in dynamical systems with zero entropy is the following result by Beck, generalizations of which are the main topic of this paper. Let us denote by R_α the rotation on the interval $\mathbb{T} = \mathbb{R} \setminus \mathbb{Z}$ by an irrational number $\alpha \in \mathbb{R}$, given by

$$R_\alpha(x) = x + \alpha \mod 1.$$

Let $f_\beta : \mathbb{T} \to \mathbb{R}$ be the indicator of the interval $[0, \beta)$ where $0 < \beta < 1$, rescaled to have mean zero with respect to the Lebesgue measure on $\mathbb{T}$, namely

$$f_\beta(x) = \mathbf{1}_{[0,\beta)}(x) - \beta.$$

The sequence $\{S_n\}$ of random variables given by the Birkhoff sums $S_n(x) = S_n(R_\alpha, f_\beta, x)$, where x is taken uniformly with respect to the Lebesgue measure, is sometimes referred to in the literature as the *deterministic random walk* driven by an irrational rotation (see for example [**?**]).

Beck proved [**?**, **?**] that if α is a quadratic irrational, and β is rational, then the pair (R_α, f_β) satisfies a temporal DLT along the orbit of $x_0 = 0$. More precisely, he shows that there exist constants C_1 and C_2 such that for all $a, b \in \mathbb{R}$, $a < b$

$$\frac{1}{n}\#\left\{1 \le k \le n : \frac{S_k(R_\alpha, f_\beta, 0) - C_1 \log n}{C_2\sqrt{\log n}} \in [a,b]\right\} \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx.$$

Beck's CLT relates to the theory of discrepancy in number theory as follows. If $\alpha \in \mathbb{R}$ is irrational, by unique ergodicity of the rotation R_α, the sequence of $\{j\alpha\}$ is *equidistributed modulo one*, i.e. in particular, for any $\beta \in [0, 1]$ if we set

$$N_k(\alpha, \beta) := \#\{0 \le j < k \mid \quad 0 \le \{j\alpha\} < \beta\},$$

where $\{x\}$ is the fractional part of x, then $N_k(\alpha, \beta)/k$ converges to β, or, equivalently, $N_k(\alpha, \beta) = k\beta + o(k)$. Discrepancy theory concerns the study of the error term in the expression $N_k(\alpha, \beta) = k\beta + o(k)$. Beck's result hence says that, when α is a *quadratic irrational* and β is *rational,* the error term $\overline{N_k}(\alpha, \beta) := N_k(\alpha, \beta) - k\beta$, when k is chosen randomly uniformly in $\{1, ..., n\}$, can be normalized so that it converges to the standard Gaussian distribution as n grows to infinity.

Let us also remark that the Birkhoff sums in the statement of Beck's theorem are related to the dynamics of the map $T_{f_\beta} : \mathbb{T} \times \mathbb{R} \to \mathbb{T} \times \mathbb{R}$, defined by

$$T_{f_\beta}(x, y) = (R_\alpha(x), y + f_\beta(x)), \qquad (x, y) \in T \times \mathbb{R},$$

since one can see that the form of the iterates of T_f is $T_f^n(x, y) = (R_\alpha^n(x), y + S_n(f_\beta, x))$. This skew product map has been studied as a basic example in infinite ergodic theory and there is a long history of results on it, starting from ergodicity (see for example [**?**, **?**, **?**, **?**, **?**, **?**]).

Recently, in [**?**], a new proof of Beck's theorem is the special case when $\beta = \frac{1}{2}$ was given. This proof, which uses dynamical and geometrical renormalization tools, is crucially based on the interpretation of the corresponding skew-product map $T_{f_{1/2}}$ as the Poincaré map of a flow on the *staircase* periodic surface, which was noticed and pointed out in [**?**]. In [**?**] this method is generalized to show that for any initial point x, any α quadratic irrational and any *rational* β, there exists a sequence $A_n := A_n(\alpha, \beta, x)$ and a constant $B := B(\alpha, \beta)$ such that

$$\frac{1}{n}\#\left\{1 \le k \le n : \frac{S_k(R_\alpha, f_\beta, x) - A_n}{B\sqrt{\log n}} \in [a,b]\right\} \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx$$

for all $a, b \in \mathbb{R}$, $a < b$. Dolgopyat and Sarig informed us that in ongoing work [**?**] they are also able to prove a temporal CLT for the case in which α *is badly approximable* and β is rational. Furthermore, they can show (also in [**?**]) that the temporal CLT fails for a full Lebesgue measure set of α.

1.3. **Main result and comments.** The main result of this paper is the following generalization of Beck's temporal CLT, in which we consider certain irrational values of β and badly approximable values of α. Let us recall that α is *badly approximable* (or equivalently, α is of *bounded type*) if there exists a constant $c > 0$ such that $|\alpha - p/q| \ge c/|q|$ for any integers p, q such that $q \ne 0$ and $\gcd(p, q) = 1$. Equivalently, α is *badly approximable* if the continued faction entries of α are uniformly bounded. For $\alpha \in (0, 1) \setminus \mathbb{Q}$ let us say that β is *badly approximable with respect to* α if there exists a constant $C > 0$ such that

$$|q\alpha - \beta - p| > \frac{C}{|q|} \qquad \forall p \in \mathbb{Z},\ q \in \mathbb{Z} \setminus \{0\},\ \gcd(p, q) = 1. \tag{1.1}$$

One can show that given a badly approximable α, the set of β which are badly approximable with respect to α have full Hausdorff dimension [**?**, Corollary 1].

Theorem 1.1. *Let $0 < \alpha < 1$ be a badly approximable irrational number. For every β badly approximable with respect to α and every $x \in \mathbb{T}$ there exists a sequence of centralizing constants $A_n := A_n(\alpha, \beta, x)$ and a sequence of normalizing constants $B_n := B_n(\alpha, \beta)$ such that for all $a < b$*

$$\frac{1}{n}\#\left\{1 \leq k \leq n : \frac{S_k(R_\alpha, f_\beta, x) - A_n}{B_n} \in [a, b]\right\} \to \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx.$$

In other words, for every α badly approximable, any β badly approximable with respect to α the pair (R_α, f) satisfies the temporal CLT along the orbit of any $x \in \mathbb{T}$. Note that the centralizing constants depend on x, while the normalizing constants do not. We will see in Section 4.1 that badly approximable numbers with respect to α can be explicitly described in terms of their *Ostrowski expansion*, using an adaptation of the continued fraction algorithm in the context of non homogenous Diophantine Approximation. Let us recall that quadratic irrationals are in particular badly approximable. Moreover, when α is badly approximable, it follows from definition that any *rational* number β is badly approximable with respect to α. Thus, this theorem, already in the special case in which α is assumed to be a quadratic irrational, since it includes *irrational* values of β, gives a strict generalization of the results mentioned above. As we already pointed out, the temporal limit theorem, fails to hold for almost every value of α [**?**] (in preparation). It would be interesting to see whether a temporal CLT holds for a larger class of values of β. Under the present assumptions, we believe that there exist constants $C, c > 0$, such that $c\sqrt{\log n} \leq B_n \leq C\sqrt{\log n}$ for all $n \in \mathbb{N}$; this will be the subject of future generalizations of this work, for which we believe that estimates of the growth rate of the variance will be crucial.

While the proof of Theorem 1.1 was inspired and motivated by an insight of Dolgopyat and Sarig and based, as theirs, on renormalization, we stress that our renormalization scheme and the formalism that we develop is different. As remarked in the previous section, the proof of Beck's theorem in [**?**, **?**] exploits a geometric renormalization which is based on the link with the staircase flow and the existence of affine diffeomorphisms which renormalize certain directions of directional flows on this surface. This geometrical insight, unfortunately, as well as the interpretation of the map T_f as the Poincaré map of a staricase flow, breaks down when β is not rational. Our proof does not rely on this geometric picture, but uses only the more classical renormalization given by the continued fraction algorithm for rotations, with the additional information encoded by Ostrowski expansions in the context of non homogeneous Diophantine approximations (see Section 2). This renormalization allows to encode the dynamics symbolically and reduce it to the formalism of adic and Vershik maps [**?**].

There is a large literature of results on limiting distributions for entropy zero dynamical systems, see for example [**?**, **?**, **?**, **?**, **?**, **?**]. Let us mention two recent results in the context of substitution systems which are related to our work. Bressaud, Bufetov and Hubert proved in [**?**] a spatial CLT for substitutions with eigenvalues of modulus one along a subsequence of times. In the same context (substitutions with eigenvalues of modulus one), Paquette and Son [**?**] recently also proved a *temporal* CLT. Let us also mention that in [**?**] a temporal CLT over quadratic irrational rotations and $\mathbb{R}^d$ valued, piecewise constant functions with rational discontinuities, is shown to hold along subsequences and in [**?**] a temporal DLT is proved for windings of horocycle flows on non-compact hyperbolic surfaces with finite area.

While we wrote this paper specifically for deterministic random walks driven by rotations, there are other entropy zero dynamical systems where this formalism applies and for which one can prove temporal limit theorems using similar techniques. For example, in work in progress, we can prove temporal limit theorems also for certain linear flows on infinite translation surfaces and some cocycles over interval exchange transformation and more in general for certain $\mathcal{S}-adic$ systems (which are non-stationary generalizations of substitution systems, see [**?**]).

1.4. **Proof tools and sketch and outline of the paper.** In Section 2 we introduce the renormalization algorithm that we use, as a key tool in the proofs: this is essentially the classical multiplicative continued fraction algorithm, with additional data which records the relative position of the break point β of the function f_β under renormalization. This renormalization acts on the underlying parameter space to be defined in what follows, as a (skew-product) extension of the Gauss map, and it produces simultaneously the continued fraction expansion entries of α and the Ostrowski expansion entries of β. Variations on this skew product have been studied by several authors (see in particular [**?**, **?**]) and it is well known that it is related to a section of the diagonal flow on the space of affine lattices (as explained in detail in [**?**]). In sections 2.4 and 2.5 we explain how the renormalization algorithm provides a way of encoding dynamics symbolically in terms of a Markov chain. More precisely, the dynamics of the map R_α we are interested in translates in symbolic language to the adic or Vershik dynamics (on a Bratelli diagram given by the Markov chain), as explained in section. The original function f_β defines under renormalization a sequence of induced functions (which correspond to Birkhoff sums of the function f_β at first return times, called special Birkhoff sums in the terminology introduced by [**?**]). The Birkhoff sums of the function f_β can be then decomposed into sums of special Birkhoff sums. This formalism and the symbolic coding allows to translate the study of the temporal visit distribution random variable to the study of a non-homogeneous Markov chain, see section 2.6. In Section 3 we provide sufficient conditions for a non-homogeneous Markov chain to satisfy the CLT. Finally, in Section 4 we prove that these conditions are satisfied for the Markov chain modeling the temporal distribution random variables.

2. RENORMALIZATION

2.1. **Preliminaries on continued fraction expansions and circle rotations.** Let $\mathcal{G}$:$(0,1) \to (0,1)$ be the Gauss map, given by $\mathcal{G}(x) = \{1/x\}$, where $\{\cdot\}$ denotes the fractional part. Recall that a regular continued fraction expansion of $\alpha \in (0,1) \setminus \mathbb{Q}$ is given by

$$\alpha = \cfrac{1}{a_0 + \cfrac{1}{a_1 + \dots}}$$

where $a_i := a(\alpha_i) = \left[\frac{1}{\alpha_i}\right]$ and $\alpha_i := G^i(\alpha) = \left\{\frac{1}{\alpha_{i-1}}\right\}$. In this case we write $\alpha = [a_0, a_1, ...]$. Setting $q_{-1} = 1$, $q_0 = a_0$, $q_n = a_n q_{n-1} + q_{n-2}$ for $n \geq 1$, and $p_{-1} = 0$, $p_0 = 1$, $p_n = a_n p_{n-1} + p_{n-2}$ for $n \geq 1$ we have $\gcd(p_n, q_n) = 1$ and

$$\cfrac{1}{a_0 + \cfrac{1}{a_1 + \dots \cfrac{1}{a_n}}} = \frac{p_n}{q_n}.$$

Let $\alpha \in (0,1) \setminus \mathbb{Q}$, $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ and $R_\alpha : \mathbb{T} \to \mathbb{T}$ be the irrational rotation given by$R_\alpha := x + \alpha \mod 1$. Then the Denjoy-Koksma inequality [**?**, **?**] states that if $f : \mathbb{T} \to \mathbb{R}$ is a function of bounded variation, then for any $n \in \mathbb{N}$,

$$\sup\{|f \circ R_\alpha^{q_n}(x)| : x \in \mathbb{T}\} \leq \bigvee\nolimits_{\mathbb{T}} f \tag{2.1}$$

where $\bigvee_{\mathbb{T}} f$ is the variation of f on $\mathbb{T}$.

In this section we define the dynamical renormalization algorithm we use in this paper, which is an extension of the classical continued fraction algorithm and hence of the Gauss map. This algorithm gives a dynamical interpretation of the notion of *Ostrowski expansion* of β relative to α in non-homogeneous Diophantine approximation. We mostly follow the conventions of the paper [**?**] by Arnoux and Fisher, in which the connection between this renormalization and homogeneous dynamics (in particular the geodesic flow on the space of lattices with a marked point, which is also known as *the scenery flow)* is highlighted. As in [**?**] we use a different convention for rotations on the circle. Let $\alpha \in (0,1) \setminus \mathbb{Q}$, $I = [-1, \alpha)$ and let $T_\alpha(x) : [-1, \alpha) \to [-1, \alpha)$ be defined by

$$T_\alpha(x) = \begin{cases} x + \alpha & x \in [-1, 0) \\ x - 1 & x \in [0, \alpha) \end{cases} \tag{2.2}$$

Note that T_α may also be viewed as a rotation on the circle $\mathbb{R}/\sim$ where the equivalence relation $\sim$ on $\mathbb{R}$ is given by $x \sim y \iff x - y \in (1+\alpha)\mathbb{Z}$. It is conjugate to the standard rotation $R_{\alpha'}$ on $\mathbb{T}$, where $\alpha' = \frac{\alpha}{1+\alpha}$, by the map

$$\psi(x) = (\alpha' + 1)\, x - 1 \tag{2.3}$$

which maps the unit interval $[0,1]$ to the interval $[-1,\alpha]$.

Remark 2.1. In what follows, we slightly abuse notation by not distinguishing between the transformation T_α and the transformation defined similarly on the interval $(-1,\alpha]$ by

$$T'_\alpha(x) = \begin{cases} x+\alpha & x \in (-1,0]\,; \\ x-1 & x \in [0,\alpha)\,; \end{cases}$$

when viewed as transformations on the circle, T_α and $T_{\alpha'}$ coincide.

Note that given an irrational rotation R_α, we can assume without loss of generality that $\alpha < \frac{1}{2}$ (otherwise consider the inverse rotation by $1-\alpha$). If we set

$$\alpha_0 := \frac{\alpha}{1-\alpha} \tag{2.4}$$

then $\mathcal{G}(\alpha) = \mathcal{G}^i(\alpha_0)$ for any $i \in \mathbb{N}$ and thus, apart from the first entry, the continued fraction entries of α and α_0 coincide. If a_0, a'_0 are correspondingly the first entries in the expansion of α_0 and α, then $a_0 = a'_0 + 1$. Furthermore, given $\beta \in (0,1)$, let

$$\beta_0 := (\alpha+1)\,\beta - 1. \tag{2.5}$$

Then the mean zero with a discontinuity at β_0, given by

$$\varphi(x) = \mathbf{1}_{[-1,\beta_0)}(x) - \frac{\beta_0+1}{\alpha_0+1} \tag{2.6}$$

is the function that corresponds to the function f_β in the introduction under the conjugation between R_α and T_{α_0}. Therefore, we are interested in the Birkhoff sums

$$\varphi_n(x) = \sum_{k=0}^{n-1} \varphi\left(T^k_{\alpha_0}(x)\right). \tag{2.7}$$

Henceforth, unless explicitly stated otherwise, we work with the transformation T_{α_0}. The sequences $(a_n)_{n=0}^\infty$, $\left(\frac{p_n}{q_n}\right)_{n=0}^\infty$ will correspond to the sequence of entries and the sequence of partial convergents in the continued fraction expansion of $\alpha = \frac{\alpha_0}{1+\alpha_0}$.

We denote by λ the Lebesgue measure on $[-1,\alpha_0)$ normalized to have total mass 1.

2.2. **Continued fraction renormalization and Ostrowski expansion.** The renormalization procedure is an inductive procedure, where at each stage we induce the original transformation T_{α_0} onto a subinterval of the interval we induced upon at the previous stage. We denote by $I^{(n)}$ the nested sequence of intervals which we induce upon, and by $T^{(n)}$ the first return map of T_{α_0} onto $I^{(n)}$. The nested sequence of intervals $I^{(n)}$ is chosen in such a way that the induced transformations $T^{(n)}$ are all irrational rotations. The next paragraph describes a step of induction given an irrational rotation T_{α_n} on the interval $I_n = [-1,\alpha_n)$ defined by (2.2). The procedure is then iterated recursively by rescaling and performing the induction step once again. In general, we keep to the convention that we use n as a superscript to denote objects related to the non-rescaled nth step of renormalization, and as a subscript for the rescaled version.

One step of renormalization. For an irrational $\alpha_n \in (0,1)$ let $I_n := [-1, \alpha_n)$, $T_{\alpha_n} : I_n \to I_n$ defined by the formula in (2.2) and $\beta_n \in I_n$. Then T_{α_n} is an exchange of two intervals of lengths α_0 and 1 respectively (namely $[0, \alpha_0)$ and $[-1, 0)$). The renormalization step consists of inducing T_{α_n} onto an interval I'_n, where I'_n is obtained by *cutting* a half-open interval of size α_0 from the left endpoint of the interval I_0, i.e. -1, as many times as possible in order to obtain an interval of the from $[-\alpha'_n, \alpha_n)$ containing zero. More precisely, let

$$a_n = [1/\alpha_n], \qquad \alpha'_n = 1 - a_n\alpha_n$$

so that $[-1, 0)$ contains exactly a_n intervals of lengths α_n plus an additional remainder of length $0 < \alpha'_n < \alpha_n$ (see Figure 2.1). If $\beta_n \in [-1, -\alpha'_n)$, let $1 \leq b_n \leq a_n$ be such that β_n belongs to the b_n^{th} copy of the interval which is cut, otherwise set $b_n := 0$, i.e. define

$$b_n := \begin{cases} [(\beta_n - (-1))/\alpha_n] + 1 = [(1 + \beta_n)/\alpha_n] + 1 & \text{if } \beta_n \in [-1, -\alpha'_n) \\ 0 & \text{if } \beta_n \in [-\alpha'_n, \alpha_n). \end{cases} \tag{2.8}$$

FIGURE 2.1. One step of the Ostrowski renormalization algorithm.

For $b_n \geq 1$, let us define x_n to be the left endpoint of the copy of the interval which contains β_n, otherwise, if $b_n = 0$, set $x_n := 0$; let also $\beta'_n := \beta_n - x_n$, so that if $b_n \geq 1$ then β'_n is the distance of β_n from the left endpoint of the interval which contains it (Figure 2.1). In formulas

$$x_n := \begin{cases} -1 + (b_n - 1)\alpha_n & \text{if } b_n \geq 1 \\ 0 & \text{if } b_n = 0 \end{cases}, \qquad \beta'_n := \begin{cases} \beta_n + 1 - (b_n - 1)\alpha_n & \text{if } b_n \geq 1 \\ \beta_n & \text{if } b_n = 0 \end{cases}. \tag{2.9}$$

Notice that $x_n = T_{\alpha_n}^{b_n}(0)$ and hence in particular it belongs to the segment $\{0, T_{\alpha_n}(0), \dots, T_{\alpha_n}^{a_n}(0)\}$ of the orbit of 0 under T_{α_n}.

Let $I'_n = [-\alpha'_n, \alpha_n)$ and note that $\beta'_n \in I'_n$ and that the induced transformation obtained as the first return map of T_{α_n} on I'_n is again an exchange of two intervals, a short one $[-\alpha'_n, 0)$ and a long one $[0, \alpha_n)$. Hence, if we *renormalize* and *flip* the picture by multiplying by $-\alpha_n$, the interval I'_n is mapped to $I_{n+1} := (-1, \alpha_{n+1}]$, where

$$\alpha_{n+1} := \frac{\alpha'_n}{\alpha_n} = \frac{1 - a_n\alpha_n}{\alpha_n} = \frac{1}{\alpha_n} - \left[\frac{1}{\alpha_n}\right] = G(\alpha_n)$$

and the transformation T_{α_n} as first return on the interval I'_n is conjugated to $T_{\alpha_{n+1}}$. We then set

$$\beta_{n+1} := -\frac{\beta'_n}{\alpha_n} = -\frac{\beta_n - x_n}{\alpha_n},$$

so that $\beta_{n+1} \in I_{n+1}$. Thus we have defined α_{n+1}, β_{n+1} and $T_{\alpha_{n+1}}$ and completed the description of the step of induction.

Notice that by definition of β'_n and β_{n+1}we have that

$$\beta_n = \begin{cases} -1 + (b_n - 1)\,\alpha_n - \alpha_n\beta_{n+1} = x_n - \alpha_n\beta_{n+1} & \text{if } b_n \geq 1; \\ -\alpha_n\beta_{n+1} & \text{if } b_n = 0; \end{cases} \tag{2.10}$$

and hence, by using equation (2.8), we get

$$\beta_{n+1} = \mathcal{H}(\alpha_n, \beta_n) := \begin{cases} -\left\{\frac{\beta_n + 1}{\alpha_n}\right\} & \text{if } b_n \geq 1; \\ -\frac{\beta_n}{\alpha_n} & \text{if } b_n = 0. \end{cases} \tag{2.11}$$

Repeating the described procedure inductively, one can prove by induction the assertions summarized in the next proposition.

Proposition 2.2. *Let $\alpha^{(n)} := \alpha_0 \cdot ... \cdot \alpha_n$ where α_i are defined inductively from α_0 by $\alpha_n = \mathcal{G}(\alpha_{n-1})$ and set $\alpha^{(-1)} = 1$. Define a sequence of nested intervals $I^{(n)}$, $n = 0, 1, ...$, by $I^{(0)} := \left[-1, \alpha^{(0)}\right)$, and*

$$I^{(n)} := \begin{cases} \left(-\alpha^{(n-1)}, \alpha^{(n)}\right] & \text{if } n \text{ is odd;} \\ \left[-\alpha^{(n)}, \alpha^{(n-1)}\right) & \text{if } n \text{ is even.} \end{cases}$$

The induced map $T^{(n)}$ of T_{α_0} on $I^{(n)}$ is conjugated to T_{α_n} on the interval $I_n = [-1, \alpha_n)$ if n is even or to T_{α_n} on $I_n = (-1, \alpha_n]$ if n is odd, where the conjugacy is given by $\psi_n : I_n \to I^{(n)}$, $\psi_n(x) = (-1)^n \alpha^{(n-1)}(x)$.

Let $\beta_0 \in I^{(0)}$ and let $(b_n)_n$ and $(\beta_n)_n$ be the sequences defined inductively[1] by the formulas (2.8) and (2.11). Then we have

$$\beta_0 = \sum_{n=0}^{\infty} x^{(n)}, \qquad \text{where } x^{(n)} = \psi_n(x_n) = \begin{cases} (-1)^n \alpha^{(n-1)}(-1 + (b_n - 1)\alpha_n) & 1 \le b_n \le a_n \\ 0 & b_n = 0 \end{cases} \tag{2.12}$$

and the reminders are given by

$$\left|\beta_0 - \sum_{k=0}^{n} x^{(k)}\right| = \left|\beta^{(n+1)}\right|, \qquad \text{where } \beta^{(n)} := \psi_n(\beta_n) = (-1)^n \alpha^{(n-1)} \beta_n. \tag{2.13}$$

The expansion in (2.12) is an *Ostrowski type* expansion for β_0 in terms of α_0. We call the integers b_n the *entries* in the Ostrowski expansion of β_0.

Remark 2.3. Partial approximations in the Ostrowski expansions have the following dynamical interpretation. It well known that, for any $n \in \mathbb{N}$, the finite segment$\left\{T^i_{\alpha_0}(0) : \; i = 0, ..., q_n + q_{n-1} - 1\right\}$ of the orbit of 0 under T_{α_0} (which can be thought of as a rotation on a circle) induces a partition of $[-1, \alpha_0)$ into intervals of two lengths (see for example [**?**]; these partitions correspond to the classical Rokhlin-Kakutani representation of a rotation as two towers over an induced rotation given by the Gauss map, see also Section 2.3 and Remark 2.6). The finite Ostrowski approximation $\sum_{k=0}^{n} x^{(k)}$ gives one of the endpoints of the unique interval of this partition which contains β_0 (if it is the left or the right one depends on the parity as well as on whether b_n is zero or not). In particular, we have that

$$\sum_{k=0}^{n} x^{(k)} \in \left\{T^i_{\alpha_0}(0) : \; i = 0, ..., q_n + q_{n-1} - 1\right\} \cup \{\alpha_0\}, \qquad n \in \mathbb{N} \cup \{0\}.$$

Remark 2.4. Since the points $\alpha^{(n)}$ are all in the orbit of the point 0 by the rotation T_{α_0}, it follows from the correspondence between T_{α_0} and R_α that the Ostrowski expansion of β_0 appearing in the previous proposition is finite, i.e. $\beta_0 = \sum_{n=0}^{N} x^{(n)}$ for some $N \in \mathbb{N}$ if and only if $\beta \in \{n\alpha \mod 1 : \; n \in \mathbb{Z}\}$. This condition is well known to be equivalent to the function f_β (and hence also φ) being a coboundary (see [**?**]) .

It follows from the description of the renormalization algorithm that

$$(\alpha_{n+1}, \beta_{n+1}) = \hat{\mathcal{G}}(\alpha_n, \beta_n) := (\mathcal{G}(\alpha_n), \mathcal{H}(\alpha_n, \beta_n))$$

where the function $\mathcal{H}$ is defined by (2.11). The ergodic properties of a variation on the map

$$\hat{\mathcal{G}} : X \to X, \quad X = \{(\alpha, \beta) : \; \alpha \in [0, 1) \setminus \mathbb{Q}, \; \beta \in [-1, \alpha)\} \tag{2.14}$$

were studied among others in [**?**].

Introduce the functions $a, b : X \to \mathbb{N}$ defined by

$$a(\alpha, \beta) := [1/\alpha], \qquad b(\alpha, \beta) := \begin{cases} [(1 + \beta)/\alpha] + 1 & \beta \in [-1, -1 + a(\alpha, \beta)\alpha) \\ 0 & \beta \in [-1 + a(\alpha, \beta)\alpha, \alpha) \end{cases}.$$

The functions are defined so that the sequences $(a_n)_n$ and $(b_n)_n$ of continued fractions and Ostrowski entries are respectively given by $a_n = a\left(\hat{\mathcal{G}}^n(\alpha_0, \beta_0)\right)$, $b_n = b\left(\hat{\mathcal{G}}^n(\alpha_0, \beta_0)\right)$ for any $n \in \mathbb{N}$.

[1]Note that given β_n, formulas (2.8) and (2.11) determine first b_n and then, as function of β_n and b_n, also β_{n+1} and hence b_{n+1}.

By Remark 2.4, the restriction of the space X to

$$\tilde{X} := \left\{ (\alpha_0, \beta_0) \in X : \ \beta \notin \{n\alpha \mod 1\} \text{ for } \alpha = \frac{\alpha_0}{\alpha_0 + 1}, \ \beta = \frac{\beta_0 + 1}{\alpha + 1} \right\} \tag{2.15}$$

is invariant with respect to $\hat{G}$ and we partition this space into three sets X_G, X_{B_-}, $X_{B_+} \subset \tilde{X}$ defined by

$$\begin{array}{ll} X_G := \{(\alpha, \beta) : \ b(\alpha, \beta) \geq 1\} & X_B := \{(\alpha, \beta) : \ b(\alpha, \beta) = 0\} \\ X_{B_+} := X_B \cap \{(\alpha, \beta) : \ \beta \geq 0\} & X_{B_-} := X_B \cap \{(\alpha, \beta) : \ \beta < 0\} \end{array} \tag{2.16}$$

Explicitly, in terms of the relative position of α, β, these sets are given by

$$\begin{aligned} X_G &= \left\{ (\alpha, \beta) \in \tilde{X} : \ \beta \in [-1, -1 + a(\alpha, \beta)\ \alpha) \right\}, \\ X_{B_-} &= \left\{ (\alpha, \beta) \in \tilde{X} : \ \beta \in [-1 + a(\alpha, \beta)\ \alpha, 0) \right\}, \\ X_{B_+} &= \left\{ (\alpha, \beta) \in \tilde{X} : \ \beta \in [0, \alpha) \right\}. \end{aligned}$$

The reason for the choice of names G, B_-, B_+ for the thee parts of parameter space, which stand for *Good* (G) and *Bad* (B), where *Bad* has two subcases, B_- and B_+ (according to whether β is positive or negative), will be made clear in Section 4.1.

2.3. **Description of the Kakutani-Rokhlin towers obtained from renormalization.** We assume throughout the present Section and Sections 2.4, 2.5 that we are given a fixed pair $(\alpha_0, \beta_0) \in \tilde{X}$. The symbols q_n used in this Section refer to the denominators of the n^{th} convergent in the continued fraction expansion of α, where α is related to α_0 via (2.4).

The renormalization algorithm described above defines a nested sequence of intervals $I^{(n)}$. We describe here below how the original transformation T_{α_0} can be represented as a union of *towers* in a *Kakutani skyscraper* (the definition is given below) with base $I^{(n)}$; the tower structure of the skyscraper corresponding to the $(n+1)^{th}$ stage of renormalization is obtained from the towers of the previous skyscraper corresponding to the n^{th} stage by a *cutting and stacking* procedure. We will use these towers to describe what we call an *adic* symbolic coding of the interval $I = [-1, \alpha_0)$ (see section 2.4). In what follows, we give a detailed description of the tower structure and the coding.

Let us first recall that if a measurable set $B \subset [-1, \alpha_0)$ and a positive integer h are such that the union $\bigcup_{i=0}^{h-1} T_{\alpha_0}^i B$ is disjoint, we say that the union is a *(dynamical) tower* of base B and height h. The union can indeed be represented as a tower with h floors, namely $T_{\alpha_0}^i B$ for $i = 0, \ldots, h-1$, so that T_{α_0} acts by mapping each point in each level except the last one, to the point directly above it. A disjoint union of towers is called a *skyscraper* (see for example [**?**]). A *subtower* of a tower of base B and height h is a tower with the same height whose base is a subset $B' \subset B$.

As it was explained in the previous section, the induced map of T_{α_0} on $I^{(n)}$ is an exchange of two intervals, a *long* and a *short* one. If n is even, the long one is given by $\left[-\alpha^{(n-1)}, 0\right)$ and the short one by $\left[0, \alpha^{(n)}\right)$. If n is odd the long and short interval are respectively given by $\left[0, \alpha^{(n-1)}\right)$ and $\left[-\alpha^{(n)}, 0\right)$. In both cases, these are the preimages of the intervals $[-1, 0)$ and $[0, \alpha_n)$ under the conjugacy map $\psi_n : I^{(n)} \to I_n$ given in Proposition 2.2. Notice also that $\beta^{(n)} = \psi_n^{-1}(\beta_n)$, the non rescaled marked point corresponding to the point $\beta_n \in I_n$, further divides the two mentioned subintervals of $I^{(n)}$ into three, by cutting either the long or the short into two subintervals. We denote these three intervals $I_M^{(n)}, I_L^{(n)}$ and $I_S^{(n)}$, where the letters M, L, S, respectively correspond to *middle* (M), *long* (L) and *short* (S), and $I_M^{(n)}$ denotes the middle interval, while $I_L^{(n)}$ and $I_S^{(n)}$ denote (what is left of) the long one and the short one, after removing the middle interval. Explicitly, it is convenient to describe the intervals in terms of the partition X_G, X_{B_-}, X_{B_+} defined in the end of the previous section. Thus, set

$$I_L^{(n)} = \psi_n^{-1}\left([-1,\beta_n)\right), \quad I_M^{(n)} = \psi_n^{-1}\left([\beta_n,0)\right), \quad I_S^{(n)} = \psi_n^{-1}\left([0,\alpha_n)\right) \quad \text{if } (\alpha_n,\beta_n) \in X_G \bigcup X_{B_-},$$
$$I_L^{(n)} = \psi_n^{-1}\left([-1,0)\right), \quad I_M^{(n)} = \psi_n^{-1}\left([0,\beta_n)\right), \quad I_S^{(n)} = \psi_n^{-1}\left[\beta_n,\alpha_n\right) \quad \text{if } (\alpha_n,\beta_n) \in X_{B_+}.$$

We claim that the first return time of T_{α_0} to the interval $I^{(n)}$ is constant on the subintervals $I_L^{(n)}$, $I_M^{(n)}$ and $I_S^{(n)}$. Moreover, the first return time over $I_L^{(n)}$ and $I_S^{(n)}$ equals to q_n and q_{n-1} respectively, while the first return time over $I_M^{(n)}$ equals either q_n or q_{n-1}, depending on whether $\beta^{(n)} \in \left[-\alpha^{(n-1)},0\right)$ or $\beta^{(n)} \in \left[0,\alpha^{(n)}\right)$ and hence on whether the middle interval was cut from the long or the short interval respectively. For $J \in \{L,M,S\}$, let us denote by $h_J^{(n)}$ the first return time of $I_J^{(n)}$ to $I^{(n)}$ under T_{α_0} and let us denote by $Z_J^{(n)}$ the tower with base $I_J^{(n)}$ and height $h_J^{(n)}$.

Let us how describe how the tower structure at stage $n+1$ of the renormalization is related to the tower structure at stage n. We will describe in detail as an example the particular case where n is odd and $\beta^{(n)} \in \left[-\alpha^{(n-1)},-\alpha^{(n)}\right)$ (i.e. $\beta^{(n)} \notin I^{(n+1)}$), or equivalently $(\alpha_n,\beta_n) \in X_G$ (see also Figure 2.2). The other cases are summarized in Proposition 2.5 below. In the considered case, the heights $h_J^{(n)}$of the three towers $Z_J^{(n)}$, $J \in \{L,M,S\}$, at stage n are given by $h_J^{(n)} = q_n$ for $J \in \{M,L\}$ and $h_S^{(n)} = q_{n-1}$. By the structure of the first return map $T^{(n)}$, the intervals $\left(T^{(n)}\right)^i\left(I_S^{(n)}\right)$, $i=1,...,a_n$ partition the interval $\left[-\alpha^{(n-1)},-\alpha^{(n)}\right) = I^{(n)} \setminus I^{(n+1)}$ into intervals of equal length, and it follows that the first return time of T_{α_0} is constant on $I_S^{(n)}$ and equals to

$$a_n \cdot h_L^{(n)} + h_S^{(n-1)} = a_n q_n + q_{n-1} = q_{n+1}.$$

FIGURE 2.2. The tower structure at step n and $n+1$ in the case when $(\alpha_n,\beta_n) \in X_G$. In this example $a_n = 4$ and $b_n = 3$.

It also follows that the tower over $I_S^{(n)} \subset I^{(n+1)}$ at stage $n+1$ is obtained by stacking the subtowers over the intervals $\left(T^{(n)}\right)^i\left(I_S^{(n)}\right)$ on top of the tower $Z_S^{(n)}$ (as shown in Figure 2.2). By construction, the point $\beta^{(n+1)}$ is obtained by vertically projecting the point $\beta^{(n)}$ from its location in the tower over $I_S^{(n)}$ down to the interval $I_S^{(n)}$. According to our definitions, $\beta^{(n+1)}$ divides $I_S^{(n)}$ into $I_L^{(n+1)} = \left[\beta^{(n)},\alpha^{(n-1)}\right)$ and $I_M^{(n+1)} = \left[0,\beta^{(n)}\right)$. As we have seen, the height of the towers at stage $n+1$ over the intervals $I_M^{(n+1)}$ and $I_L^{(n+1)}$ is the same and equals q_{n+1}, but the composition of the towers is different. The tower $Z_M^{(n+1)}$ is obtained by stacking, on top of the bottom tower $Z_S^{(n)}$, first b_n subtowers of $Z_L^{(n)}$ and then $a_n - b_n$ subtowers of $Z_M^{(n)}$ on top of them; $Z_L^{(n+1)}$ has a similar structure, with the tower $Z_S^{(n)}$ in the bottom, but with $b_n - 1$ subtowers of $Z_L^{(n)}$ on top and then $a_n - b_n + 1$ subtowers of $Z_M^{(n)}$ stacked over (see Figure 2.2). The tower over $I_S^{(n+1)} = I_M^{(n)}$ remains unchanged, i.e $Z_S^{(n+1)} = Z_M^{(n)}$.

It is convenient to describe the tower structure in the language of *substitutions*. Let us recall that a substitution τ on a finite alphabet $\mathcal{A}$ is a map which associates to each letter of $\mathcal{A}$ a finite word in the alphabet $\mathcal{A}$. To each (α,β) with β rational or $\alpha,\beta,1$ linearly independent over $\mathbb{Q}$, we associate a sequence $(\tau_n)_n$ of substitutions over the alphabet $\{L,M,S\}$, where for $J \in \{L,M,S\}$,

$$\tau_n(J) = J_0 J_1 \cdots J_k, \qquad \text{where } J, J_0, \ldots, J_k \in \{L,M,S\},$$

if and only if the tower $Z_J^{(n+1)}$ consists of subtowers of $Z_{J_i}^{(n)}$, $i=0,...,k$ stacked on top of each other in the specified order, i.e. the subtower of $Z_{J_{i+1}}^{(n)}$ is stacked on top of $Z_{J_i}^{(n)}$. More formally,

(2.17)

$$\tau_n(J) = J_0 J_1 \cdots J_k \qquad \Leftrightarrow \qquad h_J^{(n+1)} = \sum_{j=0}^{k} h_{J_j}^{(n)} \quad \textit{and} \quad (T^{(n)})^i(x) \in I_{J_i}^{(n)} \quad \forall x \in I_J^{(n+1)}, \quad i=0,...,k.$$

For example, in the case discussed above, since the tower $Z_M^{(n+1)}$ is obtained by stacking, on top of each other, in order, $Z_S^{(n)}$, then b_n subtowers of $Z_L^{(n)}$ and then $a_n - b_n$ subtowers of $Z_M^{(n)}$, we have

$$\tau_n(M) = S\underbrace{L\cdots L}_{b_n \text{ times}}\underbrace{M\cdots M}_{a_n-b_n \text{ times}}.$$

We will use the convention of writing J^n for the block $J\cdots J$ where the symbol J is repeated n times. With this convention, the above substitution can be written $\tau_n(M) = SL^{b_n}M^{a_n-b_n}$.

If ω is a word $\omega = J_0J_1\cdots J_k$where we will denote by ω_i the letter indexed by $0 \le i<|\omega|$. Using this notation, we can rewrite (2.17) as

$$h_J^{(n+1)} = \sum_{i=0}^{|\omega|-1} h^{(n)}_{\tau_n(J)_i}.$$

We summarize the tower structure and the associated sequence of substitutions in the following proposition. The substitution τ_n is determined by the location of $\beta^{(n)} \in I^{(n)}$, or equivalently, by the non-rescaled parameters (α_n, β_n) and one can check that there are three separate cases corresponding to the parameters being in X_G, X_{B_-} or X_{B_+}. One of the cases was analyzed in the discussion above, while the other cases can be deduced similarly, and the proof of the proposition is a straightforward induction on n.

Proposition 2.5. *The first return time function of T_{α_0} to $I^{(n)}$ is constant on each of the three intervals $I_J^{(n)}$, $J \in \{L, M, S\}$. Thus, for $n = 0, 1, 2, ...$,*

$$I^{(0)} = \bigcup_{J\in\{L,M,S\}} Z_J^{(n)} \qquad \text{where } Z_J^{(n)} = \bigcup_{i=0}^{h_J^{(n)}-1} T^i_{\alpha_0} I_J^{(n)}$$

where $h_J^{(n)}$ is the value of the first return time function on $I_j^{(n)}$, which is given by

$$h_L^{(n)} = q_n, \quad h_S^{(n)} = q_{n-1}, \quad h_M^{(n)} = \begin{cases} q_n & if\ \beta_n \in [-1, 0) \\ q_{n-1} & if\ \beta_n \in [0, \alpha_n) \end{cases}.$$

The sequence of substitutions associated to the pair (α, β) is given by the formulas, determined by the following cases

- *If $(\alpha_n, \beta_n) \in X_G$*
$$\begin{cases} \tau_n(L) = SL^{b_n-1}M^{a_n-b_n+1} \\ \tau_n(M) = SL^{b_n}M^{a_n-b_n} \\ \tau_n(S) = M \end{cases}$$
- *If $(\alpha_n, \beta_n) \in X_{B_-}$*
$$\begin{cases} \tau_n(L) = SL^{a_n} \\ \tau_n(M) = M \\ \tau_n(S) = L \end{cases}$$
- *If $(\alpha_n, \beta_n) \in X_{B_+}$*
$$\begin{cases} \tau_n(L) = SL^{a_n} \\ \tau_n(M) = ML^{a_n} \\ \tau_n(S) = L \end{cases}$$

Remark 2.6. It can be shown that due to irrationality of α, the levels of the towers $Z_J^{(n)}$, $J \in \{L, M, S\}$ form an increasing sequence of partitions that separates points and hence generates the Borel σ-algebra on $[-1, \alpha_0)$ (see for example [**?**]).

Let A_n, $n \in \mathbb{N}$, be the 3×3 *incidence matrix* of the substitution τ_n with entries indexed by $\{L, M, S\}$, where the entry indexed by (J_1, J_2) , which we will denote by $(A_n)_{J_1,J_2}$, gives the number of subtowers contained in $Z_{J_2}^{(n)}$ among the subtowers of level n which are stacked to form the tower $Z_{J_1}^{(n+1)}$. Equivalently, the entry $(A_n)_{J_1,J_2}$ gives the number of occurrences of the letter J_2 in the word $\tau_n(J_1)$. If we adopt the convention that the order of rows/columns of A_n corresponds to L, M, S, it follows from Proposition 2.5 that these matrices are then explicitly given by:

$$A_n = \begin{pmatrix} b_n - 1 & a_n - b_n + 1 & 1 \\ b_n & a_n - b_n & 1 \\ 0 & 1 & 0 \end{pmatrix} \qquad \text{if } (\alpha_n, \beta_n) \in G, \tag{2.18}$$

$$A_n = \begin{pmatrix} a_n & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \text{if } (\alpha_n, \beta_n) \in B_-, \tag{2.19}$$

$$A_n = \begin{pmatrix} a_n & 0 & 1 \\ a_n & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \qquad \text{if } (\alpha_n, \beta_n) \in B_+. \tag{2.20}$$

In particular, if we denote by $h^{(n)}$ the column vector of heights towers, i.e. the transpose of $\left(h_L^{(n)}, h_M^{(n)}, h_S^{(n)}\right)$, it satisfies the recursive relations

$$h^{(n+1)} = A_n h^{(n)}, \qquad n \in \mathbb{N} \cup \{0\} . \tag{2.21}$$

Remark 2.7. We remark briefly for the readers familiar with the *Vershik adic map* and the S-adic formalism (even though it will play no role in the rest of this paper), that the sequence $(\tau_n)_n$ also allows to represent the map T_{α_0} as a *Vershik adic map.* The associated Bratteli diagram is a non-stationary diagram, whose vertex sets V_n are always indexed by $\{L, M, S\}$ with $(A_n)_{J_1,J_2}$ edges from J_1 to J_k; the ordering of the edges which enter the vertex J at level n is given exactly by the substitution word $\tau_n(J)$. We refer the interested reader to the works by Vershik [**?**] and to the survey paper by Berthe and Delecroix [**?**] for further information on Vershik maps, Brattelli diagrams and S-adic formalism.

2.3.1. *Special Birkhoff sums.* Let us consider now the function φ defined by (2.6) which has a discontinuity at 0 and at β_0. In order to study its Birkhoff sums φ_n (defined in (2.7)) , we will use the renormalization algorithm described in the previous section. Under the assumption that $(\alpha, \beta) \in \tilde{X}$, φ determines a sequence of functions $\varphi^{(n)}$, where $\varphi^{(n)}$ is a real valued function defined on $I^{(n)}$ obtained by *inducing* φ on $I^{(n)}$, i.e. by setting

$$\varphi^{(n)}(x) = \sum_{i=0}^{h_J^{(n)}-1} \varphi({T_\alpha}^i(x)), \qquad \text{if } x \in I_J^{(n)}. \tag{2.22}$$

The function $\varphi^{(n)}$ is what Marmi-Moussa-Yoccoz in [**?**] started calling *special Birkhoff sums*: the value $\varphi^{(n)}(x)$ gives the Birkhoff sum of the function φ along the orbit of $x \in I_J^{(n)}$ until its first return to $I^{(n)}$, i.e. it represents the Birkhoff sum of the function along an orbit which goes from the bottom to the top of the tower $Z_J^{(n)}$.

One can see that since $\varphi^{(0)} := \varphi$ has mean zero and a discontinuity with a jump of 1 at $\beta^{(0)} := \beta_0$, its special Birkhoff sums $\varphi^{(n)}$, $n \in \mathbb{N}$, again have mean zero and a discontinuity with a jump of 1. The points $\beta^{(n)}$, $n \in \mathbb{N}$, are defined in the renormalization procedure exactly so that $\varphi^{(n)}$ has a jump of one at $\beta^{(n)}$. Moreover, the function φ is constant on each level of the towers $Z_J^{(n)}$, $J \in \{L, M, S\}$, $n \in \mathbb{N} \cup \{0\}$, and therefore, it is completely determined by a sequence of vectors

$$\left(\varphi_L^{(n)}, \varphi_M^{(n)}, \varphi_S^{(n)}\right), \qquad n \in \mathbb{N} \cup 0,$$

where $\varphi_J^{(n)} = \varphi^{(n)}(x)$, for any $x \in I_J^{(n)}$. It then follows immediately from the towers recursive structure (see equation (2.17)) that the functions $\varphi^{(n)}$ also satisfy the following recursive formulas given by the substitutions in Proposition 2.5:

$$\varphi_J^{(n+1)} = \sum_{i=0}^{k} \varphi_{J_i}^{(n)} \qquad \text{if } \tau_n(J) = J_0 \cdots J_k.$$

We finish this section with a few simple observations on the heights of the towers and on special Birkhoff sums along these towers that we will need for the proof of the main result. Let $(\alpha_0, \beta_0) \in \tilde{X}$ be the parameters associated to a given pair (α, β) via the relations (2.4) and (2.5). Under the assumption that α is badly approximable, since the heights of the towers appearing in the renormalization procedure satisfy (2.21) and $0 \le b_n \le a_n$ are bounded, there exists a constant C such that

$$C^{-1} n \le \log h_J^{(n)} \le Cn \qquad \text{for any } n \in \mathbb{N}. \tag{2.23}$$

It follows that for any $m \in \mathbb{N}$, there exists a constant $M = M(m)$, such that if $|k - n| \le m$, then

$$\frac{1}{M} < \frac{h_J^{(n)}}{h_K^{(k)}} < M \qquad \text{for any } J, K \in \{L, M, S\}. \tag{2.24}$$

Moreover, by (2.1), the special Birkhoff $\varphi_J^{(n)}$ are uniformly bounded, i.e.

$$\sup\left\{\left|\varphi_J^{(n)}\right| : \ J \in \{L, M, S\}, \ n \in \mathbb{N}\right\} < \infty. \tag{2.25}$$

2.4. **The (adic) symbolic coding.** The renormalization algorithm and the formalism defined above lead to the symbolic coding of the dynamics of T_{α_0} described in the present section. This coding is exploited in Section 2.5 to build an array of non-homogeneous Markov chains which models the dynamics.

Definition 2.8. (Markov compactum) Let $(\mathcal{S}_n)_{n=1}^{\infty}$ be a sequence of finite sets with $\sup_i |\mathcal{S}_i| < \infty$ and let $\left(A^{(n)}\right)_{n=1}^{\infty}$ be a sequence of matrices, such that $A^{(n)}$ is an $|\mathcal{S}_n| \times |\mathcal{S}_{n+1}|$ matrix whose entries $A_{s,t}^{(n)} \in \{0, 1\}$ for any $(s, t) \in \mathcal{S}_n \times \mathcal{S}_{n+1}$. The *Markov compactum* determined by $A^{(n)}$ is the space

$$X = \left\{\omega \in \prod_{n=1}^{\infty} \mathcal{S}_n : \ A_{s_n, s_{n+1}}^{(n)} = 1 \text{ for all } n \in \mathbb{N}\right\}.$$

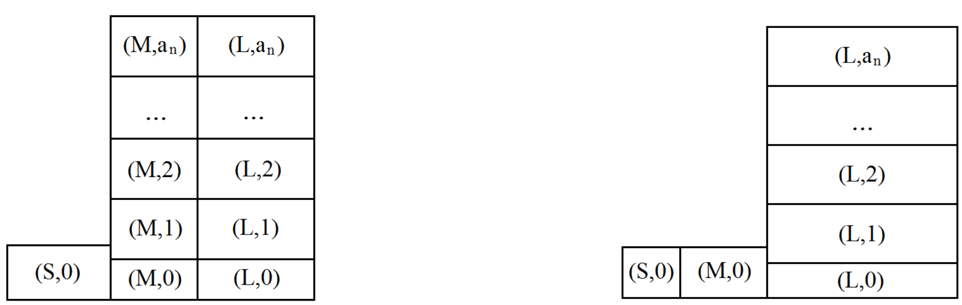


FIGURE 2.3. Labeling of the subtowers of the towers$Z_J^{(n)}$ by labels (J, i), $J \in \{L, M, S\}$.

To describe the coding, recall that for each $n \in \mathbb{N}$, each tower $Z_J^{(n)}$, where $J \in \{L, M, S\}$, is obtained by stacking at most $a_n + 1$ *subtowers* of the towers $Z_K^{(n-1)}$ (the type and order of the subtowers is completely determined by the word $\tau_{n-1}(J)$ given by the substitution τ_{n-1} as described in Proposition 2.5). We will *label* these subtowers by (J, i), where the index i satisfies $0 \le i \le a_n$ and indexes the subtowers *from bottom* *to top*: more formally, (J,i) is the label of the subtower of $Z_J^{(n)}$, with base $(T^{(n-1)})^i(I_J^{(n)})$, which is the $(i+1)^{th}$ subtower from the bottom (see Figure 2.3). Thus, for a fixed n, denoting by $|\tau_{n-1}(J)|$ the length of the word $\tau_{n-1}(J)$, the labels of the subtowers belong to

$$\{(J,i): \quad J\in\{L,M,S\},\ i=0,...,|\tau_{n-1}(J)|-1\}.$$

For $\alpha=\frac{\alpha_0}{1+\alpha_0}=[a_0,a_1,\ldots,a_n,\ldots]$ badly approximable, let a_{max} be the largest of its continued fraction entries and consider the alphabet

$$E=E(a_{max})=\{L,M,S\}\times\{0,\cdots,a_{max}\}.$$

Remark 2.9. It is not necessary for α to be badly approximable in order for the construction of the present section and the next section to be valid. If α is not badly approximable, define $E=\{L,M,S\}\times\{0,1,...,n,...\}$. This definition would make all statements of this and the following sections valid, without any further changes.

Definition 2.10. Given $x\in[-1,\alpha_0)$, for each $n\in\mathbb{N}$, x is contained in a unique tower $Z^{(n)}_{J_n(x)}$ for some $J_n(x)\in\{L,M,S\}$, and furthermore in a unique subtower of stage $n-1$ inside it, labeled by $(J_n(x),j_n(x))$ where $0\le j_n(x)\le a_n$. Let $\Psi:[-1,\alpha_0)\to E$ be the coding map defined by

$$\Psi(x):=(J_n(x),j_n(x))\in E^{\mathbb{N}}.$$

Let us recall that for word ω in the alphabet E let us denote by ω_i the letter in the word which is labeled by $0\le i<|\omega|$.

Proposition 2.11. *The image of Ψ is contained in the subspace $\Sigma\subset E^{\mathbb{N}}$ defined by*

$$\Sigma:=\left\{((J_1,j_1),...,(J_n,j_n),...)\in E^{\mathbb{N}}:\ (\tau_i(J_{i+1}))_{j_{i+1}}=J_i,\ i=1,2,...\right\}.$$

The preimage under Ψ of any cylinder $[(J_1,j_1),...,(J_n,j_n)]:=\{\omega\in\Sigma:\ \omega_i=(J_i,j_i),\ i=1,...,n\}$ satisfying the constraints $(\tau_i(J_{i+1}))_{j_{i+1}}=J_i,\ i=1,2,...,n-1$ is the set of all points on some level of the tower $Z^{(n)}_{J_n}$, i.e. there exists $0\le i<h^{(n)}_{J_n}$ such that

$$\Psi^{-1}([(J_1,j_1),...,(J_n,j_n)])=T^i_{\alpha_0}\left(I^{(n)}_{J_n}\right).$$

Moreover, Ψ is a Borel isomorphism between $[-1,\alpha_0)$ and its image, where the Borel structure on the image of Ψ is inherited from the natural Borel structure on $E^{\mathbb{N}}$ arising from the product topology on $E^{\mathbb{N}}$.

Let

$$\mathcal{S}_n:=\{(J,j)\in E:\ \exists\omega\in\Sigma\ s.t.\ \omega_n=(K,k)\}\tag{2.26}$$

be the set of symbols which appear as n^{th} coordinate in some admissible word in Σ, and note that the definition of Σ shows that Σ is a Markov compactum with state space $\prod_{i=1}^{\infty}\mathcal{S}_i$, given by a sequence of matrices $\left(A^{(n)}\right)^T$ indexed by $\mathcal{S}_n\times\mathcal{S}_{n+1}$ such that $A^{(n)}_{(K,k),(J,j)}=1$ if and only if $(\tau_n(J))_j=K$. Although we do not need it in what follows, one can explicitly describe the image $\Sigma'\subset\Sigma$ of the coding map Ψ and show that it is obtained from Σ by removing countably many sequences. We remarked in Remark 2.7 that T_α is conjugated to a the Vershik adic map. Let us add that the map Ψ provides the measure theoretical conjugacy.

Proof of Proposition 2.11. First we prove that the image of Ψ is contained in Σ. To see this, note that for $x\in[-1,\alpha_0)$, $(J_n(x),j_n(x))=(K,k)$ means that x belongs to $Z^{(n)}_K$ (since $J_n(x)=K$) and $(J_{n+1}(x),j_{n+1}(x))=(J,j)$ means that x belongs to the j^{th} subtower of $Z^{(n+1)}_J$. Hence the j^{th} subtower of $Z^{(n+1)}_J$ must be contained in $Z^{(n)}_K$. Recalling the definition of the substitutions $(\tau_n)_n$, this implies exactly the relation $(\tau_n(J))_j=K$, which in turn implies that $\Psi(x)\in\Sigma$.

To prove the second statement, namely that cylinders correspond to floors of towers, note that according to our labeling of the towers, the set of all x such that $(J_n(x),j_n(x))=(J_n,j_n)$ consists exactly of all points which belong to $h^{(n-1)}_K$ levels of the tower $Z^{(n)}_{J_n}$, where $K=(\tau_{n-1}(J_n))_{j_n}$. Proceeding by induction, one sees that for any $k = n-1, \dots, 1$, the set $\{x : (J_i(x), j_i(x)) = (J_i, j_i), \ i = k, ..., n\}$ is the set of all points contained in precisely $h_K^{(k-1)}$ levels of the tower $Z_{J_n}^{(n)}$, where $K = (\tau_{k-1}(J_k))_{j_k}$. Thus, since $h_K^{(0)} = 1$ for any $K \in \{L, M, S\}$, $\Psi^{-1}([(J_1, j_1), ..., (J_n, j_n)])$ is the set of all points on a *single* level of the tower $Z_{J_n}^{(n)}$. This argument shows that the levels of the towers $Z_J^{(n)}$, $J \in \{L, M, S\}$, are in bijective correspondence under the map Ψ with cylinders of length n in Σ.

Finally, injectivity and bi-measurability of Ψ follow since the sequence of partitions induced by the tower structure generate the Borel sets of the space $[-1, \alpha_0)$ and separates points (see Remark 2.6). □

2.5. **The Markov chain modeling towers.** In what follows, we denote by μ be the push forward by the map Ψ of the normalized Lebesgue measure λ on $[-1, \alpha_0)$, i.e. the measure given by

$$\mu := \lambda \circ \Psi^{-1}. \tag{2.27}$$

Moreover, for $J \in \{L, M, S\}$, let us define also the conditional measures

$$\mu_n^J([(J_1, j_1), ..., (J_n, j_n)]) := \mu([(J_1, j_1), ..., (J_n, j_n)] \,|J_n = J). \tag{2.28}$$

We denote by Σ_n and E_n correspondingly, the restriction of Σ and E to the first n coordinates and we endow these sets with the σ-algebras inherited from the Borel σ-algebra on $E^{\mathbb{N}}$. Let $\mathcal{S}_n$ be the set of states appearing in the n^{th} coordinate of Σ, defined by (2.26).

We define a sequence of transition probabilities, or equivallently in this discrete case, stochastic matrices $p^{(n)}_{(J,j),(K,k)}$, where $(J, j) \in \mathcal{S}_{n+1}$ and $(K, k) \in \mathcal{S}_n$, and a sequence of probability distributions π_n on $\mathcal{S}_n$ which are used to define a sequence of Markovian measures on Σ_n that model the dynamical renormalization procedure. We refer to the sequence $p^{(n)}$ as the sequence of *transition matrices associated to the pair* $(\alpha_0, \beta_0) \in \tilde{X}$.

Definition 2.12. For any $n \in \mathbb{N}$ and J, $K \in \{L, M, S\}$, if

$$\tau_n(J) = J_0 \dots J_l \text{ and } \tau_{n-1}(K) = K_0 \dots K_m,$$

we define

$$p^{(n)}_{(J,j),(K,k)} := \begin{cases} \frac{h^{(n-1)}_{K_k}}{h^{(n)}_K} & \text{if } K = J_j \text{ and } 0 \le j \le l, \ k \le m, \\ 0 & \text{otherwise;} \end{cases}$$

$$\pi_n(K, k) := \begin{cases} \lambda\left(Z_K^{(n)}\right) \cdot \frac{h^{(n-1)}_{K_k}}{h^{(n)}_K} & \text{if } 0 \le k \le m, \\ 0 & \text{otherwise.} \end{cases}$$

Moreover, for any $(L, l) \in E$, we set

$$\pi_n^K(L, l) := \begin{cases} \frac{h^{(n-1)}_{K_l}}{h^{(n)}_K} & \text{if } 0 \le l \le m \text{ and } L = K, \\ 0 & \text{otherwise.} \end{cases}.$$

Remark 2.13. The rationale behind the definition of π_n is that $\pi_n(K, k)$ is defined to be the λ- measure of the piece of the tower $Z_K^{(n)}$ labeled by (K, k); similarly $p^{(n)}_{(J,j),(K,k)}$ is non zero exactly when the k^{th} subtower inside $Z_J^{(n)}$ is contained in $Z_K^{(n-1)}$, in which case it gives the proportion of this k^{th} subtower which is contained in the subtower of $Z_K^{(n-1)}$ labeled by (J, j).

The following Proposition identifies the measures μ_n and μ_n^J as Markovian measures on Σ_n generated by the transition matrices and initial distributions indicated in the previous definition.

Proposition 2.14. *For every $n \in \mathbb{N}$, $J \in \{L, M, S\}$ and every word $((J_1, j_1), ..., (J_n, j_n)) \in \Sigma_n$ we have*

$$\mu([(J_1, j_1), ..., (J_n, j_n)]) = \pi_n(J_n, j_n) \prod_{i=1}^{n-1} p^{(i)}_{(J_{i+1}, j_{i+1}),(J_i, j_i)} \tag{2.29}$$

and

$$\mu\left(\left[\left(J_1,j_1\right),...,\left(J_n,j_n\right)\right]|J_n=J\right)=\pi_n^J\left(J_n,j_n\right)\prod_{i=1}^{n-1}p^{(i)}_{(J_{i+1},j_{i+1}),(J_i,j_i)}. \tag{2.30}$$

Proof. By Proposition 2.11, $\Psi^{-1}\left(\left[\left(J_1,j_1\right),...,\left(J_n,j_n\right)\right]\right)$ is non empty if and only if the sequence $\left(J_1,j_1\right),...,\left(J_n,j_n\right)$ satisfies the conditions $\left(\tau_i\left(J_{i+1}\right)\right)_{j_{i+1}}=J_i$ for $i=1,...,n-1$, in which case it consists of the the set of all points on a certain level of the tower $Z^{(n)}_{J_n}$, i.e.

$$\Psi^{-1}\left(\left[\left(J_1,j_1\right),...,\left(J_n,j_n\right)\right]\right)=T^1_{\alpha_0}\left(I^{(n)}_{J_m}\right)\quad\text{for some } 0\le l\le h^{(n)}_{J_n}.$$

It follows from the definition (2.27) of the measure μ that

$$\mu\left(\left[\left(J_1,j_1\right),...,\left(J_n,j_n\right)\right]\right)=\lambda\left(I^{(n)}_{J_n}\right)=\frac{\lambda\left(Z^{(n)}_{J_n}\right)}{h^{(n)}_{J_n}}.$$

Moreover, we get that the conditional measures μ^J_n given by (2.28) satisfy

$$\mu^J_n\left(\left[\left(J_1,j_1\right),...,\left(J_n,j_n\right)\right]\right)=\mu\left(\left[\left(J_1,j_1\right),...,\left(J_n,j_n\right)\right]|J_n=J\right)=\frac{1}{h^{(n)}_{J_n}}.$$

Equations (2.29) now follow by definition of π_n and $p^{(n)}$, which give that

$$p^{(i)}_{(J_{i+1},j_{i+1}),(J_i,j_i)}=\frac{h^{(i-1)}_{\tau_{i-1}(J_i)_{j_i}}}{h^{(i)}_{\tau_i(J_{i+1})_{j_{i+1}}}},\qquad \pi_n\left(J_n,j_n\right)=\lambda\left(Z^{(n)}_{J_n}\right)\cdot\frac{h^{(n-1)}_{\tau_{n-1}(J_n)_{j_n}}}{h^{(n)}_{J_n}}.$$

Hence, by the conditions $\left(\tau_i\left(J_{i+1}\right)\right)_{j_{i+1}}=J_i$ for $i=1,...,n-1$ and recalling that $J_n=J$ and $h^{(0)}_K=1$ for any $K\in\{L,M,S\}$, we have that

$$\pi_n\left(J_n,j_n\right)\prod_{i=1}^{n-1}p^{(i)}_{(J_{i+1},j_{i+1}),(J_i,j_i)}=\lambda\left(Z^{(n)}_{J_n}\right)\cdot\frac{h^{(n-1)}_{J_{n-1}}}{h^{(n)}_{J_n}}\left(\prod_{i=2}^{n-1}\frac{h^{(i-1)}_{J_{i-1}}}{h^{(i)}_{J_i}}\right)\frac{1}{h^{(1)}_{J_1}}=\lambda\left(Z^{(n)}_J\right).$$

Equations (2.30) follows in the same way by using the definition of π^J_n instead than π_n .

Finally, if the sequence $\left(J_1,j_1\right),...,\left(J_n,j_n\right)$ does not satisfy the conditions $\left(\tau_i\left(J_{i+1}\right)\right)_{j_{i+1}}=J_i$ for $i=1,...,n-1$, $\Psi^{-1}\left(\left[\left(J_1,j_1\right),...,\left(J_n,j_n\right)\right]\right)=\emptyset$ (by Proposition 2.11, as recalled above) and by definition of $p^{(n)}$ and π_n, π^J_n, we get that the right hand sides in (2.4) and (2.32) are both zero, so equations (2.4) and (2.32) hold in this case too. This completes the proof. □

For $\omega\in\Sigma$, $n\in\mathbb{N}$, we define the *coordinate random variables*

$$X_n\left(\omega\right):=\omega_n \tag{2.31}$$

Since, all cylinders of the form $\left[\left(J_1,j_1\right),...,\left(J_n,j_n\right)\right]$, with $\left(\left(J_1,j_1\right),...,\left(J_n,j_n\right)\right)\in\Sigma_n$ generate the σ-algebra of Σ_n, it immediately follows from Proposition 2.14 that for every $n\in\mathbb{N}$, $X_n,...,X_1$ form a Markov chain on Σ_n with respect to the measures μ, μ^J_n with transition probabilities $p^{(i)}$, $i=1,...,n-1$ and initial distributions π_n , π^J_n, respectively.

2.6. **The functions over the Markov chain modeling the Birkhoff sums.** Let us now define a sequence of functions $\xi_n:S_n\to\mathbb{R}$ that enables us to model the distribution of Birkhoff sums. In section 2.3.1 we introduced the notion of special Birkhoff sums of φ, i.e. Birkhoff sums of φ along the orbit of a point x in the base of a renormalization tower $Z^{(n)}_J$ up to the height of the tower, see (2.22). We will consider in this section *intermediate Birkhoff sums along a tower* (of for short, *intermediate Birkhoff sums*), namely Birkhoff sums of a point at the base of a tower $Z^{(n)}_J$ up to an intermediate height, i.e. sums of the form

$$\sum_{k=0}^{j}\varphi\left(T^k_{\alpha_0}\left(x\right)\right),\qquad\text{where}\quad x\in I^{(n)}_J,\ 0\le j<h^{(n)}_J,\qquad n\in\mathbb{N}\cup\{0\},\ J\in\{L,M,S\}.$$

The crucial Proposition (2.16) shows that intermediate Birkhoff sums can be expressed as sums of the following functions $\{\xi_n\}$ over the Markov chain (X_n).

Definition 2.15. For $n \in \mathbb{N}$, $(J,j) \in S_n$ such that $\tau_{n-1}(J) = J_0 \dots J_l$ (note that this forces $0 \le j \le l = |\tau_{n-1}(J)|$), if $n \ge 2$ set

$$\xi_n((J,j)) = \sum_{i=0}^{j-1} \varphi_{J_i}^{(n-1)}$$

where, by convention, a sum with i that runs from 0 to -1 is equal to zero. If $n=1$, set

$$\xi_1((J,j)) = \sum_{i=0}^{j} \varphi_{J_i}^{(0)}.$$

We then have the following proposition.

Proposition 2.16. *Let $J \in \{L,M,S\}$ and let ,$x_J \in I_J^{(n)}$. Then for any $A \in \mathcal{B}(\mathbb{R})$,*

$$\frac{1}{h_J^{(n)}} \# \left\{ 0 \le j \le h_J^{(n)} - 1 : \sum_{k=0}^{j} \varphi\left(T_{\alpha_0}^k(x_J)\right) \in A \right\} = \mu_n^J \left(\sum_{k=1}^{n} \xi_k(X_k) \in A \right).$$

Proof. We show by induction on n that, for any $J \in \{L,M,S\}$, any $x \in I_J^{(n)}$ and $0 \le l \le h_J^{(n)} - 1$, we have that

$$\sum_{k=0}^{l} \varphi\left(T_{\alpha_0}^k(x)\right) \in A \iff \sum_{k=1}^{n} \xi_k((J_k, j_k)) \in A. \tag{2.32}$$

where $[\omega] = [(J_1,j_1), ..., (J_n,j_n)]$ is the (unique) cylinder containing $T_{\alpha_0}^l x$. To see this, note first that for $n=1$, $J \in \{L,M,S\}$, $x \in I_J^{(1)}$ and $0 \le l \le h_J^{(1)} - 1$, we have $T^l x \in \Psi^{-1}([(J,l)])$ and by definition of ξ_1,

$$\xi_1((J,l)) = \sum_{k=1}^{l} \varphi_{(\tau_0(J))_k}^{(0)} = \sum_{k=0}^{l-1} \varphi\left(T_{\alpha_0}^k(x)\right),$$

which proves the claim for $n=1$. Now, assume that (2.32) holds for some $n \in \mathbb{N}$. Then for $J \in \{L,M,S\}$, $x \in I_J^{(n+1)}$, and $0 \le l \le h_J^{(n+1)} - 1$, let $[\omega] = [(J_1,j_1), ..., (J_{n+1}, j_{n+1})]$ be the unique cylinder such that $T_{\alpha_0}^l(x) \in \Psi^{-1}([\omega])$. Then $J_n = J$ and by Proposition 2.11 $\Psi^{-1}([\omega]) = T_{\alpha_0}^l(I_J)$. It follows from definition of the map Ψ, that $J_{n+1} = J$ and

$$\sum_{i=0}^{j_{n+1}-1} h_{(\tau(J))_i}^{(n)} < l \le \sum_{i=0}^{j_{n+1}} h_{(\tau(J))_i}^{(n)}.$$

Thus, setting $l' = l - \sum_{i=1}^{j_{n+1}-1} h_{(\tau(J))_i}^{(n)}$, $x' = \left(T^{(n)}\right)^{j_{n+1}}(x)$ and using the definition of ξ_{n+1}, we may write

$$\sum_{k=0}^{l} \varphi\left(T_{\alpha_0}^k(x)\right) = \sum_{k=0}^{j_{n+1}-1} \varphi_{(\tau_n(J))_k}^{(n)} + \sum_{k=0}^{l'} \varphi\left(T_{\alpha_0}^k(x')\right) = \xi_{n+1}((J, j_{n+1})) + \sum_{k=0}^{l'} \varphi\left(T_{\alpha_0}^k(x')\right). \tag{2.33}$$

The previous equality is obtained by splitting the Birkhoff sum up to l of a point at the base of the tower $Z_{J_{n+1}}^{(n+1)}$ into special Birkhoff sums over towers obtained at the n^{th} stage of the renormalization procedure and a remainder given by $\sum_{k=0}^{l'} \varphi\left(T_{\alpha_0}^k(x')\right)$. Now, by definition of the coding map Ψ, $T_{\alpha_0}^{l'}(x') \in \Psi^{-1}([(J_1,j_1), ..., (J_n,j_n)])$. Thus, if for $y \in \mathbb{R}$, we let $A - y$ denote the set $\{a - y : a \in A\}$, (2.33) implies,

$$\sum_{k=0}^{l} \varphi\left(T_{\alpha_0}^k(x)\right) \in A \iff \sum_{k=0}^{l'} \varphi\left(T_{\alpha_0}^k(x')\right) \in A - \xi_{n+1}(J, j_{n+1}).$$

and the equality (2.32) now follows from the hypothesis of induction, which gives

$$\sum_{k=0}^{l'}\varphi\left(T_{\alpha_0}^{k}\left(x'\right)\right)\in A-\xi_{n+1}\left(J,j_{n+1}\right)\iff\sum_{k=1}^{n}\xi_{k}\left(X_{k}\left(J_{k},j_{k}\right)\right)\in A-\xi_{n+1}\left(J,j_{n+1}\right).$$

Since by Proposition 2.14, $\mu_n^J\left([\omega]\right)=\frac{1}{h_J^{(n)}}$ for any $\omega\in\Sigma_n$, and since by the proof of Proposition 2.11, the levels of the tower $Z_J^{(n)}$ are in bijective correspondence with cylinders of length n in Σ_n, the proof is complete. □

3. THE CLT FOR MARKOV CHAINS

In the previous section we established that the study of intermediate Birkhoff sums can be reduced to the study of (in general) non-homogeneous Markov chains. In this section we establish some (mostly well-known) statements about such Markov chains which we use in the proof of our temporal CLT. The main result which we need is the CLT for non-homogeneous Markov chains. To the best of our knowledge, this was initially established by Dobrushin [?, ?] (see also [?] for a proof using martingale approximations). Dobrushin's CLT is not directly valid in our case (since it assumes that the contraction coefficient is strictly less than 1 for every transition matrix in the underlying chain, while under our assumptions this is only valid for a product of a constant number of matrices). While the proof of Dobrushin's theorem can be reworked to apply to our assumptions, we do not do it here, and instead use a general CLT for φ-mixing triangular arrays of random variables by Utev [?].

3.1. **Contraction coefficients, mixing properties and CLT for Markov chains.** In this section we collect some probability theory results for (arrays of) non-homogeneous Markov chains that we will use in the next section.

Let $(\Omega,\mathcal{B},P)$ be a probability space and let $\mathcal{F}$, $\mathcal{G}$ be two sub σ-algebras of $\mathcal{B}$. For any σ-algebra $\mathcal{A}\subset\mathcal{B}$, denote by $\mathcal{L}^2\left(\mathcal{A}\right)$ the space of square integrable, real functions on Ω, which are measurable with respect to $\mathcal{A}$. We use two measures of dependence between $\mathcal{F}$ and $\mathcal{G}$, the so called φ-coefficient and ρ-coefficient, defined by

$$\varphi\left(\mathcal{F},\mathcal{G}\right):=\sup_{A\in\mathcal{G},B\in\mathcal{F}}\left|P\left(A\,|B\right)-P\left(A\right)\right|$$

and

$$\rho\left(\mathcal{F},\mathcal{G}\right):=\sup_{f\in\mathcal{L}^2(\mathcal{F}),g\in\mathcal{L}^2(\mathcal{G})}\left|\frac{Cov\left(f,g\right)}{\sqrt{Var\left(f\right)Var\left(g\right)}}\right|.$$

It is a well-known fact (see [?]) that

$$\rho\left(\mathcal{F},\mathcal{G}\right)\leq 2\left(\varphi\left(\mathcal{F},\mathcal{G}\right)\right)^{\frac{1}{2}}. \tag{3.1}$$

In what follows, let $Y=\left\{Y_1^{(n)},...,Y_n^{(n)}:n\geq 1\right\}$ be a triangular array of mean zero, square integrable random variables such that the random variables in each row are defined on the same probability space $(\Omega,\mathcal{B},P)$. For any set $\mathcal{Y}$ of random variables defined on $(\Omega,\mathcal{B},P)$, let us denote by $\sigma\left(\mathcal{Y}\right)$ the σ-algebra generated by all the random variables in $\mathcal{Y}$.

Set $S_n=\sum_{k=1}^{n}Y_k^{(n)}$ and $e_n=E\left(S_n\right)$, $\sigma_n=\sqrt{Var\left(S_n\right)}$. For any $n,k\in\mathbb{N}$ let

$$\varphi_n\left(k\right):=\sup_{1\leq s,s+k\leq n}\varphi\left(\sigma\left(Y_i^{(n)},i\leq s\right),\sigma\left(Y_i^{(n)},i\geq s+k\right)\right),$$

$$\varphi\left(k\right):=\sup_{n}\varphi_n\left(k\right).$$

The array Y is said to be *φ-mixing* if $\varphi\left(k\right)\to 0$ as k tends to infinity.

The following CLT for φ-mixing arrays of random variables, which follows from a more general CLT for such arrays in [?] is the main result that we use to prove our distributional CLT.

Theorem 3.1. *Let Y be a φ-mixing array of square integrable random variables and assume that*

$$\lim_{n\to\infty} \sigma_n^{-2} \sum_{k=1}^{n} E\left(Y_k^{(n)} \mathbf{1}_{\left\{\left|Y_k^{(n)}\right| > \epsilon\sigma_n\right\}} \right) = 0 \tag{3.2}$$

for every $\epsilon > 0$. Then

$$\frac{S_n - e_n}{\sigma_n}$$

converges in law to the standard normal distribution.

Let T, S be finite sets and P a stochastic matrix with entries indexed by $T \times S$. The contraction coefficient of P is defined by

$$\tau(P) = \frac{1}{2} \sup_{x_1, x_2 \in T} \sum_{s\in S} \left| P_{x_1,s} - P_{x_2,s} \right|. \tag{3.3}$$

It is not difficult to see that $\tau(P) = 0$ if and only if the entry $P_{s,t}$ does not depend on s and that

$$\tau(PQ) \leq \tau(P)\,\tau(Q) \tag{3.4}$$

for any pair of stochastic matrices P and Q such that their product is defined.

For $n \in \mathbb{N}$, let $X_1^{(n)}, ..., X_n^{(n)}$ be a Markov chain with each $X_i^{(n)}$ taking values in a finite state space $\mathcal{S}_i$, determined by an initial distribution π_n and transition matrices $P_i^{(n)}$, $i = 1, ...n$ (thus, each matrix $P_i^{(n)}$ has dimension $|\mathcal{S}_i| \times |\mathcal{S}_{i+1}|$).

Proposition 3.2. *Assume that there exist $0 \leq \delta < 1$ and $s \in \mathbb{N}$ such that for every $n \in \mathbb{N}$*

$$\tau\left(P_k^{(n)} \cdots P_{k+s}^{(n)}\right) < \delta, \qquad \textit{for any } 1 \leq k \leq n - s.$$

Then $X = X_0^{(n)}, ..., X_n^{(n)}$ is φ-mixing and $\varphi(k)$ tends to 0 as $k \to \infty$ with exponential rate.

Proof. This is a direct consequence of the inequality

$$\varphi\left(\sigma\left(Y_i^{(n)}, i \leq j\right), \sigma\left(Y_i^{(n)}, i \geq j+k\right)\right) \leq \tau\left(P_j^{(n)} ... P_{j+k}^{(n)}\right) \qquad \text{for } 1 \leq j \leq n-k$$

(see relation (1.1.2) and Proposition 1.2.5 in [?]) and the fact that $\tau\left(P_j^{(n)} ... P_{j+k}^{(n)}\right) \leq \delta^{\left[\frac{k}{s}\right]}$, which immediately follows from the assumption and (3.4). □

Now, let $\xi_i^{(n)} : \mathcal{S}_i \to \mathbb{R}$, with $1 \leq i \leq n$ for any $n \in \mathbb{N}$, be an array of functions and set $Y_i^{(n)} = \xi_i^{(n)}\left(X_i^{(n)}\right)$. Henceforth, we assume that

$$\sup\left\{\left|\xi_i^{(n)}(s)\right| : \, n \in \mathbb{N}, \, i = 0, ..., n, \, s \in \mathcal{S}_i\right\} = M < \infty. \tag{3.5}$$

An application of Theorem 3.1 yields the following corollary.

Corollary 3.3. *Under the conditions of Proposition 3.2, assume further that and $\sigma_n \to \infty$. Then $\frac{S_n - e_n}{\sigma_n}$ converges in law to the standard normal distribution.*

Proof. It is enough to remark that the condition (3.2) in Theorem 3.1 holds trivially for n large in virtue of the bound in (3.5) since by assumption the variance $\sigma_n \to \infty$. □

Let now $\tilde{\pi}_n$ be a sequence of probability distributions on $\mathcal{S}_1$, and let $\tilde{X}_1^{(n)}, ..., \tilde{X}_n^{(n)}$ be an array of Markov chains generated by initial distributions $\tilde{\pi}_n$ and transition matrices $P_i^{(n)}$. Let $\tilde{S}_n = \sum_{k=0}^{n-1} \xi_i\left(\tilde{X}_i\right)$ and let $\tilde{e}_n = E\left(\tilde{S}_n\right)$, $\tilde{\sigma}_n = \sqrt{Var\left(\tilde{S}_n\right)}$.

Proposition 3.4. *Under the conditions of Proposition 3.2, there exists a constant C, independent of the sequences π_n and $\tilde{\pi}_n$, such that $|e_n - \tilde{e}_n| \leq C$ and $\left|\sigma_n^2 - \tilde{\sigma}_n^2\right| \leq C$ for all $n \in \mathbb{N}$.*

Proof. The assumption implies that there exists a constant M and a sequence of rank 1 stochastic matrices (i.e stochastic matrices with all rows being identical) $V_i^{(n)}$ such that

$$\sup\left\{\left\|V_i^{(n)}-\prod_{j=1}^{i}P_j^{(n)}\right\|,\ i=1,...,n,\ n\in\mathbb{N}\right\}\leq M\delta^{\frac{i}{s}}$$

(see [?, Chapter 4, Cor. 2]), where for two matrices P, Q indexed by $S\times T$, $\|P-Q\|=\max\{|P_{s,t}-Q_{s,t}|:\ (s,t)\in S\times T\}$. Using (3.5) it follows that there exists a constant $\tilde{C}$ which depends only on the array of matrices $P_i^{(n)}$ and functions $\xi_i^{(n)}$, such that

$$\left|E\left(\xi_i^{(n)}\left(X_i\right)\right)-E\left(\xi_i^{(n)}\left(\tilde{X}_i\right)\right)\right|=\left|\sum_{s\in S_1}\sum_{t\in S_i}\left(\pi_n\left(s\right)-\tilde{\pi}_n\left(s\right)\right)\left(P_1^{(n)}...P_{i-1}^{(n)}\right)_{s,t}\cdot\xi_i^{(n)}\left(t\right)\right|\leq\tilde{C}\delta^{\frac{n}{s}}.$$

Since the right hand side of the last inequality is a general term of a summable geometric series, we have proved that there exists a constant C, such that $|e_n-\tilde{e}_n|\leq C$ for all $n\in\mathbb{N}$.

To prove the inequality for the variances, we first note that it follows from (3.1) and (3.5) that there exists a constant C' independent of π_n, such that

$$\left|\sum_{1\leq i<j\leq n}Cov\left(\xi_i^{(n)}\left(X_i^{(n)}\right)\cdot\xi_j^{(n)}\left(X_j^{(n)}\right)\right)\right|<C' \tag{3.6}$$

for all $n\in\mathbb{N}$. An analogous inequality hence holds also for the array $\tilde{X}_i^{(n)}$ instead of $\tilde{X}_i^{(n)}$, so that

$$\left|\sum_{1\leq i<j\leq n}Cov\left(X_i^{(n)},X_j^{(n)}\right)-Cov\left(\tilde{X}_i^{(n)},\tilde{X}_j^{(n)}\right)\right|<2C'. \tag{3.7}$$

Moreover, since $\sup_n|\mu_n-\tilde{\mu}_n|<\infty$, one can also prove that

$$\sup_n\left|\sum_{i=1}^{n}Var\left(\xi_i^{(n)}\left(X_i^{(n)}\right)\right)-Var\left(\xi_i^{(n)}\left(\tilde{X}_i^{(n)}\right)\right)\right|<\infty. \tag{3.8}$$

Now, write

$$\begin{aligned}\left|\sigma_n^2-\tilde{\sigma}_n^2\right|\leq&\left|\sum_{i=1}^{n}\left(Var\left(\xi_i^{(n)}\left(X_i^{(n)}\right)\right)-Var\left(\xi_i^{(n)}\left(\tilde{X}_i^{(n)}\right)\right)\right)\right|\\&+\left|\sum_{1\leq i<j\leq n}Cov\left(\xi_i^{(n)}\left(X_i^{(n)}\right),\xi_j^{(n)}\left(X_j^{(n)}\right)\right)-Cov\left(\xi_i^{(n)}\left(\tilde{X}_i^{(n)}\right),\xi_j^{(n)}\left(\tilde{X}_j^{(n)}\right)\right)\right|.\end{aligned}$$

The proof of the Lemma hence follows by (3.7) and (3.8). □

4. PROOF OF THE TEMPORAL CLT

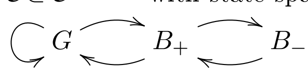
In this section we give the proof of Theorem 1.1. We need to show that we can apply the results on Markov chains summarized in the previous section (and in particular Corollary 3.3) to the Markov chains that model the dynamics. In order to check that the required assumptions are verified, we first show, in section 4.1 a result on positivity of the product of finitely many transition matrices, which follows from the assumption that α is badly approximable and β is badly approximable with respect to α. Then, in section 4.2 we prove that the variance grows. Finally, the proof of the Theorem is given in section 4.3.

4.1. **Positivity of products of incidence matrices.** Let us recall that in Section 2.2 we described a renormalization procedure that, to a pair of parameters (α,β) (under the assumption that $(\alpha,\beta)\in\tilde{X}$), in particular associates a sequence $(A_n)_n$ of matrices (given by equations (2.18), (2.19) and (2.20) respectively), which are the incidence matrices of the sequence of substitutions $(\tau_n)_n$ which describe the tower structure. In this section, we develop conditions on the pair $(\alpha,\beta)\in\tilde{X}$ that ensure that we may split the sequence of incidence matrices $(A_n)_n$ associated to (α,β) into consecutive blocks of uniformly bounded length, so that the product of matrices in each block is strictly positive. This fact is used for showing that the Markov chain associated to (α,β) satisfies the assumption of the previous section needed to prove the CLT.

Under the assumption that $(\alpha,\beta)\in\tilde{X}$, the orbit $\hat{G}^n(\alpha,\beta)$ of the point (α,β) under the transformation $\hat{G}$ defined in (2.14) is infinite and one can consider its *itinerary* with respect to the partition $\left\{X_G, X_{B_-}, X_{B_+}\right\}$ defined in Section 2.2: the itinerary is the sequence $(s_n)_n\in\mathcal{S}^{\mathbb{N}\cup\{0\}}$, where $\mathcal{S}:=\{G,B_-,B_+\}$, defined by

$$s=J\in\{G,B_-,B_+\}\iff \hat{G}^n(\alpha,\beta)\in X_J,\quad n=0,1,2,.... \tag{4.1}$$

We will call $\mathcal{S}:=\{G,B_-,B_+\}$ the set of *states* and we will say that $s(\alpha,\beta):=(s_n)_n\in\mathcal{S}\ {}^{\mathbb{N}\cup\{0\}}$ the infinite sequence of *states* associated to $(\alpha,\beta)\in\tilde{X}$. From the definitions in Section 2.2, $s_n=G$ (or B_-,B_+ respectively) if and only if the incidence matrix A_n is of the form (2.18) (or (2.19), (2.20) respectively). It can be easily deduced from the description of the renormalization procedure that not all sequences in $\mathcal{S}^{\mathbb{N}}$ are itineraries of some pair $(\alpha,\beta)\in\tilde{X}$. The sequences $s\in\mathcal{S}^{\mathbb{N}\cup\{0\}}$ such that $s=s(\alpha,\beta)$, for some $(\alpha,\beta)\in\tilde{X}$) form a stationary Markov compactum $\tilde{\mathcal{S}}\subset\mathcal{S}^{\mathbb{N}\cup\{0\}}$ with state space determined by the graph,



namely $s=(s_n)_n\in\tilde{\mathcal{S}}$ if and only if for any $n\geq 0$ there is an oriented edge from the state $s_n\in\mathcal{S}$ to the state $s_{n+1}\in\mathcal{S}$ in the graph above.

Since at this point we are interested solely in positivity of the incidence matrices and not in the values themselves, we define a function $F:\mathcal{S}\to M_3(\mathbb{Z})$, where $M_3(\mathbb{Z})$ are 3×3 matrices, by

$$\begin{aligned} F(s)&=\begin{bmatrix}0&1&1\\1&0&1\\0&1&0\end{bmatrix} &&\text{if}\quad s=G;\\ F(s)&=\begin{bmatrix}1&0&1\\0&1&0\\1&0&0\end{bmatrix} &&\text{if}\quad s{=}B_-;\\ F(s)&=\begin{bmatrix}1&0&1\\1&1&0\\1&0&0\end{bmatrix} &&\text{if}\quad s{=}B_+.\end{aligned} \tag{4.2}$$

Note that F is defined in such a way, so that some entry of the matrix $F(s_n)$ is 1, if and only if the corresponding entry of incidence matrix A_n which corresponds to the state $s(\alpha,\beta):=(s_n)_n$ has a non-zero value, independently of a_n and b_n (for example a_n and a_n-b_n+1 are always greater than 1 or $b_n\geq 1$ when $(\alpha,\beta)\in G$). Note that the other implication is not necessarily true, namely some entries of $F(s)$ could be 0 even if the corresponding entry of the incidence matrices are positive (in such cases the positivity depends on the values of a_n and b_n, for example a_n-b_n is zero if $a_n=b_n$). Thus, for any $n,k\in\mathbb{N}\bigcup\{0\}$,

$$F(s_{n+k})\,F(s_{n+k-1})\,...F(s_n)>0\implies A_{n+k}A_{n+k-1}...A_n>0.$$

It immediately follows from the topology of the transition graph that every itinerary $s\in\mathcal{S}^{\mathbb{N}\cup\{0\}}$ can be written in the form

$$s=W_1\,(B_-B_+)^{n_1}\,W_2\,(B_-B_+)^{n_2}\,...W_k\,(B_-B_+)^{n_k}\,... \tag{4.3}$$

where W_k, $k\in\mathbb{N}$ are words in the alphabet $\mathcal{S}$ which do not contain B_- (i.e. they are words in G and B_+), and W_k is not empty for $k\geq 2$. Note that it may be that the number of appearances of B_- in the above representation is finite. This means that there exists K such that $n_k = 0$ for $k \geq K$ and in this case the above representation reduces to

$$s = W_1 (B_- B_+)^{n_1} W_2 (B_- B_+)^{n_2} ... W_K (B_- B_+)^{n_K} W_{K+1} \tag{4.4}$$

where the length of W_{K+1} is infinite.

Definition 4.1. Let $(\alpha, \beta) \in \tilde{X}$. We say that β *is of Ostrowski bounded type with respect to* α if the decomposition of $s(\alpha, \beta) \in \mathcal{S}^{\mathbb{N}}$ given by (4.3) or (4.4) satisfies $\sup \{n_k\} = M < \infty$, where the supremum is taken over $k \in \mathbb{N}$ in the first case, and over $k \in \{1, ..., K\}$ in the second case. We say in both cases that β *is of Ostrowski bounded type of order* M.

Proposition 4.2. *Let β be of Ostrowski bounded type of order M with respect to α and let $(A_i)_i$ bee i the sequence of incidence matrices associated to (α, β) by the Ostrowski renormalization. Then for any k, and any $n \geq 5M$, we have that $A_{k+n} A_{k+n-1} ... A_k > 0$.*

Proof. Let $W_1 (B_- B_+)^{n_1} W_2 (B_- B_+)^{n_2} ... W_k (B_0 B_1)^{n_k} ...$ the decomposition of $s(\alpha, \beta)$ described above. Direct calculation gives that the product of matrices which corresponds to an admissible word of length 5 (or more) which does not contain B_- is strictly positive. Also, any word of length 5 which starts with $B_- B_+ G$ gives a transition matrix which is strictly positive. Note that it follows from the transition graph that each W_i, $i \geq 2$ must start with G and must be of length strictly greater than 1. Since any subword of length greater than $5M$ must contain a block of the form $B_- B_+ W_i B_-$, or a block of length at least 5 where there is no occurrence of B_-, the claim follows. □

Lemma 4.3. *If $0 < \alpha < \frac{1}{2}$ is badly approximable and $\beta \in (0, 1)$ is badly approximable with respect to α, then the pair (α_0, β_0), related to (α, β) via equations (2.4) and (2.5), satisfies $(\alpha_0, \beta_0) \in \tilde{X}$ and β_0 is of Ostrowski bounded type with respect to α_0.*

Proof. Let $\sum_{k=0}^{\infty} x^{(k)}$ be the Ostrowski expansion of β_0 in terms of α_0 given by Proposition 2.2. Then by Remark 2.3 $\sum_{k=0}^{n} x^{(k)} \in \left\{T_{\alpha_0}^j (0) : \ 0 \leq j \leq q_{n-1} + q_n\right\} \bigcup \{\alpha_0\}$ where q_n are the denominators of the n^{th} convergent in the continued fraction expansion of α. Since under the conjugacy ψ (which was defined in (2.3)) between T_{α_0} and R_α (where both maps are viewed as rotations on a circle), the (equivalence class of the) points 0 and α_0 in the domain of T_{α_0} correspond respectively to the (equivalence class of) points $1 - \alpha$ and 1 in the domain of R_α, we obtain that $\psi^{-1}\left(\sum_{k=0}^{n} x^{(k)}\right) \in \left\{R_\alpha^j (1 - \alpha) : \ 0 \leq j \leq q_{n-1} + q_n - 1\right\}$. It follows that the Ostrowski expansion of β_0 is infinite, since otherwise, if there exists an n such that $\beta_0 = \sum_{k=0}^{n} x^{(k)}$, we would get that $\beta = \psi^{-1}(\beta_0) = 1 - \alpha + j\alpha \mod 1$ for some $j \in \mathbb{N} \bigcup \{0\}$, which obviously contradicts (1.1). Thus, $(\alpha_0, \beta_0) \in \tilde{X}$.

Fix $M \in \mathbb{N}$ and let $s = s(\alpha_0, \beta_0)$ be defined by (4.1). We claim that, if for some $n \in \mathbb{N}$, $s_{n+i} \in \{B_-, B_+\}$ for all $1 \leq i \leq M$, then there exist a constant C, which does not depend on n, and $0 \leq k \leq q_n + q_{n-1}$, $p \in \mathbb{Z}$, such that

$$|\beta - (k-1)\alpha - p| \leq \frac{C}{q_{n+M}}. \tag{4.5}$$

The second assertion of the Lemma follows immediately from this and the fact that $\frac{q_{n+M}}{q_n} \to \infty$ uniformly in n as M tends to ∞.

To see that the claim holds, suppose that $s_{n+i} \in \{B_-, B_+\}$ for all $0 \leq i \leq M$. Recalling the description of the renormalization procedure in section 2.2, this is equivalent to $x^{(n+i)} = 0$ for all $0 \leq i \leq M$, so that $\sum x^{(k)} = \sum_{k=0}^{n+M} x^{(k)}$. Thus, by the estimate of the reminder in an Ostrowski expansion given by Proposition 2.2, we obtain that

$$\left|\beta_0 - \sum_{k=0}^{n} x^{(k)}\right| = \left|\beta_0 - \sum_{k=0}^{n+M} x^{(k)}\right| \leq \left|\beta^{(n+M+1)}\right| \leq \alpha^{(n+M)}.$$

Since α is badly approximable, $\alpha^{(n)} = \mathcal{G}^n(\alpha) \leq \frac{C}{q_n}$ for all n, where C is a constant which depends only on α. Since the conjugacy map ψ is affine, the previous inequality yields that there exists a constant C, such that

$$\left|\beta-\psi^{-1}\left(\sum_{k=0}^{n}x^{(k)}\right)\right|\leq\frac{C}{q_{n+M}}.$$

Since $\psi^{-1}\left(\sum_{k=0}^{n}x^{(k)}\right)\in\left\{R_{\alpha}^{j}\left(1-\alpha\right):\ 0\leq j<q_{n-1}+q_{n}\right\}$, we obtain that

$$\psi^{-1}\left(\sum_{k=0}^{n}x^{(k)}\right)=1-\alpha+k\alpha+p$$

where $0\leq k<q_{n}+q_{n-1}$, and $p\in\mathbb{Z}$. Thus, combining the last two equations, we proved (4.5). This completes the proof of the Lemma. □

Let $0<\alpha<\frac{1}{2}$ be badly approximable, let $\beta\in(0,1)$ be badly approximable with respect to α and let (α_0,β_0) be related to (α,β) via equations (2.4) and (2.5). Since by the previous proposition $(\alpha_0,\beta_0)\in\tilde{X}$, the sequence of transition matrices $p^{(n)}$ associated to the pair (α_0,β_0) given by Definition 2.12 is well defined. Recall that $\tau\left(P\right)$, where P is a stochastic matrix, denotes the contraction coefficient defined by (3.3).

Corollary 4.4. *Let $0<\alpha<\frac{1}{2}$ be badly approximable, $\beta\in(0,1)$ be badly approximable with respect to α and let (α_0,β_0) be related to (α,β) via equations (2.4) and (2.5). Then if $p^{(n)}$ is the sequence of transition matrices associated to (α_0,β_0) (see Definition 2.12), there exist $M\in\mathbb{N}$, and $0\leq\delta<1$, such that*

$$\tau\left(p^{(n+M)}\cdot p^{(n+M-1)}\cdot\ldots\cdot p^{(n)}\right)<\delta\qquad\textit{for all } n\in\mathbb{N}.$$

Proof. Lemma 4.3 implies that β_0 is of Ostrowski bounded type. By definition of the transition matrices $p^{(n)}$ (see Definition 2.12), for any $(K,k)\in S_{n+M+1}$, $(J,j)\in S_n$

$$\left(p^{(n+M)}\cdot p^{(n+M-1)}\cdot\ldots\cdot p^{(n)}\right)_{(K,k),(J,j)}>0$$

if and only if

$$\left(A_{n+M}A_{n+M-1}\ldots A_{n}\right)_{\left(\tau_{n+M}(K)\right)_k,J}>0.$$

This should be interpreted as the statement that the probability to pass from a state $(K,k)\in\mathcal{S}_{M+n+1}$ to some state $(J,j)\in\mathcal{S}_n$ is positive if and only if, the intersection of the tower $Z_J^{(n)}$ with the subtower of $Z_K^{(n+M+1)}$ labelled by (K,k) is non-empty. Thus, Proposition 4.2 implies that there exists $M\in\mathbb{N}$ such that $p^{(n+M)}\cdot\ldots\cdot p^{(n)}$ is strictly positive for any $n\in\mathbb{N}$. From α being badly approximable (see inequality (2.24)) and by the fact that by definition, every positive entry of $p^{(n+M)}\cdot\ldots\cdot p^{(n)}$ is a ratio between the heights of tower at the $(n+M)^{th}$ and n^{th} stage of the renormalization, it follows that there exists $\delta>0$ which is indpendent of n, such that every entry of $p^{(n+M)}\cdot\ldots\cdot p^{(n)}$ is not less than δ. Note that it follows from the definition of the coefficient τ (see (3.3)) that if $P_{n\times m}$ is a stochastic matrix such that there exists $\delta>0$, for which $P_{i,j}>\delta$, for all $1\leq i\leq n$, $1\leq j\leq m$, then $\tau\left(P\right)<1-\delta$. Thus, the proof is complete. □

4.2. **Growth of the variance.** In this section we consider the random variables $\xi_k\left(X_k\right)$, $k\in\mathbb{N}$, constructed in Section 2.6 (see equation (2.31) therein). Recall that the array is well defined for any given pair of parameters $(\alpha_0,\beta_0)\in\tilde{X}$ and, by the key Proposition 2.16, models Birkhoff sums over the transformation T_{α_0} of the function φ defined by (2.6), which has a jump at β_0. The goal in the present section is to show that if φ is not a coboundary, then the variance $Var_{\mu_n}\left(\sum_{k=1}^{n}\xi_k\left(X_k\right)\right)$ tends to infinity as n tends to infinity, where $Var_{\mu_n}\left(\sum_{k=1}^{n}\xi_k\left(X_k\right)\right)$ is the variance of $\sum_{k=1}^{n}\xi_k\left(X_k\right)$ with respect to the measure μ_n.

Let us first recall the definition of tightness and a criterion which characterizes coboundaries.

Definition 4.5. Let $(\Omega,\mathcal{B},P)$ be a probability space. A sequence of random variables $\{Y_n\}$ defined on Ω and taking values in a Polish space $\mathcal{P}$ is *tight* if for every $\epsilon>0$, there exists a compact set $C\subseteq\mathcal{P}$ such that $\forall n\in\mathbb{N}$, $P\left(Y_n\in C\right)>1-\epsilon$.

Let $(X,\mathcal{B},m,T)$ be a probability preserving system and let $f:X\to\mathbb{R}$ be a measurable function. We say that f is a *coboundary* if there exists a measurable function $g:X\to\mathbb{R}$ such that the equality $f\left(x\right)=g\left(x\right)-g\circ T\left(x\right)$ holds almost surely. Let us recall the following characterization of coboundaries on $\mathbb{R}$ (see [**?**]).

Theorem 4.6. *The sequence* $\left\{\sum_{k=0}^{n-1} f\circ T^k\right\}$ *is tight if and only if* f *is a coboundary.*

Set $e_n = E_{\mu_n}\left(\sum_{k=1}^n X_k\right)$, $\sigma_n = \sqrt{Var_{\mu_n}\left(\sum_{k=1}^n X_k\right)}$. We will now prove the following lemma.

Lemma 4.7. *Assume that there exists a strictly increasing sequence of positive integers* $\{n_j\}_{j=1}^\infty$ *such that*

$$\sup\left\{\sigma_{n_j}:\ j=1,2,...\right\}<\infty.$$

Then the sequence $\varphi_n=\sum_{k=0}^{n-1}\varphi\circ T_{\alpha_0}^k$ *is tight.*

Thus, combining Theorem 4.6 and Lemma 4.7 we have the following.

Corollary 4.8. *If* σ_n *does not tend to infinity as* $n\to\infty$, *then* φ *must be a coboundary.*

Proof of Lemma 4.7. Fix $\epsilon>0$. By the Chebychev's inequality, the assumption that $\sup\left\{\sigma_{n_j}:\ j=1,2,...\right\}<\infty$, implies that there exists a constant A such that for every $j\in\mathbb{N}$,

$$\mu_{n_j}\left(\left|\sum_{k=1}^{n_j}\xi_k\left(X_k\right)-e_{n_j}\right|>A\right)<\epsilon. \tag{4.6}$$

Let $n\in N$ and fix j such that $n<\epsilon h_J^{(n_j)}$ for any $J\in\{L,M,S\}$ (this is possible since the heights of the towers $h_J^{(n)}$ tend to infinity with n). Let x be any point on level l of the tower $Z_J^{(n_j)}$ and consider the Birkhoff sums $\varphi_n(x)$. Then there exists a point $x_0=x_0(x)$ in the base of the tower $I_J^{(n_j)}$ such that $\varphi_n\left(x\right)=\varphi_{n+l}\left(x_0\right)-\varphi_l\left(x_0\right)$. Since the values of $S_l(x_0)$ for $x_0\in I_J^{(n_j)}$ and $0\le l\le h_J^{(n_j)}$ do not depend on x_0, we can choose any point $x_J\in I_J^{(n_j)}$ and by the triangle inequality we have that $\left|\varphi_n\left(x\right)\right|>2A$ implies that $\left|\varphi_{n+l}\left(x_J\right)-e_{n_j}\right|>A$ or $\left|\varphi_l\left(x_J\right)-e_{n_j}\right|>A$ for any point x on level l of the tower $Z_J^{(n_j)}$ with $0\le l<h_J^{(n_j)}-n$. Thus,

$$\begin{aligned}\lambda\Bigg(|\varphi_n(x)|>2A\Bigg|\ x\in Z_J^{(n_j)}\Bigg)&\\ &\le\lambda\left(I_J^{(n_j)}\right)\left(\#\left\{0\le l<h_J^{(n_j)}-n:\ \left|\varphi_{n+l}\left(x_J\right)-e_{n_j}\right|>A\text{ or }\left|\varphi_l\left(x_J\right)-e_{n_j}\right|>A\ \right\}+n\right)\\ &\le\frac{1}{h_J^{(n_j)}}\#\left\{0\le l<h_J^{(n_j)}-n:\ \left|\varphi_{n+l}\left(x_J\right)-e_{n_j}\right|>A\text{ or }\left|\varphi_l\left(x_J\right)-e_{n_j}\right|>A\right\}+\epsilon.\end{aligned}$$

where the last inequality follows by using that $\lambda\left(I_J^{(n_j)}\right)h_J^{(n_j)}=\lambda\left(Z_J^{(n_j)}\right)\le1$ and recalling that by choice of n_j we have that $n/h_J^{(n_j)}<\epsilon$. Furthermore, by a change of indices,

$$\begin{aligned}\frac{1}{h_J^{(n_j)}}\#\Bigg\{0\le l<h_J^{(n_j)}-n:\ \left|\varphi_l\left(x_J\right)-e_{n_j}\right|>A\text{ or }\left|\varphi_{n+l}\left(x_J\right)-e_{n_j}\right|>A\Bigg\}&\\ &\le\frac{2}{h_J^{(n_j)}}\#\left\{0\le l<h_J^{(n_j)}:\ \left|\varphi_l\left(x_J\right)-e_{n_j}\right|>A\ \right\}\\ &=2\mu_{n_j}^J\left(\left|\sum_{k=1}^{n_j}\xi_k\left(X_k\right)-e_{n_j}\right|>A\right),\end{aligned}$$

where the last equality follows from Proposition 2.16. Therefore, from the relation between the measures μ_n^J and μ_n (see Definition 2.12) it follows that

$$\begin{aligned}\lambda\left(|\varphi_n|>2A\right) &= \sum_{J\in\{L,M,S\}}\lambda\left(|\varphi_n|>2A\,\middle|Z_J^{(n_j)}\right)\cdot\lambda\left(Z_J^{(n_j)}\right)\\ &\leq 3\epsilon+\sum_{J\in\{L,M,S\}}2\mu_{n_j}^J\left(\left|\sum_{k=1}^{n_j}\xi_k\left(X_k\right)-e_{n_j}\right|>A\right)\lambda\left(Z_J^{(n_j)}\right)\\ &= 3\epsilon+2\mu_{n_j}\left(\left|\sum_{k=1}^{n_j}\xi_k\left(X_k\right)-e_{n_j}\right|>A\right).\end{aligned}$$

It follows from (4.6) that $\lambda\left(\{x:|\varphi_n(x)|>2A\}\right)<5\epsilon$. Since ϵ was chosen arbitrarily, this shows that φ_n is tight. □

4.3. **Proof of Theorem 1.1.** We begin this section with a few observations that summarize the results obtained in the preceding sections in the form that is used in order to prove Theorem 4.9 below from which the main theorem follows.

Let $0<\alpha<\frac{1}{2}$ be badly approximable and $\beta\in(0,1)$ be badly approximable with respect to α. By Lemma 4.3 the pair (α_0,β_0) related to (α,β) via equations (2.4) and (2.5), satisfies $(\alpha_0,\beta_0)\in\tilde{X}$. To each such pair, in Section 2.5 we associated a Markov compactum given by a sequence of transition matrices $\{A_n\}$ (which are incidence matrices for the substitutions $\{\tau_n\}$ which describe the Rokhlin tower structure) and Markov measures $\{\mu_n\}$ with transition matrices $\left\{p^{(n)}\right\}$ (defined in 2.27 and Definition 2.12 respectively). Let$\{X_k\}$ be the coordinate functions on the Markov compactum (see 2.31) and $\{\xi_k\}$ be the functions also defined therein (see Definition 2.15), which can be used to study the behavior of Birkhoff sums of the function φ defined by (2.6) over T_{α_0} in virtue of as proved in Proposition 2.16. We set

$$e_n:=E_{\mu_n}\left(\sum_{k=1}^{n}\xi_k\left(X_k\right)\right),\qquad \sigma_n:=\sqrt{Var_{\mu_n}\left(\sum_{k=1}^{n}\xi_k\left(X_k\right)\right)},$$

where the subscript μ_n in E_{μ_n} and Var_{μ_n} mean that all integrals are taken with respect to the measure μ_n.

Since the function φ defined by (2.6) is not a coboundary (see Remark 2.4), Corollary 4.8 implies that $\sigma_n\to\infty$. By definition of ξ_k, combining the assumption that α is badly approximable with the inequality (2.25), we obtain that

$$\sup\left\{\xi^k\left(J,j\right):\quad k\in\mathbb{N},\ (J,j)\in\mathcal{S}_k\right\}<\infty.$$

Finally, for any $n\in\mathbb{N}$, set $\xi_k^{(n)}:=\xi_k$, $X_k^{(n)}:=X_n$, for $k=n,...,1$. Let us then define a Markov array $\left\{X_k^{(n)}:\ n\in\mathbb{N},\ k=n,...,1\right\}$, where $Prob\left(\left(X_1^{(n)},...,X_n^{(n)}\right)\in A\right)=\mu_n(A)$ for every set A in the Borel σ-algebra of the space Σ_n. The observations above together with Corollary 4.4 show that all assumptions of Corollary 3.3 hold for this array. Thus

$$\lim_{n\to\infty}\mu_n\left\{\frac{\sum_{k=1}^{N}\xi_k\left(X_k\right)-e_N}{\sigma_N}\in[a,b]\right\}=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx. \tag{4.7}$$

Moreover, by Proposition 3.4 (and the fact the $\sigma_n\to\infty$), (4.7) holds with μ_n replaced by μ_n^J, for any $J\in\{L,M,S\}$ (where μ_n^J are the conditional measures defined by (2.28)).

We can now deduce the temporal CLT for Birkhoff sums. Fix $x\in[-1,\alpha_0)$. Let us first define the centralizing and normalizing constants for the Birkhoff sums $\varphi_n(x)$. For $n\in\mathbb{N}$, let $N=N(n):=\min\left\{k:\ n\leq h_S^{(k)}\right\}$. Let $Z_J^{(N)}$ be the tower at stage N of the renormalization which contains the point x and let l_n be the level of the tower $Z_J^{(N)}$ which contains x, i.e. l_n satisfies $x\in T^{l_n}\left(I_J^{(N)}\right)$. Set $c_n(x):=\varphi_{l_n}(x')$ where x' is any point in $I_J^{(N)}$, i.e. $c_n(x)$ is the Birkhoff sum over the tower $Z_J^{(N)}$ from the bottom of the tower and up to the level that contains x.

We will prove the following temporal DLT, from which Theorem 1.1 follows immediately recalling the correspondence between R_α and T_{α_0} and the functions f_β and φ (refer to the beginning of Section (2.2)).

Theorem 4.9. *For any $a<b$,*

$$\lim_{n\to\infty}\frac{1}{n}\#\left\{0\le k\le n-1:\ \frac{\varphi_k(x)-c_n(x)-e_{N(n)}}{\sigma_{N(n)}}\in[a,b]\right\}=\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx.$$

The above formulation, in particular, shows that the centralizing constants depend on the point x and have a very clear dynamical meaning. The proof of this Theorem, which will take the rest of the section, is based on a quite standard decomposition of a Birkhoff sums into special Birkhoff sums. For each intermediate Birkhoff sum along a tower, we then exploit the connection with the Markov chain given by Proposition 2.16 and the convergence given by (4.7).

Proof. Fix $0<\epsilon<1$, $a,b\in\mathbb{R}$, $a<b$ and let $n\in\mathbb{N}$. By definition of $N=N(n)$, the points $\{x,...,T^{n-1}x\}$ are contained in at most two towers obtained at the N^{th} level of the renormalization. Let K be defined by $K:=K(n)=\max\left\{k:\ h_L^{(k)}\le\epsilon n\right\}$. Evidently, $K\le N$, and by (2.23) there exists $C>0$ which depends on ϵ but not on n, such that $N-K\le C$.

Thus, since towers of level N are decomposed into towers of level K, we can decompose the orbit $\{x,...,T^{n-1}x\}$ into blocks which are each contained in a tower of level K. More precisely, as shown in Figure (4.1), there exist $0=k_0\le k_1<...,<k_t\le n$ and towers $\left(Z_{J_{k_i}}^{(K)}\right)_{i=0}^{t}$ appearing at the K^{th} stage of renormalization, such that $\{T^{k_i}x,...,T^{k_{i+1}-1}x\}\subseteq Z_{J_{k_i}}^{(K)}$ for $i=0,...,t$. Moreover, for $i=1,...,t-1$, the set $\{T^{k_i}x,...,T^{k_{i+1}-1}x\}$ contains exactly $h_{J_{k_i}}^{(K)}$ points, i.e. $k_{i+1}-k_i=h_{J_{k_i}}^{(K)}$ and the points T^{k_i+j}, $j=0,...,k_{i+1}-1$ belong to the $j+1$ level of the tower $Z_{J_{k_i}}^{(K)}$. Since the orbit segment is contained in at most two towers of level N and each tower of level N contains at most $h_L^{(N)}/h_S^{(K)}$ towers of level K, we have that $t=t(n)\le 2h_L^{(N)}/h_S^{(K)}$ and hence is uniformly bounded in n.

It follows from this decomposition that, for any interval $I\subset\mathbb{R}$,

$$\begin{aligned}\frac{1}{n}\#\{0\le k\le n-1:\ \varphi_k(x)\in I\}&=\frac{1}{n}\sum_{i=0}^{t-1}\#\{k_i\le k<k_{i+1}:\ \varphi_k(x)\in I\}\\&\le\frac{1}{n}\sum_{i=1}^{t-1}\#\{k_i\le k<k_{i+1}:\ \varphi_k(x)\in I\}+2\epsilon,\end{aligned}\tag{4.8}$$

where the last inequality follows from the fact that $h_{J_{k_0}}^{(K)}$ and $h_{J_{k_t}}^{(K)}$ are both not greater than $n\epsilon$. Evidently, we also have the opposite inequality

$$\frac{1}{n}\#\{0\le k\le n-1:\ \varphi_k(x)\in I\}\ge\frac{1}{n}\sum_{i=1}^{t-1}\#\{k_i\le k<k_{i+1}:\ \varphi_k(x)\in I\}.\tag{4.9}$$

For $i=1,...,t-1$, and $k_i<k\le k_{i+1}$, write

$$\varphi_k(x)=\varphi_k(x)-\varphi_{k_i}(x)+\varphi_{k_i}(x)=\varphi_{k-k_i}(x')+\varphi_{k_i}(x)$$

where x' is any point in $I_{J_{k_i}}^{(K)}$.

By definition of $c_n(x)$, $\varphi_{k_i}(x)+c_n(x)=\varphi_{k_i}(x_0)$ where x_0 belongs to the base $I^{(N)}$ (see Figure 4.1), thus $\varphi_{k_i}(x)+c_n(x)$ is a sum of special Birkhoff sums over subtowers of $Z_J^{(K)}$, $J\in\{L,M,S\}$. Hence,

$$|\varphi_{k_i}(x)+c_n(x)|\le\left(h_L^{(N)}/h_S^{(K)}\right)\sup_J\left|\varphi_J^{(K)}\right|,$$

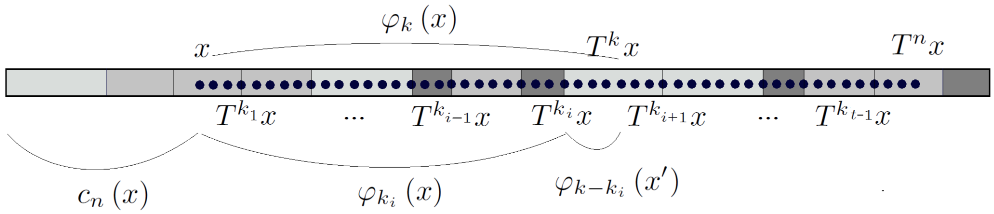


FIGURE 4.1. Decomposition of the orbit segment $\{x, ..., T^{n-1}x\}$ into Birkhoff sums along towers of level $N-K$.

by (2.25), there exists a constant $\tilde{C} := \tilde{C}(\epsilon)$ which does not depend on n, such that $|\varphi_{k_i}(x) + c_n(x)| \leq \tilde{C}$. It follows from Proposition 2.16 that

$$\frac{\#\left\{k_i \leq k < k_{i+1} :\ \frac{\varphi_k(x)-e_N-c_n(x)}{\sigma_N} \in [a,b]\right\}}{h^{(K)}_{J_{k_i}}} = \frac{\#\left\{0 \leq k < h^{(K)}_{J_{k_i}} :\ \frac{\varphi_k(x')-\varphi_{k_i}(x)-e_N-c_n(x)}{\sigma_N} \in [a,b]\right\}}{h^{(K)}_{J_{k_i}}}$$
$$= \mu_K^{J_{k_i}}\left(\frac{\sum_{k=1}^{K}\xi_k(X_k) - \varphi_{k_i}(x) - e_N - c_n(x)}{\sigma_N} \in [a,b]\right).$$

Since $|N-K| = |N(n) - K(n)| < C$, we have that $\sup_n\{|e_N - e_K|\} < \infty$ and $\frac{\sigma_N}{\sigma_K} \underset{n\to\infty}{\longrightarrow} 1$. Moreover, since $|\varphi_{k_i}(x) + c_n(x)| \leq \tilde{C}$, it follows from (4.7), that for any $J \in \{L, M, S\}$

$$\lim_{n\to\infty}\frac{1}{h^{(K)}_J}\#\left\{k_i \leq k < k_{i+1} :\ \frac{\varphi_k(x) - e_N - c_n(x)}{\sigma_N} \in [a,b]\right\} = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx.$$

Let n_0 be such that for all $n > n_0$ and any $J \in \{L, M, S\}$,

$$\left|\frac{1}{h^{(K)}_J}\#\left\{0 \leq k < h^{(k)}_J - 1 :\ \frac{\varphi_k(x) - e_N - c_n(x)}{\sigma_N} \in [a,b]\right\} - \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx\right| < \epsilon. \tag{4.10}$$

Then if $n > n_0$, by (4.8) and (4.10), recalling that $\sum_{i=1}^{t-1} h_{J_{k_i}} \leq n$,

$$\left|\frac{1}{n}\#\Big\{0 \leq k \leq n-1 :\ \frac{\varphi_k(x) - e_N - c_n(x)}{\sigma_N} \in [a,b]\Big|\Big\}$$
$$\leq \frac{1}{n}\sum_{i=1}^{t-1}\#\left\{k_i \leq k < k_{i+1} :\ \frac{\varphi_k(x) - e_N - c_n(x)}{\sigma_N} \in [a,b]\right\} + 2\epsilon$$
$$\leq \frac{1}{n}\sum_{i=1}^{t-1} h_{J_{k_i}}\left(\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\mathrm{dx}\cdot\frac{1}{n}\sum_{i=1}^{t-1} h_{J_{k_i}} + \epsilon\right) + 2\epsilon$$
$$\leq \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx + 3\epsilon.$$

Similarly, by (4.9), if $n > n_0$, using this time that $\sum_{i=1}^{t-1} h_{J_{k_i}} \geq n(1-2\epsilon)$, we obtain the lower bound

$$\begin{aligned}
\Big|\frac{1}{n}\#\Big\{0 \leq k \leq n-1 :\ & \frac{\varphi_k(x) - e_N - c_n(x)}{\sigma_N} \in [a,b]\Big|\Big\} \\
& \geq \frac{1}{n}\sum_{i=1}^{t-1} \#\left\{k_i \leq k < k_{i+1} :\ \frac{\varphi_k(x) - e_N - c_n(x)}{\sigma_N} \in [a,b]\right\} \\
& \geq \frac{1}{n}\sum_{i=1}^{t-1} h_{J_{k_i}} \left(\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx - \epsilon\right) \\
& \geq (1-2\epsilon)\left(\frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}dx - \epsilon\right).
\end{aligned}$$

This completes the proof. □

Acknowledgments. We would like to thank Jon Aaronson, Dima Dolgopyat, Jens Marklof and Omri Sarig for useful discussions and for their interest in our work. We would also like to thank the referee for carefully reading the paper and for valuable comments. Both authors are supported by the ERC Starting Grant ChaParDyn. C. U. is also supported by the Leverhulm Trust through a Leverhulme Prize and by the Royal Society through a Wolfson Research Merit Award. The research leading to these results has received funding from the European Research Council under the European Union Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement n. 335989.