MÖBIUS DISJOINTNESS FOR NON-UNIQUELY ERGODIC SKEW PRODUCTS

ZHIREN WANG

Abstract. For $\tau > 2$, let T be a C^τ skew product map of the form $(x+\alpha, y+h(x))$ on $\mathbb{T}^2$ over a rotation of the circle. We show that if T preserves a measurable section, then it is disjoint to the Möbius sequence. This in particular implies that any non-uniquely ergodic C^τ skew product map on $\mathbb{T}^2$ has a finite index factor that is disjoint to the Möbius sequence.

Contents

1. Introduction

Let T be a skew product map on $\mathbb{T}^2$ over a rotation of the circle $\mathbb{T}^1$. That is,

$$T(x,y) = (x+\alpha, y+h(x)), \tag{1.1}$$

where $h : \mathbb{T}^1 \mapsto \mathbb{T}^1$ is continuous, and $\alpha \in [0,1)$.

A measurable invariant section of T is a graph $(x, g(x))$ where $g : \mathbb{T}^1 \mapsto \mathbb{T}^1$ is measurable, such that $T(x, g(x))$ is still in the graph for Lebesgue almost every x.

Our main result is the following Mobius disjointness property:

Theorem 1.1. *Suppose h is C^τ for some real value $\tau > 2$ and T preserves a measurable invariant section. Then for all $(x,y) \in \mathbb{T}^2$, and all continuous functions $f \in C(\mathbb{T}^2)$,*

$$\frac{1}{N}\sum_{n\le N} \mu(n) f(T^n(x,y)) \to 0 \text{ as } N \to \infty. \tag{1.2}$$

One important feature of the theorem is that it holds for all α, without assuming any Diophantine condition.

By a dichotomy of Furstenberg [Fur61, Lemma 2.1], if a map T of the form (1.1) is not uniquely ergodic, then for some positive integer ξ, the solution $g(x+\alpha)-g(x)=\xi h(x)$ has a measurable solution $g:\mathbb{T}^1\mapsto\mathbb{T}^1$. Let $\pi_\xi:\mathbb{T}^2\mapsto\mathbb{T}^2$ be the ξ-to-one projection $\pi(x,y)=\pi(x,\xi y)$. Then the transform $T_\xi(x,y)=(x+\alpha,y+\xi h(x))$ is a topological factor of T through π, in other words, $\pi_\xi\circ T=T_\xi\circ\pi_\xi$. One can easily check that the graph $(x,g(x))$ is a measurable invariant section for T_ξ. Hence Theorem 1.1 implies:

Corollary 1.2. *Suppose h is C^τ for some real value $\tau>2$ and T is not uniquely ergodic. Then there exists $\xi\in\mathbb{N}$, such that the ξ-to-one topological factor T_ξ of T via the projection π_ξ satisfies*

$$\frac{1}{N}\sum_{n\le N}\mu(n)f(T_\xi^n(x,y))\to 0 \text{ as } N\to\infty.$$

The Möbius function is the multiplicative function

$$\mu(n)=\begin{cases}(-1)^{\#\text{ of prime factors of }n}, & \text{if } n \text{ is square free}\\ 0, & \text{otherwise.}\end{cases}$$

It is expected that the Möbius function captures much of the randomness in the distribution of prime numbers. This is charactertized by Sarnak's Mobius Disjointness Conjecture, which has been the focus of much research in number theory and dynamical systems in recent years.

Conjecture 1.3. *(Möbius Disjointness Conjecture,* [Sar09]*) For a continuous transformation $T:X\mapsto X$ with zero topological entropy, where X is a compact metric space and T is continuous, then for all continuous $f\in C(X)$ and all $x\in X$, $\frac{1}{N}\sum_{n\le N}\mu(n)f(T^n(x))\to 0$.*

It should be emphasized the conjecture addresses all points $x\in X$, instead of almost every x with respect to a measure.

When X is a single point, Conjecture 1.3 gives the prime number theorem. The prime number theorem in arithmetic progressions corresponds to the case of a rotation on the finite abelian group $\mathbb{Z}/q\mathbb{Z}$. Davenport's classic theorem [Dav37] that $\frac{1}{N}\sum_{n\le N}\mu(n)e(\alpha n)\to 0$ is the case of a rotation of the circle.

Researches during the last a few years confirmed numerous cases of Conjecture 1.3. To list a few: [MR10, Gre12, GT12, Bou13a, Bou13b, BSZ13, KPL15, EALdlR14, MMR14, Pec15, FKPLM15]. Most of these results use Vinogradov's bilinear method, or its modern variant due to Vaughan [Vau77]. The Bourgain-Sarnak-Ziegler criterion, developed in [BSZ13], offers a different variant that allows to interprete bilinear averages in as ergodic averages in joinings of dynamical systems.

All the special case listed above are regular dynamical systems. A system (X,T) is regular if the ergodic average $\frac{1}{N}\sum_{n\le N}f(T^n(x))$ converges for every

$x \in X$ and every continuous function f. In particular, all uniquely ergodic systems are regular by Birkhoff's ergodic theorem.

The only irregular dynamical systems for which Möbius disjointness has been studied come from the family (1.1). In this family, for generic choices of α that satisfy certain diophantine conditions, a $C^{1+\epsilon}$-differentiable T is still regular, as well as the joinings needed for applying the Bourgain-Sarnak-Ziegler criterion. For such α, Conjecture 1.3 was proved by Kułaga-Przymus and Lemańczyk [KPL15].

On the other hand, for Liouville choices of α, the map T can be irregular. Such a counterexample was first constructed by Furstenberg in [Fur61]. Furstenberg's example is real analytic, but it is not hard to modify it to get a counterexample of finite differentiability.

Irregular transformation from the family 1.1 were first studied in [LS15] by Liu and Sarnak. They proved Conjecture 1.3 for a class of analytic skew products of the form (1.1) under an additional assumption that $\hat{h}(m)$ decays not too fast. In [Wan15], the author proved Conjecture 1.3 for all real analytic maps of the form (1.1). One main ingredient from [Wan15] was the use of the estimate of averages of non-pretentious multiplicative functions on typical short intervals by Matomäki, Radziwiłł and Tao [MRT15]. The use in [Wan15] of this new tool was quantitative, and as a consequence analyticity, or at least the weaker condition that $|\hat{h}(m)| \ll e^{-\tau|m|^{\frac{1}{2}+\epsilon}}$, had to be assumed to control the estimates from [MRT15].

The main aim of Theorem 1.1 is to prove Conjecture 1.3 for some irregular dynamical systems of finite differentiability.

Our proof of Theorem 1.1, except for a few standard reductions at the beginning, doesn't use bilinear method. Instead, as in [Wan15], the endgame of the proof relies on the theorem of Matomäki-Radziwiłł-Tao. To be accurate, we use an application that Matomäki-Radziwiłł-Tao derived from their theorem, saying that when R is sufficiently large and N is sufficiently large compared to R, for most $1 \leq L \leq N$, the correlation between $\mu(n)$ and $e(\alpha n)$ on the short interval $[L, L+R]$ is small (see Proposition 9.3). However, unlike in [Wan15], the use of [MRT15] in this paper is only qualitative, eliminating the need of analyticity.

The dynamical analysis in this paper is quite different from that in [Wan15]. The main strategy in [Wan15] was to prove that the dynamics is almost periodic at a single step length for a very long time. In contrast, to prove Theorem 1.2, we analyse the trajectory at multiple scales and show that it has a structure that looks like a sum of independent random variables at different scales. These scales are determined by the continued fraction expansion of α. Thanks to the mutual independence, in order to have a measurable invariant section, the total variance of these random variables have to be bounded. This allows to study the dynamics only on the first finitely many scales. After such a reduction, the dynamics resembles a linear flow when n is restricted to a short interval of given length. At this stage, [MRT15] can be applied.

Notations.

- $\mathrm{m}_{\mathbb{T}^d}$ = the Lebesgue probability measure on $\mathbb{T}^d$;
- $e(\theta) = e^{2\pi i\theta}$;
- $\|\theta\| = \mathrm{dist}(\theta, \mathbb{Z})$ for $\theta \in \mathbb{R}$. Remark that $\|\theta\| \ll e(\theta) - 1 \ll \|\theta\|$;
- $\mathbb{E}_F$, Var_F, $\mathbb{P}_F$ respectively stand for the expectation, variance and probability of a function/event defined on a finite set F with respect to the uniform probability measure on F.

Acknowledgments. I am grateful to Jean Bourgain, Jon Chaika and Peter Sarnak for helpful discussions. This work was done when the author was a member at the Institute for Advanced Study, and I would like to thank the IAS for its support. The author was also supported by NSF grants DMS-1451247 and DMS-1501295.

2. Elimination of non-resonant frequencies

In the remainder of the paper, we will assume that T is given by (1.1) where $h : \mathbb{T}^1 \to \mathbb{T}^1$ is C^τ regular with $\tau \in (2, \infty)$.

Liu and Sarnak [LS15] showed that if α is rational, then Conjecture 1.3 holds for T. So we will always assume α is irrational.

Since T has a measurable invariant section, by a result of Furstenberg [Fur61, Lemma 2.2] h is homotopically trivial. Under this restriction, h can be realized as a C^τ function from $\mathbb{T}^1$ to $\mathbb{R}$ and be written as

$$h(x) = \sum_{m\in\mathbb{Z}} \hat{h}(m)e(mx), \tag{2.1}$$

where the convergence is uniform and the equality holds pointwise for all $x \in \mathbb{T}^1$. Moreover, as h is C^τ, we have

$$|\hat{h}(m)| \ll |m|^{-\tau}, \forall m \in \mathbb{Z}\backslash\{0\}. \tag{2.2}$$

Take the continued fraction expansion

$$\alpha = [0; a_1, a_2, \cdots] = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cfrac{1}{a_3 + \cdots}}}$$

and let $\frac{p_k}{q_k} = [0; a_1, \cdots, a_{k-1}]$ be the corresponding convergents[1]. Because α is irrational, the expansion is infinite. Denote $\theta_k = q_k\alpha - p_k$.

Remark 2.1. *The following standard facts can be found in* [Khi97]*:*

(1) $p_1 = 0$, $q_1 = 1$; $p_2 = 1$, $q_2 = a_1$; $p_{k+1} = a_k p_k + p_{k-1}$ *and* $q_{k+1} = a_k q_k + q_{k-1}$ *for* $k \geq 2$;

(2) p_k *is coprime to* q_k;

[1] It should be noted that our enumeration here differs from the more commonly used one in the literature, namely $\frac{p_k}{q_k} = [0; a_1, \cdots, a_k]$.

(3) $\frac{1}{q_{k+1}+q_k} < \|q_k\alpha\| < \frac{1}{q_{k+1}}$*;*
(4) If $|\alpha - \frac{p}{q}| < \frac{1}{2q^2}$ *for* $p, q \in \mathbb{Z}$*, then* $\frac{p}{q}$ *coincides with one of the* $\frac{p_k}{q_k}$*'s.*
(5) The sequence $\{\theta_k\}$ *has alternating signs and* $\|q_k\alpha\| = |\theta_k|$*.*

One consequence to Remark 2.1.(1) is that $q_{k+2} \geq q_{k+1} + q_k \geq 2q_k$. In particular, q_k grows exponentially:

$$q_k \gg 2^{\frac{k}{2}} \text{ and } q_{k+j} \gg 2^{\frac{j}{2}} q_k. \tag{2.3}$$

Set

$$M = \bigcup_{k: q_{k+1} > q_k^{\frac{\tau}{2}}} \{\pm m_k q_k : m_k = 1, \cdots, a_k\} \tag{2.4}$$

What we prove next is essentially [LS15, Lemma 4.1], with slightly finer estimates.

Lemma 2.2. *If* (2.2) *holds, then*

$$\sum_{m \notin M} \hat{h}(m) \frac{1}{e(m\alpha) - 1} e(mx)$$

converges uniformly to a continuous function $\psi(x)$*.*

Proof. If $m \notin M$, then we are in at least one of the three situations below:

(1) $m = 0$. Since this involves only one frequecy, we can ignore this case in the study of convergence.

(2) For some k, $q_k \leq |m| < q_{k+1}$ but $q_k \nmid |m|$. Then $\|m\alpha\| \geq \frac{1}{2|m|}$. This is because otherwise, by Remark 2.1, $|m| = aq_j$ and $\|m\alpha\| = |m\alpha - ap_j|$ for some index $j \leq k$ and $a \in \mathbb{Z}$. Since $q_k \nmid |m|$, $j < k$. Hence we have

$$\|m\alpha\| = |a| \cdot \|q_j\alpha\| > \frac{|a|}{q_{j+1} + q_j} \geq \frac{1}{2q_k} \geq \frac{1}{2m}$$

anyway in this case.

Therefore for any given k

$$\begin{aligned}
&\sum_{\substack{q_k \leq |m| < q_{k+1} \\ q_k \nmid m}} \left| \hat{h}(m) \frac{1}{e(m\alpha) - 1} e(mx) \right| \\
\ll &\sum_{\substack{q_k \leq |m| < q_{k+1} \\ q_k \nmid m}} (|m|^{-\tau} \cdot |m| \cdot 1) \ll \sum_{m=q_k}^{q_{k+1}} m^{-(\tau-1)} \\
\ll &q_k^{-(\tau-2)} - q_{k+1}^{-(\tau-2)}.
\end{aligned} \tag{2.5}$$

(3) $m = \pm m_k q_k$ where $m_k \in \{1, \cdots, a_k\}$ but $q_{k+1} \leq q_k^{\frac{\tau}{2}}$. Since $m_k\|q_k\alpha\| \leq a_k \frac{1}{q_{k+1}} < \frac{1}{q_k}$, $\|m\alpha\|$ is given by $m_k\|q_k\alpha\|$ for $k \geq 2$. Thus, we have for all $k \geq 2$ that

$$\begin{aligned}
&\sum_{\substack{q_k\le|m|<q_{k+1}\\ q_k|m}}\left|\hat{h}(m)\frac{1}{e(m\alpha)-1}e(mx)\right|\\
\ll&2\sum_{m_k=1}^{a_k}\left((m_kq_k)^{-\tau}\cdot\frac{1}{m_k\cdot\frac{1}{q_{k+1}+q_k}}\cdot 1\right)\\
\ll&\sum_{m_k=1}^{a_k}m_k^{-(\tau+1)}q_{k+1}q_k^{-\tau}\ll\sum_{m_k=1}^{\infty}m_k^{-(\tau+1)}q_k^{-\frac{\tau}{2}}\\
\ll&q_k^{-\frac{\tau}{2}}.
\end{aligned}\tag{2.6}$$

The last inequality here is because $\tau-1>1$.

We sum both estimates (2.5) and (2.6) over all $k\ge 2$. Since only finitely many terms are neglected in doing this, and the estimates are independent of x, to prove the lemma it suffices to know that both the resulting series are convergent. This is indeed the case, respectively because $\tau-2>0$ and $\frac{\tau}{2}>1$. □

Corollary 2.3. *Assuming* (2.2), *Conjecture 1.3 holds for T if and only it holds for the map $T_1(x,y)=(x+\alpha,y+h_1(x))$ on $\mathbb{T}^2$, where*

$$h_1(x)=\sum_{m\in M\cup\{0\}}\hat{h}(m)e(mx).$$

In other words, in order to prove Theorem 1.1, one may assume that $\hat{h}$ is supported on $M\cup\{0\}$. The proof of the corollary is the same as that of [Wan15, Corollary 3.3]. In fact, it suffices to notice that T is continously conjugate to T_1 by the map $(x,y)\mapsto(x,y+\psi(x))$.

In the same spirit, one may assume that M is infinite. In fact, if M is finite then

$$\sum_{m\neq 0}\hat{h}(m)\frac{1}{e(m\alpha)-1}e(mx)$$

differs from the series in Lemma 2.2 by only finitely many terms and hence also defines a continuous function $\tilde{\psi}$. And T is continously conjugate to the Kronecker flow $(x,y)\mapsto(x+\alpha,y+\hat{h}(0))$ by $(x,y)\mapsto(x,y+\tilde{\psi}(x))$. So it suffices to show (1.2) for this linear flow, which is known by the work of Davenport [Dav37].

To summarize the reductions above, in order to prove Theorem 1.1, we may assume the following:

Hypothesis 2.4. *$T=(x+\alpha,y+h(x))$ with α irrational and $h\in C^\tau$ where $\tau>2$. Moreover,*

1. *h is homotopically trivial;*
2. *The subset $M\subset\mathbb{Z}$ defined by* (2.4) *is infinite and $\hat{h}$ is supported on M;*

(3) The map T preserves a measurable section $(x, g(x))$.

We will always work under Hypothesis 2.4 hereafter.

3. Approximation of trajectories: non-zero frequencies

For $n \in \mathbb{Z}$, write the n-th iterate of T as

$$T^n(x,y) = \big(x + n\alpha, y + H_n(x)\big). \tag{3.1}$$

Then for $n \geq 0$,

$$H_n(x) = \sum_{l=0}^{n-1} h(x + l\alpha) = \sum_{l=0}^{n-1} \sum_{m\in\mathbb{Z}} \hat{h}(m)e(mx)e(lm\alpha).$$

By exchanging the order of sums, we get

$$\begin{aligned} H_n(x) =& \sum_{m\in\mathbb{Z}} \hat{h}(m)\big(\sum_{l=0}^{n-1} e(lm\alpha)\big)e(mx) \\ =& n\hat{h}(0) + \sum_{m\in\mathbb{Z}\setminus\{0\}} \hat{h}(m)\frac{e(nm\alpha)-1}{e(m\alpha)-1}e(mx). \end{aligned} \tag{3.2}$$

When $n < 0$,

$$H_n(x) = \sum_{l=n}^{-1} -h(x + l\alpha) = \sum_{l=n}^{-1} \sum_{m\in\mathbb{Z}} -\hat{h}(m)e(mx)e(lm\alpha).$$

In this case, one can easily check that (3.2) remains true. Thus (3.2) holds for all $n \in \mathbb{Z}$.

Given $x \in \mathbb{T}^1$, we shall choose its unique representative from $[-\alpha, 1-\alpha)$, which we denote by x indifferently. To better approximate the series (3.2) we expand n and x in their Ostrowski numerations [Ost22] with respect to α, the definitions of which we now recall.

Definition 3.1. (Ostrowski numeration for integers) *Every non-negative integer n can be uniquely written as a sum $\sum_{k\geq 1} n_k q_k$, where:*

(1) n_k is an integer, $0 \leq n_k \leq a_k$. When $k = 1$, $0 \leq n_1 \leq a_1 - 1$;
(2) $n_k = 0$ if $n_{k+1} = a_{k+1}$;
(3) $n_k = 0$ for all but finitely many k's.

Definition 3.2. (Ostrowski numeration for real values) *Every value $x \in [-\alpha, 1-\alpha)$ can be uniquely written as a convergent series $\sum_{k\geq 1} \tilde{x}_k \theta_k$, where:*

(1) $\tilde{x}_k$ is an integer, $0 \leq \tilde{x}_k \leq a_k$ in general, but $0 \leq \tilde{x}_1 \leq a_1 - 1$;
(2) $\tilde{x}_k = 0$ if $\tilde{x}_{k+1} = a_{k+1}$;
(3) $\tilde{x}_k < a_k$ infinitely often.

We denote $x_k = \tilde{x}_k \theta_k$.

For an introduction to Ostrowski numerations, see the survey [Ber01].

Given positive integers $2 \leq k_- \leq k_+$, we are interested in estimating (3.2) for positive integers n whose Ostrowski numerations have the form $n = \sum_{k=k_-}^{k_+} n_k q_k$.

More generally, we may assume

$$n = \sum_{k=k_-}^{k_+} n_k q_k,\ n_k \in \mathbb{Z},\ |n_k| \ll a_k, \tag{3.3}$$

without requiring the decomposition to be an Ostrowski numeration.

With (3.3), we will denote partial sums by

$$\bar{n}_k = \sum_{j=k_-}^{k} n_j q_j. \tag{3.4}$$

The main idea of this section is that the interaction between the component $n_k q_k$ of n and the Fourier component at frequency $m_j q_j$ of h really matters for the dynamics of T only when $k = j$.

Lemma 3.3. *For k_-, k_+, n as in* (3.3) *and any $x \in \mathbb{T}^1$, $H_n(x)$ is approximated by the sum*

$$H_n^{(1)}(x) = n\hat{h}(0) + \sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \leq m_k \leq a_k \\ m_k \neq 0}} \hat{h}(m_k q_k) \frac{e(n_k m_k q_k^2 \alpha) - 1}{e(m_k q_k \alpha) - 1} e(m_k q_k (x + \bar{n}_{k-1}\alpha)),$$

with an error term of order $O(2^{-\frac{k_-}{4}})$.

Proof. By definition $H_n(x) = \sum_{k=k_-}^{k_+} H_{n_k q_k}(x + \bar{n}_{k-1}\alpha)$. Since we are assuming Hypothesis 2.4, it follows that

$$\begin{aligned} &H_n(x) \\ =&n\hat{h}(0) + \sum_{k=k_-}^{k_+} \sum_{j=1}^{\infty} \sum_{\substack{-a_j \leq m_j \leq a_j \\ m_j \neq 0}} \hat{h}(m_j q_j) \frac{e(n_k q_k m_j q_j \alpha) - 1}{e(m_j q_j \alpha) - 1} e(m_j q_j (x + \bar{n}_{k-1}\alpha)). \end{aligned} \tag{3.5}$$

So $H_n(x) - H_n^{(1)}(x)$ is equal to

$$\sum_{k=k_-}^{k_+} \sum_{\substack{j \geq 1 \\ j \neq k}} \sum_{\substack{-a_j \leq m_j \leq a_j \\ m_j \neq 0}} \hat{h}(m_j q_j) \frac{e(n_k q_k m_j q_j \alpha) - 1}{e(m_j q_j \alpha) - 1} e(m_j q_j (x + \bar{n}_{k-1}\alpha)).$$

Therefore

$$
\begin{aligned}
&|H_n(x) - H_n^{(1)}(x)| \\
\le &\sum_{k=k_-}^{k_+} \sum_{\substack{j\ge 1\\ j\ne k}} \left| \sum_{\substack{-a_j\le m_j\le a_j\\ m_j\ne 0}} \hat{h}(m_j q_j) \frac{e(n_k q_k m_j q_j \alpha)-1}{e(m_j q_j \alpha)-1} e(m_j q_j (x+\bar{n}_{k-1}\alpha)) \right| \\
\ll &\sum_{k=k_-}^{k_+} \sum_{\substack{j\ge 1\\ j\ne k}} \sum_{\substack{-a_j\le m_j\le a_j\\ m_j\ne 0}} |\hat{h}(m_j q_j)| \frac{\|n_k q_k m_j q_j \alpha\|}{\|m_j q_j \alpha\|}
\end{aligned} \tag{3.6}
$$

We distinguish between the cases $j<k$ and $j>k$.
When $j<k$,

$$
\begin{aligned}
&\sum_{\substack{-a_j\le m_j\le a_j\\ m_j\ne 0}} |\hat{h}(m_j q_j)| \frac{\|n_k q_k m_j q_j \alpha\|}{\|m_j q_j \alpha\|} \\
\ll &\sum_{\substack{-a_j\le m_j\le a_j\\ m_j\ne 0}} |m_j|^{-\tau} q_j^{-\tau} \frac{|n_k|\cdot|m_j|\cdot q_j\cdot|\theta_k|}{|m_j|\cdot|\theta_j|} \\
\ll &\sum_{\substack{-a_j\le m_j\le a_j\\ m_j\ne 0}} |m_j|^{-\tau} q_j^{-\tau} \frac{|n_k| q_j q_{j+1}}{q_{k+1}} \ll q_j^{-(\tau-1)} \frac{q_{j+1}}{q_k}
\end{aligned} \tag{3.7}
$$

Here we used $q_{j+1}^{-1} \ll |\theta_j| \ll q_{j+1}^{-1}$, which is guaranteed by Remark 2.1, and that $\frac{|n_k|}{q_{k+1}} \ll \frac{a_k}{q_{k+1}} < \frac{1}{q_k}$.

One can further bound the estimate above using the exponential growth rate from (2.3). If $1 \le j < \frac{k}{2}$, then (3.7) $\ll \frac{q_{j+1}}{q_k} \ll 2^{-\frac{k-j-1}{2}}$. If $\frac{k}{2} \le j < k$, then (3.7) $\ll q_j^{-(\tau-1)} \ll 2^{-\frac{j(\tau-1)}{2}}$ by (2.3). It follows that

$$
\begin{aligned}
&\sum_{k=k_-}^{k_+} \sum_{j=1}^{k-1} \sum_{\substack{-a_j\le m_j\le a_j\\ m_j\ne 0}} |\hat{h}(m_j q_j)| \frac{\|n_k q_k m_j q_j \alpha\|}{\|m_j q_j \alpha\|} \\
\ll &\sum_{k=k_-}^{k_+} \left(\sum_{j=1}^{\lceil \frac{k}{2}\rceil -1} 2^{-\frac{k-j-1}{2}} + \sum_{j=\lceil \frac{k}{2}\rceil}^{k-1} 2^{-\frac{j(\tau-1)}{2}} \right) \\
\ll &\sum_{k=k_-}^{k_+} (2^{-\frac{k}{4}} + 2^{-\frac{(\tau-1)k}{4}}) \ll \sum_{k=k_-}^{k_+} 2^{-\frac{k}{4}} \ll 2^{-\frac{k_-}{4}}.
\end{aligned} \tag{3.8}
$$

And when $j > k$,

$$\begin{aligned}
&\sum_{\substack{-a_j\le m_j\le a_j\\ m_j\neq 0}} |\hat{h}(m_jq_j)|\frac{\|n_kq_km_jq_j\alpha\|}{\|m_jq_j\alpha\|}\\
\ll &\sum_{\substack{-a_j\le m_j\le a_j\\ m_j\neq 0}} |m_j|^{-\tau}q_j^{-\tau}n_kq_k \ll q_j^{-\tau}q_{k+1} \qquad (3.9)\\
\ll &q_j^{-\tau}q_j = q_j^{-(\tau-1)}.
\end{aligned}$$

Thus, once again thanks to the exponential growth of $\{q_k\}$,

$$\begin{aligned}
&\sum_{k=k_-}^{k_+}\sum_{j=k+1}^{\infty}\sum_{\substack{-a_j\le m_j\le a_j\\ m_j\neq 0}} |\hat{h}(m_jq_j)|\frac{\|n_kq_km_jq_j\alpha\|}{\|m_jq_j\alpha\|}\\
\ll &\sum_{k=k_-}^{k_+}\sum_{j=k+1}^{\infty} q_j^{-(\tau-1)} \ll \sum_{k=k_-}^{k_+} q_{k+1}^{-(\tau-1)} \qquad (3.10)\\
\ll &q_{k_-+1}^{-(\tau-1)}.
\end{aligned}$$

Because $q_{k_-+1} \gg 2^{\frac{k_-}{2}}$ and $\tau - 1 > 1$, (3.10) is dominated by (3.8). By feeding both of them into (3.6), we obtain that $|H_n(x) - H_n^{(1)}(x)| \ll 2^{-\frac{k_-}{4}}$. $\square$

From (3.9) we deduce an approximation by truncated Fourier series, that will become useful in a later part of this paper.

Lemma 3.4. *Suppose $n = \sum_{k=k_-}^{k_+} n_kq_k$ where $|n_k| \ll a_k$, define a truncaction of $h(x)$ by $h^*(x) = \hat{h}(0) + \sum_{j=1}^{k_+}\sum_{\substack{-a_k\le m_k\le a_k\\ m_k\neq 0}} \hat{h}(m_kq_k)e(m_kq_kx)$. Also define $H_n^*(x)$ in the same way as $H_n(x)$ using $h^*(x)$ instead of $h(x)$, then*

$$\|H_n^*(x) - H_n(x)\| \ll q_{k_++1}^{-(\tau-1)}.$$

Proof. The difference between $H_n(x)$ and $H_n^*(x)$ is the sum of all terms in (3.2) involving frequencies m_jq_j, $j > k_+$, or

$$\sum_{k=k_-}^{k_+}\sum_{j=k_++1}^{\infty}\sum_{\substack{-a_j\le m_j\le a_j\\ m_j\neq 0}} \hat{h}(m_jq_j)\frac{e(n_kq_km_jq_j\alpha)-1}{e(m_jq_j\alpha)-1}e(m_jq_j(x+\bar{n}_{k-1}\alpha)).$$

By the second line in (3.9),

$$\begin{aligned}
\|H_n^*(x) - H_n(x)\| \ll &\sum_{k=k_-}^{k_+}\sum_{j=k_++1}^{\infty} q_j^{-\tau}q_{k+1} \ll \sum_{k=k_-}^{k_+} q_{k_++1}^{-\tau}q_{k+1}\\
\ll &q_{k_++1}^{-\tau}q_{k_++1} = q_{k_++1}^{-(\tau-1)}.
\end{aligned}$$

□

The approximation $H_n^{(1)}(x)$ can be further refined by exploiting the Ostrowski numeration of x.

Lemma 3.5. *For k_-, k_+, and n as in (3.3), $H_n(x)$ is approximated by the sum*

$$H_n^{(2)}(x) = n\hat{h}(0) + \sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \le m_k \le a_k \\ m_k \ne 0}} \hat{h}(m_k q_k) \frac{e((n_k + \tilde{x}_k) m_k q_k^2 \alpha) - e(\tilde{x}_k m_k q_k^2 \alpha)}{e(m_k q_k \alpha) - 1},$$

with an error term of order $O(2^{-\frac{k_-}{4}})$. Here $\{\tilde{x}_k\}$ is the Ostrowski numeration of x.

Proof. The proof is in a sense symmetric to that of Lemma 3.5, by interchanging perspectives of the variables n and x.

Write first

$$\begin{aligned} x + \bar{n}_{k-1}\alpha =& \sum_{j=1}^{\infty} \tilde{x}_j \theta_j + \sum_{j=1}^{k-1} n_j q_j \alpha \\ \equiv& \tilde{x}_k q_k \alpha + \sum_{j=1}^{k-1} (n_j + \tilde{x}_j) q_j \alpha + \sum_{j=k+1}^{\infty} \tilde{x}_j q_j \alpha \pmod 1. \end{aligned} \tag{3.11}$$

For simplicity, write

$$\tilde{x}'_j = \tilde{x}_j + \mathbf{1}_{j \le k} n_j,$$

and

$$z_j = \tilde{x}_k q_k \alpha + \sum_{l=1}^{\min(j,k-1)} \tilde{x}'_l q_l \alpha + \sum_{l=k+1}^{j} \tilde{x}'_l q_l \alpha.$$

Then $z_0 = \tilde{x}_k q_k \alpha$; and $z_j \to x + \bar{n}_{k-1}\alpha$ as $j \to \infty$ in $\mathbb{T}^1$; and $z_j - z_{j-1} = \tilde{x}'_j q_j \alpha$ except when $j = k$, for which $z_k = z_{k-1}$. Also notice that $|\tilde{x}'_j| \ll 2a_j$.

Therefore we have a decomposition

$$\begin{aligned} & e(m_k q_k (x_k + \bar{n}_{k-1}\alpha)) \\ =& e(m_k q_k \tilde{x}_k q_k \alpha) + \sum_{\substack{j \ge 1 \\ j \ne k}} e(m_k q_k z_j) \big(e(m_k q_k \tilde{x}'_j q_j \alpha) - 1 \big). \end{aligned} \tag{3.12}$$

This series converges uniformly as with k fixed, the j-th term is of the same order as $\|\tilde{x}'_j q_j \alpha\| \ll a_j \theta_j \ll a_j q_{j+1}^{-1} \ll q_j^{-1}$ and decays exponentially fast.

After plugging (3.12) into the expression of $H_n^{(1)}(x)$ and comparing with $H_n^{(2)}(x)$, we see that

$$
\begin{aligned}
&|H_n^{(1)}(x) - H_n^{(2)}(x)| \\
\le&\Bigg| \sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \le m_k \le a_k \\ m_k \ne 0}} \Big(\hat{h}(m_k q_k) \frac{e(n_k m_k q_k^2 \alpha) - 1}{e(m_k q_k \alpha) - 1} \cdot \\
&\quad \sum_{\substack{j \ge 1 \\ j \ne k}} e(m_k q_k z_j) \big(e(m_k q_k \tilde{x}'_j q_j \alpha) - 1 \big) \Big) \Bigg| \\
\ll& \sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \le m_k \le a_k \\ m_k \ne 0}} \sum_{\substack{j \ge 1 \\ j \ne k}} |\hat{h}(m_k q_k)| \cdot \left| \frac{e(m_k q_k \tilde{x}'_j q_j \alpha) - 1}{e(m_k q_k \alpha) - 1} \right| \\
\ll& \sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \le m_k \le a_k \\ m_k \ne 0}} \sum_{\substack{j \ge 1 \\ j \ne k}} |\hat{h}(m_k q_k)| \frac{\|m_k q_k \tilde{x}'_j q_j \alpha\|}{\|m_k q_k \alpha\|}.
\end{aligned}
\tag{3.13}
$$

We again distinguish the cases $j < k$ and $j > k$.

Suppose first $j < k$. Similar to (3.9) we have

$$
\begin{aligned}
&\sum_{\substack{-a_k \le m_k \le a_k \\ m_k \ne 0}} |\hat{h}(m_k q_k)| \frac{\|m_k q_k \tilde{x}'_j q_j \alpha\|}{\|m_k q_k \alpha\|} \\
\ll& \sum_{\substack{-a_k \le m_k \le a_k \\ m_k \ne 0}} |m_k|^{-\tau} q_k^{-\tau} \cdot |\tilde{x}'_j| q_j \ll q_k^{-\tau} q_{j+1}.
\end{aligned}
\tag{3.14}
$$

Here we used that $|\tilde{x}'_j| q_j \ll a_j q_j < q_{j+1}$.

Because q_k has exponential growth, it follows that

$$
\begin{aligned}
&\sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \le m_k \le a_k \\ m_k \ne 0}} \sum_{j=1}^{k-1} |\hat{h}(m_k q_k)| \frac{\|m_k q_k \tilde{x}'_j q_j \alpha\|}{\|m_k q_k \alpha\|} \\
\ll& \sum_{k=k_-}^{k_+} q_k^{-\tau} q_k = \sum_{k=k_-}^{k_+} q_k^{-(\tau-1)} \ll q_{k_-}^{-(\tau-1)}.
\end{aligned}
\tag{3.15}
$$

Assume now $j > k$. Because $|\tilde{x}'_j| \ll a_j$, similar to (3.7) we have

$$\begin{aligned}
&\sum_{\substack{-a_k \le m_k \le a_k \\ m_k \neq 0}} |\hat{h}(m_k q_k)| \frac{\|m_k q_k \tilde{x}'_j q_j \alpha\|}{\|m_k q_k \alpha\|} \\
\ll &\sum_{\substack{-a_k \le m_k \le a_k \\ m_k \neq 0}} |m_k|^{-\tau} q_k^{-\tau} \frac{a_j |m_k| q_k \cdot |\theta_j|}{|m_k| \cdot |\theta_k|} \\
\ll &\sum_{\substack{-a_k \le m_k \le a_k \\ m_k \neq 0}} |m_k|^{-\tau} q_k^{-\tau} \frac{a_j q_k q_{k+1}}{q_{j+1}} \ll q_k^{-(\tau-1)} \frac{q_{k+1}}{q_j}.
\end{aligned} \tag{3.16}$$

Hence

$$\begin{aligned}
&\sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \le m_k \le a_k \\ m_k \neq 0}} \sum_{j=k+1}^{\infty} |\hat{h}(m_k q_k)| \frac{\|m_k q_k \tilde{x}'_j q_j \alpha\|}{\|m_k q_k \alpha\|} \\
\ll &\sum_{k=k_-}^{k_+} \sum_{j=k+1}^{\infty} q_k^{-(\tau-1)} q_{k+1} q_j^{-1} \ll \sum_{k=k_-}^{k_+} q_k^{-(\tau-1)} q_{k+1} q_{k+1}^{-1} \\
\ll &q_{k_-}^{-(\tau-1)}.
\end{aligned} \tag{3.17}$$

By adding (3.15) and (3.17) and comparing with (3.13), we know that

$$|H_n^{(1)}(x) - H_n^{(2)}(x)| \ll q_{k_-}^{-(\tau-1)}. \tag{3.18}$$

Because $\tau - 1 > 1$ and $q_{k_-} \gg 2^{\frac{k_-}{2}}$, $q_{k_-}^{-(\tau-1)} \ll 2^{-\frac{k_-}{4}}$. The lemma is verified by combining (3.18) with Lemma 3.3. □

When $|n_k|$ is small enough, H_n can be approximated directly by $n\hat{h}(0)$.

Lemma 3.6. *For k_-, k_+, n and $\{n_k\}$ as in* (3.3),

$$\left| \sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \le m_k \le a_k \\ m_k \neq 0}} \hat{h}(m_k q_k) \frac{e(n_k m_k q_k^2 \alpha) - 1}{e(m_k q_k \alpha) - 1} \right| \ll \sum_{k=k_-}^{k_+} |n_k| q_k^{-(\tau-1)}.$$

and for all $x \in \mathbb{T}^1$,

$$|H_n(x) - n\hat{h}(0)| \ll \sum_{k=k_-}^{k_+} |n_k| q_k^{-(\tau-1)} + 2^{-\frac{k_-}{4}}.$$

Proof. The first quantity is bounded by

$$\sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \le m_k \le a_k \\ m_k \neq 0}} |m_k|^{-\tau} q_k^{-\tau} \cdot |n_k| q_k \ll \sum_{k=k_-}^{k_+} |n_k| q_k^{-(\tau-1)}.$$

For the second inquality, it suffices to note

$$|H_n^{(1)} - n\hat{h}(0)| \leq \sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \leq m_k \leq a_k \\ m_k \neq 0}} \left| \hat{h}(m_k q_k) \frac{e(n_k m_k q_k^2 \alpha) - 1}{e(m_k q_k \alpha) - 1} \right|$$

and combine with Lemma 3.3. □

4. Diophantine relation between rotation numbers

The map T has horizontal rotation number x and vertical rotation number $\hat{h}(0)$. The main result of this part, Proposition 4.2, shows that these two numbers have similar Diophantine profile under our assumption that T has a measurable invariant section.

Assume for a measurable function $g : \mathbb{T}^1 \mapsto \mathbb{T}^1$,

$$g(x+\alpha) - g(x) = h(x), \text{ a.e. } x \in \mathbb{T}^1. \tag{4.1}$$

Lemma 4.1. *There is $\xi_1 \in \mathbb{Z}$, such that for the function $e_{(\xi_1,1)}(x,y) = e(\xi_1 x + y)$, for $\mathrm{m}_{\mathbb{T}^1}$-a.e. x and all y,*

$$\mathop{\mathbb{E}}_{n=1}^{N} (e_{(\xi_1,1)} \circ T^n)(x,y) \not\to 0, \text{ as } N \to \infty.$$

Proof. Notice that measurable function $f(x,y) = e(-g(x)+y)$ satisfies $|f(x,y)| = 1$ and $f \circ T = f$ for almost every x and all y. Decompose $e(g(x))$ as $\sum_{\xi_1 \in \mathbb{Z}} c_{\xi_1} e(\xi_1 x)$ in L^2. Then $f = \sum_{\xi_1 \in \mathbb{Z}} c_{\xi_1} e_{(\xi_1,1)}$ in $L^2(\mathrm{m}_{\mathbb{T}^2})$.

It follows that for at least one ξ_1, $\mathbb{E}_{n=1}^N e_{(\xi_1,1)} \circ T^n$ does not converge to 0 in L^2. Otherwise, $\mathbb{E}_{n=1}^N f \circ T^n$ would converge to 0 in L^2 as well, contradicting the fact that it is always equal to f.

Remark that for the same x, and different y, y',

$$\mathop{\mathbb{E}}_{n=1}^{N} (e_{(\xi_1,1)} \circ T^n)(x,y) = e(y-y') \cdot \mathop{\mathbb{E}}_{n=1}^{N} (e_{(\xi_1,1)} \circ T^n)(x,y').$$

In other words, given x, whether the ergodic average converges to 0 is independent of y. Denote by Ω the set of x for which the averages converge to 0. Then Ω is invariant under $x \mapsto x+\alpha$, and have Lebesgue measure 0 or 1 by ergodicity. Assume $\mathrm{m}_{\mathbb{T}^1}(\Omega) = 1$, then $\mathbb{E}_{n=1}^N e_{(\xi_1,1)} \circ T^n$ converges to 0 pointwisely, and hence also in L^2 by dominated convergence theorem. This contradicts the choice of x. Hence $\mathrm{m}_{\mathbb{T}^1}(\Omega) = 0$. The lemma is proved. □

Proposition 4.2. *For all $\delta > 0$, there exists $k_0 = k_0(\delta) \in \mathbb{N}$ with the following property:*

If $k_0 \leq k_- \leq k_+$, n is as in (3.3), and $|n_k| \leq \min(q_k^{\frac{\tau-1}{2}}, a_k)$ for all $k_- \leq k \leq k_+$, then $\|n\hat{h}(0)\| < \delta$.

The proof follows a similar approach as [Wan15, Lemma 4.1].

Proof. By Luzin's theorem, there is a compact subset $\Omega \subset \mathbb{T}^1$ with $\mathrm{m}_{\mathbb{T}^1}(\Omega) > \frac{1}{2}$, on which g is continuous. There is an $\epsilon > 0$, such that if $x, x' \in \Omega$ satisfies $\|x - x'\| < \epsilon$, then $\|g(x) - g(x')\| < \frac{\delta}{2}$.

We claim that when k_- is sufficiently large, $\|n\alpha\| < \epsilon$. This is because

$$\begin{aligned}\|n\alpha\| = &\left\| \sum_{k=k-}^{k_+} n_k q_k \alpha \right\| \leq \sum_{k=k_-}^{k_+} |n_k| \cdot |\theta_k| < \sum_{k=k_-}^{k_+} a_k q_{k+1}^{-1} < \sum_{k=k_-}^{k_+} q_k^{-1} \\ \ll &q_k^{-1}.\end{aligned}$$

Now assume k_- is large enough and thus $\|n\alpha\| < \epsilon$. The set $\{x \in \mathbb{T}^1 : x, x + n\alpha \in \Omega\}$ at least has Lebesgue measure $1 - 2\mathrm{m}_{\mathbb{T}^1}(\Omega^c) > 0$, and is thus non-empty. Fix a point x from this set, then $\|g(x + n\alpha) - g(x)\| < \frac{\delta}{2}$ by the choice of ϵ.

Since $g(x + n\alpha) = g(x) + H_n(x)$ is true almost everywhere, we can assume it is satisfied by the chosen x. Therefore, $\|H_n(x)\| < \frac{\delta}{2}$.

On the other hand, by Lemma 3.5 and Lemma 3.6,

$$\begin{aligned}&|H_n(x) - n\hat{h}(0)| \\ \ll &\left(\sum_{k=k_-}^{k_+} q_k^{\frac{\tau-1}{2}} q_k^{-(\tau-1)} + 2^{-\frac{k_-}{4}} \right) = \left(\sum_{k=k_-}^{k_+} q_k^{-\frac{\tau-1}{2}} + 2^{-\frac{k_-}{4}} \right) \\ \ll &(q_{k_-}^{-\frac{\tau-1}{2}} + 2^{-\frac{k_-}{4}})\end{aligned}$$

is less than $\frac{\delta}{2}$ when k_- is sufficiently large.

By adding the two estimates above, we see that, given δ,

$$\|n\hat{h}(x)\| \leq \|H_n(x)\| + |H_n(x) - n\hat{h}(0)| < \frac{\delta}{2} + \frac{\delta}{2} = \delta$$

when k_- is sufficiently large. This completes the proof. $\square$

5. Trajectories as sums of random variables

Lemma 3.5 provides an approximation $H_n^{(2)}$ to $H_n(x)$. Next, we examine $H_n^{(2)}$ more carefully and show that it is approximately the sum of a sequence of independent random variables:

$$H_n^{(2)}(x) = \sum_{k=k_-}^{k_+} \big(\tilde{\phi}_k(n_k + \tilde{x}_k) - \tilde{\phi}_k(\tilde{x}_k) \big), \tag{5.1}$$

where

$$\tilde{\phi}_k(l) = l q_k \hat{h}(0) + \sum_{\substack{-a_k \leq m_k \leq a_k \\ m_k \neq 0}} \hat{h}(m_k q_k) \frac{e(l m_k q_k^2 \alpha) - 1}{e(m_k q_k \alpha) - 1}. \tag{5.2}$$

Lemma 5.1. *For all $\delta > 0$, there exists $k_1 \in \mathbb{N}$ with the following property:*

Suppose $k_1 \leq k_- \leq k_+$; $n = \sum_{k=k_-}^{k_+} n_k q_k$, $n' = \sum_{k=k_-}^{k_+} n'_k q_k$ where n_k, n'_k are integers of absolute value bounded by $4a_k$, and $n_k \equiv n'_k (\mathrm{mod}\ a_k)$ for all $k_- \leq k \leq k$, then

$$\left\| \sum_{k=k_-}^{k_+} \left(\tilde{\phi}_k(n_k) - \tilde{\phi}_k(n'_k)\right) \right\| \leq \frac{3\delta}{8}.$$

Proof. Let $n_k - n'_k = b_k a_k$ with $|b_k| \leq 8$. Then

$$(n_k - n'_k) = b_k a_k q_k = b_k q_{k+1} - b_k q_{k-1}. \tag{5.3}$$

Decompose

$$\begin{aligned} &\sum_{k=k_-}^{k_+} \left(\tilde{\phi}_k(n) - \tilde{\phi}_k(n'\right) = \sum_{k=k_-}^{k_+} (n_k - n'_k) q_k \hat{h}(0) + \\ &\quad \sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \leq m_k \leq a_k \\ m_k \neq 0}} \hat{h}(m_k q_k) \frac{e((n_k - n'_k) m_k q_k^2 \alpha) - 1}{e(m_k q_k \alpha) - 1} e(n'_k m_k q_k^2 \alpha). \end{aligned} \tag{5.4}$$

The absolute value of the second term in (5.4) is bounded by

$$\begin{aligned} &\sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \leq m_k \leq a_k \\ m_k \neq 0}} \left| \hat{h}(m_k q_k) \frac{e((n_k - n'_k) m_k q_k^2 \alpha) - 1}{e(m_k q_k \alpha) - 1} \right| \\ &= \sum_{k=k_-}^{k_+} \sum_{\substack{-a_k \leq m_k \leq a_k \\ m_k \neq 0}} \left| \hat{h}(m_k q_k) \frac{e((b_k q_{k+1} - b_k q_{k-1}) m_k q_k \alpha) - 1}{e(m_k q_k \alpha) - 1} \right| \\ &\leq \sum_{k=k_-+1}^{k_++1} \sum_{\substack{-a_{k-1} \leq m_{k-1} \leq a_{k-1} \\ m_{k-1} \neq 0}} \left| \hat{h}(m_{k-1} q_{k-1}) \frac{e(b_{k-1} q_k m_{k-1} q_{k-1} \alpha) - 1}{e(m_{k-1} q_{k-1} \alpha) - 1} \right| \\ &\quad + \sum_{k=k_--1}^{k_+-1} \sum_{\substack{-a_{k+1} \leq m_{k+1} \leq a_{k+1} \\ m_{k+1} \neq 0}} \left| \hat{h}(m_{k+1} q_{k+1}) \frac{e(b_{k+1} q_k m_{k+1} q_{k+1} \alpha) - 1}{e(m_{k+1} q_{k+1} \alpha) - 1} \right|. \end{aligned} \tag{5.5}$$

Here in the last inequality we used the basic fact that $|e(s_1 + s_2) - 1| \leq |e(s_1) - 1| + |e(s_2) - 1|$.

Because $|b_{k-1}|, |b_{k+1}| \leq 8 \ll a_k$, both terms in (5.5) are bounded by the estimate (3.6), respectively after replacing the interval of indices $[k_-, k_+]$ with $[k_- + 1, k_+ + 1]$ and with $[k_- - 1, k_- - 1]$. From the bound on (3.6) in the proof of Lemma 3.3, we get

$$(5.5) \ll 2^{-\frac{k_- - 1}{4}} + 2^{-\frac{k_+ - 1}{4}} \ll 2^{-\frac{k_-}{4}} < \frac{\delta}{8}, \tag{5.6}$$

when k_1 is sufficiently large.

We now focus on the first term from the right hand side in (5.4), which is equal to

$$\begin{aligned}&\sum_{k=k_-}^{k_+}(b_kq_{k+1}-b_kq_{k-1})\hat{h}(0)\\ =&\Big(\sum_{k=k_--1}^{k_+-1}b_{k-1}q_k\Big)\hat{h}(0)+\Big(\sum_{k=k_--1}^{k_+-1}b_{k-1}q_k\Big)\hat{h}(0).\end{aligned} \tag{5.7}$$

If k_1 is large enough, then Proposition 4.2 applies to both terms since $k_- - 1 \geq k_1$. In consequence,

$$\|(5.7)\| \leq \frac{\delta}{8}+\frac{\delta}{8}=\frac{\delta}{4}. \tag{5.8}$$

The lemma follows by adding (5.6) to (5.8). □

Lemma 5.1 says the projection of the function $\tilde{\phi}_k$ to $\mathbb{R} \mapsto \mathbb{Z}$ is almost periodic with period a_k.

Definition 5.2. *Define a function $\phi_k : \mathbb{Z}/a_k\mathbb{Z} \mapsto \mathbb{T}^1$ by*

$$\phi_k(l \bmod a_k) = \tilde{\phi}_k(l) \bmod \mathbb{Z},\ \textit{for } l = 0, \cdots, a_k - 1.$$

We shall identify ϕ_k with an a_k-periodic function on $\mathbb{Z}$ without further notice. No confusion should arise from doing so.

Then Lemma 5.1 actually asserts that

Corollary 5.3. *In the settings of Lemma 5.1,* $\left\| \sum_{k=k_-}^{k_+} \left(\tilde{\phi}_k(n) - \phi_k(n)\right)\right\| \leq \frac{3\delta}{8}$.

We will view the digits n_k from the Ostrowski numeration of an integer as random variables, and $\phi_k(n_k)$ another sequence of random variables decided by n_k. It will be shown in the next section that these variables are approximately independent of each other when k varies.

Proposition 5.4. *Assume Hypothesis 2.4 holds. For all $\delta > 0$, there exists $k_1 = k_1(\delta) \geq 10$, such that:*

If $k_1 \leq k_- < k_+$, then for:

- *all integer sequences $\{\tilde{x}_k\}$ such that $|\tilde{x}_k| \leq 2a_k$ and $\sum_{k=1}^{\infty} \tilde{x}_k\theta_k$ converges to a real value x;*
- *all integers $n = \sum_{k=k_-}^{k_+} n_kq_k$ where $|n_k| \leq a_k$,*

we have

$$\left\| H_n(x) - \sum_{k=k_-}^{k_+} \left(\phi_k(n_k + \tilde{x}_k) - \phi_k(\tilde{x}_k)\right)\right\| < \delta.$$

Proof. By (5.1), the left hand sided is at most

$$\begin{aligned}\left\|H_n(x)-H_n^{(2)}(x)\right\| &+ \left\|\sum_{k=k_-}^{k_+}\left(\tilde{\phi}_k(n_k+\tilde{x}_k)-\phi_k(n_k+\tilde{x}_k)\right)\right\| \\ &+\left\|\sum_{k=k_-}^{k_+}\left(\tilde{\phi}_k(\tilde{x}_k)-\phi_k(\tilde{x}_k)\right)\right\|\end{aligned} \tag{5.9}$$

The first term is bounded by $O(2^{-\frac{k_-}{4}})$ by Lemma 3.5. Because $|\tilde{x}_k|\le 2a_k$ and $|n_k+\tilde{x}_k|\le 3a_k$, the second term and third terms are bounded by $\frac{3\delta}{8}$ by Corollary 5.3. The proposition follows, as for sufficiently large k_1, one can make the $O(2^{-\frac{k_-}{4}})$ less than $\frac{\delta}{4}$. □

6. Independence between digits in Ostrowski numerations

In this section, we take a brief detour to show that the digits n_k, for indices k at which a_k is large, are distributed in an almost independent way for randomly chosen n.

Definition 6.1. *Let $I_{k_-}^{k_+}$ denote the set of integers $n\ge 0$ whose Ostrowski numeration $\{n_k\}$ is supported on $k_-\le k\le k_+$.*

Remark that

$$I_1^k=\{0,\cdots,q_k-1\}. \tag{6.1}$$

Lemma 6.2. *Without loss of generality we can assume $k_-\ge 2$. For all pairs $k_-\le k_+$ and for all $\psi:\mathbb{N}\mapsto\mathbb{C}$ with absolute value bounded by 1,*

$$\left|\mathop{\mathbb{E}}_{n\in I_{k_-}^{k_+}}\psi(n_{k_-})-\mathop{\mathbb{E}}_{l=0}^{a_{k_-}-1}\psi(l)\right|\ll\frac{1}{a_{k_-}}.$$

The implied constant is absolute.

Proof. The lemma is obvious when $k_-=k_+$, in which case n_{k_-} is uniformly distributed on $\{0,\cdots,a_{k_-}\}$. We shall assume $k_-<k_+$ below.

Write $J_l=\{n\in I_{k_-}^{k_+}:n_{k_-}=l\}$ for $l=0,\cdots,a_{k_-}$. And further decompose J_0 as the union of $J_0'=\{n\in J_0:n_{k_-+1}\ne a_{k_-+1}\}$ and $J_0''=\{n\in J_0: n_{k_-+1}=a_{k_-+1}\}$. Recall that if a finitely supported sequence $\{n_k\}$ is known to satisfy $0\le n_k\le a_k$ and $1\le n_1\le a_1-1$, the only restriction for it to be an Ostrowski numeration is that $n_k=0$ if $n_{k+1}=a_{k+1}$. This fact yields a few observations:

(1) If $1\le l,l'\le a_k$, then the operation of replacing n_{k_-} with l' in the Ostrowski numeration is a bijection from J_l to $J_{l'}$;
(2) If $1\le l\le a_k$, then the operation of replacing n_{k_-} with l in the Ostrowski numeration is an injection from J_0' to J_l.

(3) If $1 \le l \le a_k$, then the operation of replacing n_{k_-} with l and n_{k_-+1} with 0 in the Ostrowski numeration is an injection from J_0'' to J_l.

So $|J_0'|, |J_0''| \le |J_1| = |J_2| = \cdots = |J_{a_{k_-}}|$. Since $I_{k_-}^{k_+}$ is the disjoint union of $J_0', J_0'', J_1, \cdots, J_{a_{k_-}}$, the lemma follows. □

For $\mathbf{r} = (r_0, \cdots, r_T) \in \prod_{t=0}^T I_{k_t}^{k_{t+1}-1}$, denote the concatenation of the Ostrowski numerations of the r_t's by

$$\tilde{\mathbf{r}} = \Big((r_0)_{k_0}, \cdots, (r_0)_{k_2-1}, (r_1)_{k_2}, \cdots, (r_1)_{k_3-1}, \cdots, (r_T)_{k_T}, \cdots, (r_T)_{k_{T+1}-1}\Big).$$

Lemma 6.3. *Suppose $k_- = k_0 < k_2 < \cdots < k_{T+1} = k_+ + 1$. If $\psi : \mathbb{N}^{k_+ +1-k_-} \mapsto \mathbb{C}$ is of absolute value bounded by 1, then*

$$\left| \mathop{\mathbb{E}}_{\mathbf{r}\in\prod_{t=0}^T I_{k_t}^{k_{t+1}-1}} \psi(\tilde{\mathbf{r}}) - \mathop{\mathbb{E}}_{n\in I_{k_-}^{k_+}} \psi(n_{k_-}, \cdots, n_{k_+}) \right| \ll \sum_{t=1}^T \frac{1}{a_{k_t}}.$$

The implied constant is absolute.

Proof. Define a map $S : \prod_{t=0}^T I_{k_t}^{k_{t+1}-1} \mapsto \mathbb{N}$ and by $S(r_0, \cdots, r_T) = \sum_{t=0}^T r_t$. Also define $P = (P_0, \cdots, P_T) : \mathbb{N} \mapsto \prod_{t=0}^T I_{k_t}^{k_{t+1}-1}$ by $P_t(n) = \sum_{k=k_t}^{k_{t+1}-1} n_k q_k$. Then the restriction of $S \circ P$ to $I_{k_-}^{k_+}$ is the identity map. Therefore, to prove the proposition, it suffices to show that the complement

$$\Big(\prod_{t=0}^T I_{k_t}^{k_{t+1}-1}\Big) \backslash P(I_{k_-}^{k_+}) \tag{6.2}$$

at most occupies an $O(\sum_{t=1}^T \frac{1}{a_{k_t}})$-portion in $\prod_{t=0}^T I_{k_t}^{k_{t+1}-1}$.

Remark that $(r_0, \cdots, r_T) \in \prod_{t=0}^T I_{k_t}^{k_{t+1}-1}$ is in the set (6.2) if and only if the concatenation $\tilde{\mathbf{r}}$ fails to be an Ostrowski numeration. This can happen only if for some $1 \le t \le T$, $(r_{t-1})_{k_t-1} \neq 0$ but $(r_t)_{k_t} = a_{k_t}$. Hence

$$(6.2) \subset \bigcup_{t=1}^T \Big(J_{k_t}^{k_{t+1}-1} \times \prod_{\substack{0\le t'\le T \\ t'\neq T}} I_{k_{t'}}^{k_{t'+1}-1} \Big), \tag{6.3}$$

where

$$J_{k_t}^{k_{t+1}-1} = \Big\{ r_t \in I_{k_t}^{k_{t+1}-1} : (r_t)_{k_t} \neq a_{k_t} \Big\}.$$

Hence to establish the claim, one only needs to show for all $1 \le t \le T$ that

$$\frac{\left| I_{k_t}^{k_{t+1}-1} \backslash J_{k_t}^{k_{t+1}-1} \right|}{\left| I_{k_t}^{k_{t+1}-1} \right|} \ll \frac{1}{a_{k_t}}. \tag{6.4}$$

This is guaranteed by the previous lemma. □

Proposition 6.4. *For all sequences $k_- \leq k_2 < \cdots < k_T \leq k_+$, and all functions $\psi : \mathbb{N}^T \mapsto \mathbb{C}$ with $|\psi| \leq 1$,*

$$\left| \mathop{\mathbb{E}}_{\substack{0\leq l_t\leq a_{k_t}-1\\ t=1,\cdots,T}} \psi(l_1,\cdots,l_T) - \mathop{\mathbb{E}}_{n\in I_{k_-}^{k_+}} \psi(n_{k_2},\cdots,n_{k_T}) \right| \ll \sum_{t=1}^{T} \frac{1}{a_{k_t}},$$

where $\{n_k\}$ is the Ostrowski numeration of n. The implied constant is absolute.

Proof. By Lemma 6.3,

$$\left| \mathop{\mathbb{E}}_{\substack{r_t\in I_{k_t}^{k_{t+1}-1}\\ t=1,\cdots,T}} \psi\big((r_1)_{k_2},\cdots,(r_T)_{k_T}\big) - \mathop{\mathbb{E}}_{n\in I_{k_-}^{k_+}} \psi(n_{k_2},\cdots,n_{k_T}) \right| \ll \sum_{t=1}^{T} \frac{1}{a_{k_t}}. \quad (6.5)$$

For each given s, by Lemma 6.2,

$$\Bigg| \mathop{\mathbb{E}}_{\substack{r_t\in I_{k_t}^{k_{t+1}-1}\\ t=0,\cdots,s}} \mathop{\mathbb{E}}_{\substack{0\leq l_t\leq a_{k_t}-1\\ t=s+1,\cdots,T}} \psi\big((r_1)_{k_2},\cdots,(r_{s-1})_{k_{s-1}},(r_s)_{k_s},l_{s+1},\cdots,l_T\big)$$
$$- \mathop{\mathbb{E}}_{\substack{r_t\in I_{k_t}^{k_{t+1}-1}\\ t=0,\cdots,s-1}} \mathop{\mathbb{E}}_{\substack{0\leq l_t\leq a_{k_t}-1\\ t=s,\cdots,T}} \psi\big((r_1)_{k_2},\cdots,(r_{s-1})_{k_{s-1}},l_s,l_{s+1},\cdots,l_T\big)\Bigg|$$
$$\ll \frac{1}{a_{k_s}}. \quad (6.6)$$

The proposition is proved by summing (6.6) over $s = 1, \cdots, T$ and adding (6.5). □

7. Invariant section and bounded variance

Recall that Hypothesis 2.4 is assumed throughout, and T has a measurable invariant section. We will see that in consequence, the L^2-variances of the variables $\phi_k : \mathbb{Z}/a_k\mathbb{Z} \mapsto \mathbb{T}^1$ are summable over k.

Definition 7.1. *Denote by $B_{k_-}^{k_+}$ the set of finitely supported integer sequences $\mathbf{n} = \{n_k\}$ such that $0 \leq n_k \leq a_k - 1$ if $k_- \leq k \leq k_+$ and $q_{k+1} > q_k^{\frac{\tau}{2}}$, and $n_k = 0$ otherwise.*

The symbol $[\mathbf{n}]$ will denote $n = \sum_k n_k q_k$ for a finitely supported sequence $\mathbf{n} = \{n_k\}$.

Lemma 7.2. *For k_- sufficiently large, the infinite product*

$$\prod_{\substack{k\geq k_-\\ q_{k+1}>q_k^{\frac{\tau}{2}}}} \left| \mathop{\mathbb{E}}_{l=0}^{a_k-1} e\big(\phi_k(l)\big) \right|$$

converges to a non-zero value.

Proof. Fix an $x \in [-\alpha, 1-\alpha)$ and ξ_1 for which Lemma 4.1 holds. In particular, the conclusion holds for the point $(x, 0)$, i.e. the ergodic average $\mathbb{E}_{n=1}^{N}(e_{(\xi_1,1)} \circ T^n)(x,0)$ does not converge to 0. To be precise, for some $c > 0$, there is a subsequence $N_i \to \infty$, such that

$$\left| \mathop{\mathbb{E}}_{n=1}^{N_i} (e_{(\xi_1,1)} \circ T^n)(x,0) \right| > c. \tag{7.1}$$

For a choice of $k_0' = k_0'(c) \in \mathbb{N}$ that will be specified later and $k_0' \le k_- \le k_+$, choose $N = N_i$ from the subsequence above, such that

$$\frac{q_{k_+ +1}}{N} < \frac{c}{8}. \tag{7.2}$$

Use ensembles of the form $\{L + [\mathbf{n}] : \ \mathbf{n} \in B_{k_-}^{k_+}\}$, $L = 1, \cdots, N$ to cover $\{1, \cdots, N\}$. Here elements in the ensemble are counted with multiplicity.

Each $1 \le l \le N$ are covered exactly $|B_{k_-}^{k_+}|$ times, except for those numbers of distance less than or equal to $\max_{\mathbf{n} \in B_{k_-}^{k_+}} [\mathbf{n}]$ from either 1 or N. The number of such exceptions is bounded by

$$2\sum_{k=1}^{k_+}(a_k - 1)q_k = 2\sum_{k=1}^{k_+}(q_{k+1} - q_{k-1} - q_k) \le 2q_{k_+ +1}.$$

Therefore $\mathbb{E}_{n=1}^{N}\left(e_{(\xi_1,1)} \circ T^n\right)(x,0)$ is approximated by the decomposed average $\mathbb{E}_{L=1}^{N} \mathbb{E}_{\mathbf{n} \in B_{k_-}^{k_+}} \left(e_{(\xi_1,1)} \circ T^{L+[\mathbf{n}]}\right)(x,0)$ up to an error term that is at most $\frac{2q_{k_+ +1}}{N} < \frac{c}{4}$. Hence by (7.1),

$$\mathop{\mathbb{E}}_{L=1}^{N} \left| \mathop{\mathbb{E}}_{\mathbf{n} \in B_{k_-}^{k_+}} \left(e_{(\xi_1,1)} \circ T^{L+[\mathbf{n}]}\right)(x,0) \right| \ge \left| \mathop{\mathbb{E}}_{n=1}^{N_i} (e_{(\xi_1,1)} \circ T^n)(x,0) \right| - \frac{c}{4} > \frac{3c}{4}.$$

In particular, there is some $L \in \{1, \cdots, N\}$ for which

$$\left| \mathop{\mathbb{E}}_{\mathbf{n} \in B_{k_-}^{k_+}} \left(e_{(\xi_1,1)} \circ T^{L+[\mathbf{n}]}\right)(x,0) \right| > \frac{3c}{4}. \tag{7.3}$$

Write $T^L(x, 0) = (x + L\alpha, y)$. For $\mathbf{n} \in B_{k_-}^{k_+}$,

$$\begin{aligned}(e_{(\xi_1,1)} \circ T^{L+[\mathbf{n}]})(x,0) =& e_{(\xi_1,1)}(x + (L + [\mathbf{n}])\alpha, y + H_{[\mathbf{n}]}(x + L\alpha)) \\ =& e(\xi_1(x + L\alpha) + y)e(\xi_1[\mathbf{n}]\alpha + H_{[\mathbf{n}]}(x + L\alpha)).\end{aligned}$$

So (7.3) is equivalent to

$$\left| \mathop{\mathbb{E}}_{\mathbf{n} \in B_{k_-}^{k_+}} e(\xi_1[\mathbf{n}]\alpha + H_{[\mathbf{n}]}(x + L\alpha)) \right| > \frac{3c}{4}. \tag{7.4}$$

We will choose k_0' to be greater than $k_0(\frac{c}{16\pi})$ where k_0 is as in Proposition 5.4. Then with $\{\tilde{x}_k'\}$ denoting the Ostrowski numeration of $x + L\alpha$,

$$\left\| H_{[\mathbf{n}]}(x + L\alpha) - \sum_{k=k_-}^{k_+} \big(\phi_k(n_k + \tilde{x}_k') - \phi_k(\tilde{x}_k')\big) \right\| < \frac{c}{16\pi}. \tag{7.5}$$

In addition, for ξ_1 is fixed,

$$\|\xi_1[\mathbf{n}]\alpha\| = \left\| \xi_1 \sum_{k=k_-}^{k_+} n_k q_k \alpha \right\| \leq \xi_1 \sum_{k=k_-}^{k_+} a_k |\theta_k| < \xi_1 \sum_{k=k_-}^{k_+} q_k^{-1} \ll q_{k_-}^{-1}.$$

So for sufficiently large k_0',

$$\|\xi_1[\mathbf{n}]\alpha\| < \frac{c}{16\pi}. \tag{7.6}$$

Using that $s \mapsto e(s)$ is 2π-Lipschitz, we deduce from (7.4), (7.5), (7.6) that

$$\left| \mathop{\mathbb{E}}_{\mathbf{n} \in B_{k_-}^{k_+}} e\Big(\sum_{k=k_-}^{k_+} \big(\phi_k(n_k + \tilde{x}_k') - \phi_k(\tilde{x}_k')\big)\Big) \right| > \frac{3c}{4} - 2\pi(\frac{c}{16\pi} + \frac{c}{16\pi}) = \frac{c}{2}.$$

By construction of $B_{k-}^{k_+}$, the left hand side is equal to

$$\begin{aligned}
&\prod_{\substack{k_- \leq k \leq k_+ \\ q_{k+1} > q_k^{\frac{\tau}{2}}}} \left| \mathop{\mathbb{E}}_{l=0}^{a_k - 1} e\Big(\big(\phi_k(l + \tilde{x}_k') - \phi_k(\tilde{x}_k')\big)\Big) \right| \\
= &\prod_{\substack{k_- \leq k \leq k_+ \\ q_{k+1} > q_k^{\frac{\tau}{2}}}} \left| \mathop{\mathbb{E}}_{l=0}^{a_k - 1} e\big(\phi_k(l + \tilde{x}_k')\big) \right| \\
= &\prod_{\substack{k_- \leq k \leq k_+ \\ q_{k+1} > q_k^{\frac{\tau}{2}}}} \left| \mathop{\mathbb{E}}_{l=0}^{a_k - 1} e\big(\phi_k(l)\big) \right|,
\end{aligned}$$

where the last step is because ϕ_k is defined on $\mathbb{Z}/a_k\mathbb{Z}$. Hence

$$\prod_{\substack{k_- \leq k \leq k_+ \\ q_{k+1} > q_k^{\frac{\tau}{2}}}} \left| \mathop{\mathbb{E}}_{l=0}^{a_k - 1} e\big(\phi_k(l)\big) \right| > \frac{c}{2}. \tag{7.7}$$

This is true for all $k_0' \leq k_- \leq k_+$.

Given that $\frac{c}{2} > 0$ and $\left|\mathbb{E}_{l=0}^{a_k-1} e\big(\phi_k(l)\big)\right| \leq 1$, the lemma follows. $\square$

Corollary 7.3. *For all $\delta > 0$, there exists $k_1' = k_1'(\delta) \in \mathbb{N}$ such that for all $k_+ \geq k_- \geq k_1'$, there exists a constant z with $|z| \leq 1$, such that*

$$\mathbb{P}_{\mathbf{n} \in B_{k_-}^{k_+}} \left(\left| e\Big(\sum_{k=k_-}^{k_+} \phi_k(n_k) \Big) - z \right| \geq \delta \right) < \delta.$$

Proof. It follows directly from Lemma 7.2 that, when k_- is sufficiently large, for all $k_+ \geq k_-$,

$$\mathbb{E}_{\mathbf{n} \in B_{k_-}^{k_+}} \left| e\Big(\sum_{k=k_-}^{k_+} \phi_k(n_k) \Big) \right| > \sqrt{1 - \delta^3}.$$

Recall the variance for a random variable X is

$$\mathrm{Var}(X) = \mathbb{E}|X|^2 - |\mathbb{E}X|^2 = \mathbb{E}|X - \mathbb{E}X|^2.$$

Since the random variable $X = e\Big(\sum_{k=k_-}^{k_+} \phi_k(n_k) \Big)$ with respect to the uniform probability on $B_{k_-}^{k_+}$ always has absolute value 1, its variance is less than δ^3. Thus

$$\mathbb{P}(|X - \mathbb{E}X| \geq \delta) \leq \frac{\mathbb{E}|X - \mathbb{E}X|^2}{\delta^2} < \delta. \tag{7.8}$$

This proves the Proposition with $z = \mathbb{E}X$, which satisfies $|\mathbb{E}X| \leq 1$. □

The next step is to replace the ensemble $B_{k_-}^{k_+}$ with the natural subset $I_{k-}^{k_+}$ of $\mathbb{N}$ without heavily distorting the probability distribution.

Proposition 7.4. *For all $\delta > 0$, there exists $k_2 = k_2(\delta) \in \mathbb{N}$ such that for all $k_3 \leq k_- \leq k_+$, there exists a constant z with $|z| \leq 1$, such that*

$$\mathbb{P}_{n \in I_{k_-}^{k_+}} \Big(\Big| e\Big(\sum_{\substack{k_- \leq k \leq k_+ \\ q_{k+1} > q_k^{\frac{\tau}{2}}}} \phi_k(n_k) \Big) - z \Big| \geq \delta \Big) < \delta.$$

Proof. By Corollary 7.3 and Proposition 6.4, for sufficiently large k_2 and all $k_+ \geq k_- \geq k_2$, there exists z with $|z| \leq 1$ such that

$$\mathbb{P}_{n \in I_{k_-}^{k_+}} \Big(\Big| e\Big(\sum_{\substack{k_- \leq k \leq k_+ \\ q_{k+1} > q_k^{\frac{\tau}{2}}}} \phi_k(n_k) \Big) - z \Big| \geq \frac{\delta}{4} \Big) < \frac{\delta}{4} + O\Big(\sum_{\substack{k_- \leq k \leq k_+ \\ q_{k+1} > q_k^{\frac{\tau}{2}}}} \frac{1}{a_k} \Big). \tag{7.9}$$

We can choose the lower bound k_2 such that the term $O(\sum_{\substack{k_- \leq k \leq k_+ \\ q_{k+1} > q_k^{\frac{\tau}{2}}}} \frac{1}{a_k})$ is less than $\frac{\delta}{4}$ for all $k_+ \geq k_- \geq k_2$. This is because for the indices k involved,

$$a_k = \frac{q_{k+1} - q_{k-1}}{q_k} \geq \frac{q_{k+1}}{q_k} - 1 \gg q_k^{\frac{\tau}{2} - 1}, \tag{7.10}$$

which implies a_k has exponential growth with respect to k. Therefore (7.9) is bounded by $\frac{\delta}{2}$.

When $q_{k+1} \le q_k^{\frac{\tau}{2}}$, by Hypothesis 2.4 and the definition of ϕ_k, $\phi_k(n_k) = \tilde{\phi}_k(n_k) = n_k q_k \hat{h}(0)$, thus

$$\sum_{\substack{k_- \le k \le k_+ \\ q_{k+1} \le q_k^{\frac{\tau}{2}}}} \phi_k(n_k) = \Big(\sum_{\substack{k_- \le k \le k_+ \\ q_{k+1} \le q_k^{\frac{\tau}{2}}}} n_k q_k \Big) \hat{h}(0).$$

Note $0 \le n_k \le a_k \le q_{k+1} q_k^{-1} \le q_k^{\frac{\tau}{2}-1}$ if $q_{k+1} \le q_k^{\frac{\tau}{2}}$. So by Proposition 4.2, if k_- is sufficiently large, then for all $n \in I_{k_-}^{k_+}$,

$$\Big| \sum_{\substack{k_- \le k \le k_+ \\ q_{k+1} \le q_k^{\frac{\tau}{2}}}} \phi_k(n_k) \Big| < \frac{\delta}{4\pi},$$

which yields that

$$\Big| e\Big(\sum_{\substack{k_- \le k \le k_+ \\ q_{k+1} > q_k^{\frac{\tau}{2}}}} \phi_k(n_k) \Big) - e\Big(\sum_{k=k_-}^{k_+} \phi_k(n_k) \Big) \Big| < 2\pi \cdot \frac{\delta}{4\pi} = \frac{\delta}{2}. \tag{7.11}$$

The statement of the proposition follows by adding this bound to the inequaltiy (7.9) $< \frac{\delta}{2}$. $\square$

8. Reduction to finitely many scales

In order to show for the map T on $\mathbb{T}^2$, it suffices to consider test functions $e_{(\zeta_1,\zeta_2)}(x,y) = e(\zeta_1 x + \zeta_2 y)$, $\zeta_1, \zeta_2 \in \mathbb{Z}$, since they span a dense subspace in $C^0(\mathbb{T}^2)$. By taking complex conjugate, we can also assume $\zeta_2 \ge 0$. From now on, fix such a function $f(x,y) = e_{(\zeta_1,\zeta_2)}(x,y)$.

The aim is to show that for any given $\delta > 0$, for suffciently large N,

$$\Big| \mathop{\mathbb{E}}_{n=0}^{N-1} \mu(n)(f \circ T^n)(x,y) \Big| \tag{8.1}$$

is bounded by δ for all x and y.

Define $I_1^{k_+}$ according to Definition 6.1 and notice that it coincides with $\{0, \cdots, q_{k_++1} - 1\}$. We decompose the correlation average between f and μ into short averages along intervals of length $q_{k_++1} - 1$. In fact,

$$\Big| \mathop{\mathbb{E}}_{n=0}^{N-1} \mu(n)(f \circ T^n)(x,y) - \mathop{\mathbb{E}}_{L=0}^{N-1} \mathop{\mathbb{E}}_{n=1}^{q_{k_++1}-1} \mu(L+n)(f \circ T^{L+n})(x,y) \Big| \ll \frac{q_{k_++1}}{N}, \tag{8.2}$$

since the two averages only differ at worst by an $O(\frac{q_{k_++1}}{N})$-portion of elements.

Together with (8.2), Lemma 6.3 shows

$$\begin{aligned}(8.1) \leq \mathop{\mathbb{E}}_{L=0}^{N-1} &\left| \mathop{\mathbb{E}}_{r\in I_1^{k_--1}} \mathop{\mathbb{E}}_{s\in I_{k_-}^{k_+}} \mu(L+r+s)(f\circ T^{L+r+s})(x,y)\right| \\ &+ O\Big(\frac{q_{k_++1}}{N}+\frac{1}{a_{k_-}}\Big).\end{aligned} \tag{8.3}$$

One can write $(f\circ T^{L+r+s})(x,y)$ as

$$\begin{aligned}&e_{(\zeta_1,\zeta_2)}\Big(x+(L+r+s)\alpha, \\ &\qquad y+H_L(x)+H_r(x+L\alpha)+H_s(x+(L+r)\alpha)\Big) \\ =&e\Big(\zeta_1(r+s)\alpha+\zeta_2H_r(x+L\alpha)+\zeta_2H_n(x+(L+r)\alpha)\Big)\cdot e(\beta_{x,L}),\end{aligned} \tag{8.4}$$

where $\beta_{x,L}$ is independent of n and r. Therefore

$$\begin{aligned}&\left|\mathop{\mathbb{E}}_{r\in I_1^{k_--1}} \mathop{\mathbb{E}}_{s\in I_{k_-}^{k_+}} \mu(L+r+s)(f\circ T^{L+r+s})(x,y)\right| \\ =&\Bigg|\mathop{\mathbb{E}}_{r\in I_1^{k_--1}} \mathop{\mathbb{E}}_{s\in I_{k_-}^{k_+}} \mu(L+r+s) \\ &\qquad \cdot e\Big(\zeta_1(r+s)\alpha+\zeta_2H_r(x+L\alpha)+\zeta_2H_n(x+(L+r)\alpha\Big)\Bigg|\end{aligned} \tag{8.5}$$

Lemma 8.1. *For all $\delta>0$, if k_- is large enough, then for all $k_+\geq k_-$ and all $x\in\mathbb{T}^1$, $L\in\mathbb{N}$,*

$$\begin{aligned}&\left|\mathop{\mathbb{E}}_{r\in I_1^{k_--1}} \mathop{\mathbb{E}}_{s\in I_{k_-}^{k_+}} \mu(L+r+s)(f\circ T^{L+r+s})(x,y)\right| \\ \leq&\left|\mathop{\mathbb{E}}_{r\in I_1^{k_--1}} \mathop{\mathbb{E}}_{s\in I_{k_-}^{k_+}} \mu(L+r+s)e\Big(\zeta_1(r+s)\alpha+\zeta_2H_r(x+L\alpha)\Big)\right|+\frac{\delta}{8}.\end{aligned}$$

Proof. By Proposition 7.4, when $k_3\geq k_2(\frac{\delta}{16\zeta_2+16})$, there is a constant z_1 that is independent of x, L, r and s, such that $|z_1|\leq 1$, and with probability $1-\frac{\delta}{16}$ or higher for a randomly chosen $s\in I_{k_-}^{k_+}$,

$$\left|e\Big(\sum_{k=k_-}^{k_+}\phi_k(s_k)\Big)-z_1\right|<\frac{\delta}{16\zeta_2+16}. \tag{8.6}$$

We write $x+(L+r)\alpha\in\mathbb{T}^1$ as $\sum_{k=1}^{\infty}(\tilde{x}_k+L_k+r_k)\theta_k$. Remark that by construction, $r_k=0$ for $k\geq k_-$. Then, again with probability at least

$1-\frac{\delta}{16}$,

$$\left|e\Big(\sum_{k=k_-}^{k_+}\phi_k\big(s_k+(\tilde{x}_k+L_k)\big)-\phi_k(\tilde{x}_k+L_k)\Big)-z_2\right|<\frac{\delta}{16\zeta_2+16}, \tag{8.7}$$

where $z_2=z_1e(-\phi_k(\tilde{x}_k+L_k))$ depends on x and L but not on r and n. Here we used the fact that ϕ_k is a function defined on $\mathbb{Z}/a_k\mathbb{Z}$.

By Proposition 5.4, (8.7) becomes

$$\begin{aligned}\left|e\big(H_s(x+(L+r)\alpha)\big)-z_2\right|<&\frac{\delta}{16\zeta_2+16}+\frac{\delta}{16\zeta_2+16}\\=&\frac{\delta}{16\zeta_2+16},\end{aligned} \tag{8.8}$$

if we choose $k_3\geq k_0(\frac{1}{2\pi}\cdot\frac{\delta}{16\zeta_2+16})$. Note that Proposition 5.4 applies because $\{n_k\}$ is an Ostrowski numeration, and $|\tilde{x}_k+L_k+r_k|\leq 3|a_k|$ for all k.

Remark a simple fact: if $|w_1|,|w_2|\leq 1$, then

$$|w_1^{\zeta_2}-w_2^{\zeta_2}|=|w_1-w_2|\cdot\Big|\sum_{j=0}^{\zeta_2}w_1^jw_2^{\zeta_2-j}\Big|\leq(\zeta_2+1)\cdot|w_1-w_2|.$$

So

$$\left|e\big(\zeta_2H_s(x+(L+r)\alpha)\big)-z_2^{\zeta_2}\right|<(\zeta_2+1)\cdot\frac{\delta}{16\zeta_2+16}<\frac{\delta}{16}. \tag{8.9}$$

Remember that this does not necessarily hold always, but with a probability higher than $1-\frac{\delta}{16\zeta_2+16}$ with respect to $s\in I_{k_-}^{k_+}$.

Plugging this estimate into (8.5), we see that, thanks to the fact that $z_2^{\zeta_2}$ doesn't depend on r or n,

$$\begin{aligned}&\Bigg|\mathop{\mathbb{E}}_{r\in I_1^{k_--1}}\mathop{\mathbb{E}}_{s\in I_{k_-}^{k_+}}\mu(L+r+s)(f\circ T^{L+r+s})(x,y)\\&\quad-z_2^{\zeta_2}\mathop{\mathbb{E}}_{r\in I_1^{k_--1}}\mathop{\mathbb{E}}_{s\in I_{k_-}^{k_+}}\mu(L+r+s)e\Big(\zeta_1(r+s)\alpha+\zeta_2H_r(x+L\alpha)\Big)\Bigg|\\<&\frac{\delta}{16\zeta_2+16}+\frac{\delta}{16}<\frac{\delta}{8}.\end{aligned} \tag{8.10}$$

We obtain the lemma because $|z_2|\leq 1$ and $\zeta_2\geq 0$. □

For each $n\geq 0$, truncate its Ostrowski numeration before the k_--th digit, defining a function $r=r(n)=\sum_{k=1}^{k_--1}n_kq_k\in I_1^{k_--1}$.

Corollary 8.2. *For all $\delta>0$, there exists $k_3=k_3(\alpha,\delta,\zeta_1,\zeta_2)\in\mathbb{N}$, such that:*

For all $k_+ \geq k_- \geq k_3$*, and all* $x \in \mathbb{T}^1$*,*

$$\begin{aligned}&\mathop{\mathbb{E}}_{n=0}^{N-1} \mu(n)(f \circ T^n)(x,y)\\ \leq &\mathop{\mathbb{E}}_{L=0}^{N-1} \left| \mathop{\mathbb{E}}_{n \in I_1^{k_+}} \mu(L+n) e\big(\zeta_1 n\alpha + \zeta_2 H_{r(n)}(x + L\alpha)\big) \right| + \frac{\delta}{8} + O\Big(\frac{q_{k_+ + 1}}{N} + \frac{1}{a_{k_-}}\Big)\end{aligned}$$

Proof. This follows from the inequality (8.3), Lemma 8.1, and one more application of Lemma 6.3. □

9. Proof of the main theorem

We can think of $r(n)$ as a quasi-periodic residue of n. Indeed, from the construction of the Ostrowski numeration, one can easily see the following fact.

Lemma 9.1. *For each* $r \in I_1^{k_- - 1} = \{0, \cdots, q_{k_-} - 1\}$*, there is an arc*[2] $D_r \subset \mathbb{T}^1$ *such that:*

(1) $\mathbb{T}^1$ *decomposes as a disjoint union* $\bigsqcup_{r=0}^{q_{k_-}-1} D_r$*;*
(2) For $n \geq 0$*,* $r(n) = r$ *if and only if* $n\alpha \in D_r$ *modulo* $\mathbb{Z}$*.*

The lemma allows to interprete $e\big(\zeta_1 n\alpha + \zeta_2 H_{r(n)}(x + L\alpha)\big)$ as a function of $n\alpha$.

Proposition 9.2. *For all* $\delta > 0$ *and* $f(x,y) = e_{(\zeta_1,\zeta_2)}(x,y)$ *, there exists* $k_3 = k_3(\alpha, \delta, \zeta_1, \zeta_2) \in \mathbb{N}$*, such that:*

For all $k_- \geq k_3$ *, there exist* $A = A(\alpha, \delta, k_-)$*,* $B = B(\alpha, \delta, k_-)$*and* $k_4 = k_4(\alpha, \delta, k_-)$*, such that for all* $(x,y) \in \mathbb{T}^2$ *and* $k_+ \geq k_4$*,*

$$\begin{aligned}&\left| \mathop{\mathbb{E}}_{n=0}^{N-1} \mu(n)(f \circ T^n)(x,y) \right|\\ \leq &B \cdot \max_{\xi=-A}^{A} \mathop{\mathbb{E}}_{L=0}^{N-1} \left| \mathop{\mathbb{E}}_{n=0}^{q_{k_+}-1} \mu(L+n) e\big((\zeta_1 + \xi) n\alpha\big) \right| + \frac{\delta}{4} + O\Big(\frac{q_{k_+ + 1}}{N} + \frac{1}{a_{k_-}}\Big).\end{aligned}$$

Proof. Suppose k_- is fixed, for every $r = 0, \cdots, q_{k_-} - 1$, choose a continuous function $\psi_r : \mathbb{T}^1 \mapsto [0,1]$ approximating $\mathbf{1}_{D_r}$ in the sense that $\psi_r = \mathbf{1}_{D_r}$ except in two short arcs U_r^- and U_r^+ respectively around both ends of D_r with total length $|U_r^-| + |U_r^+| < \frac{\delta}{32 q_{k_-}}$.

Furthermore, because trigonometric polynomials are dense in $C^0(\mathbb{T}^1)$, one can find $A_r \in \mathbb{N}$ such that $\psi_r(w)$ is approximated by a trigonometric polynomial $\sum_{\xi=-A_r}^{A_r} \theta_{r,\xi} e(\xi w)$ up to error $\frac{\delta}{16 q_{k_-}}$ in C^0 norm. By taking maximum over all r, we can make $A_r = A$ a constant that is determined by α, δ and k_-.

[2]Here an arc can be either open or closed at each of its endpoints.

When L and x are given, the function

$$\sum_{r=0}^{q_{k_-}-1} e(\zeta_2 H_r(x+L\alpha))\mathbf{1}_{D_r}(w) \tag{9.1}$$

is approximated by

$$\sum_{\xi=-A}^{A} c_{\xi,L,x} e(\xi w) := \sum_{r=0}^{q_{k_-}-1} e(\zeta_2 H_r(x+L\alpha)) \sum_{\xi=-A}^{A} \theta_{r,\xi} e(\xi w) \tag{9.2}$$

up to an error of $q_{k_-} \cdot \frac{\delta}{16 q_{k_-}} = \frac{\delta}{16}$, unless w lies in the exceptional set $U = \bigcup_{r=0}^{q_{k_-}-1}(U_r^- \cup U_r^+)$.

Let $B_0 = \max_{\xi=-A}^{A} \sum_{r=0}^{q_{k_-}-1} |\theta_{r,\xi}|$, then B_0 is determined by α, δ and k_-, and satisfies

$$\max_{\xi=-A}^{A} |c_{\xi,L,x}| \leq B_0, \forall L \geq 0, \forall x \in \mathbb{T}^1. \tag{9.3}$$

By Lemma 9.1, $e(\zeta_2 H_{r(n)}(x+L\alpha)) = \sum_{r=0}^{q_{k_-}-1} e(\zeta_2 H_r(x+L\alpha))\mathbf{1}_{D_r}(n\alpha)$. Therefore,

$$\begin{aligned}&\left|\mu(L+n)e\big(\zeta_1 n\alpha + \zeta_2 H_{r(n)}(x+L\alpha)\big) - \sum_{\xi=-A}^{A} c_{\xi,L,x}\mu(L+n)e\big((\zeta_1+\xi)n\alpha\big)\right|\\ \leq&\frac{\delta}{16}\end{aligned}$$

unless $n\alpha \in U$.

Because $|U| = \frac{\delta}{32}$ and the sequence $\{n\alpha\}$ is equidistributed in $\mathbb{T}^1$, for some Q_0 that depends on U, $\mathbb{P}_{0\leq n\leq Q-1}(n\alpha \in U) \leq \frac{\delta}{16}$ for all $Q \geq Q_0$. Choose the smallest k_4 such that $q_{k_4} > Q_0$, then k_4 is determined by α, δ and the choice of U, which in turn depends on k_-. Therefore, for all L, x and $k_+ \geq k_4$,

$$\begin{aligned}&\Bigg|\mathop{\mathbb{E}}_{n=0}^{q_{k_+}-1} \mu(L+n)e\big(\zeta_1 n\alpha + \zeta_2 H_{r(n)}(x+L\alpha)\big)\\ &\qquad - \mathop{\mathbb{E}}_{n=0}^{q_{k_+}-1} \sum_{\xi=-A}^{A} c_{\xi,L,x}\mu(L+n)e\big((\zeta_1+\xi)n\alpha\big)\Bigg|\\ \leq&\frac{\delta}{16}+\frac{\delta}{16}=\frac{\delta}{8}.\end{aligned} \tag{9.4}$$

Plugging (9.4) into the inequality in Corollary 8.2 yields that

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{n=0}^{N-1}\mu(n)(f\circ T^n)(x,y)\\
\leq&\mathop{\mathbb{E}}_{L=0}^{N-1}\left|\mathop{\mathbb{E}}_{n=0}^{q_{k_+}-1}\sum_{\xi=-A}^{A}c_{\xi,L,x}\mu(L+n)e\big((\zeta_1+\xi)n\alpha\big)\right|+\frac{\delta}{4}+O\Big(\frac{q_{k_+ +1}}{N}+\frac{1}{a_{k_-}}\Big)\\
\leq&\mathop{\mathbb{E}}_{L=0}^{N-1}\sum_{\xi=-A}^{A}|c_{\xi,L,x}|\cdot\left|\mathop{\mathbb{E}}_{n=0}^{q_{k_+}-1}\mu(L+n)e\big((\zeta_1+\xi)n\alpha\big)\right|+\frac{\delta}{4}+O\Big(\frac{q_{k_+ +1}}{N}+\frac{1}{a_{k_-}}\Big)\\
\leq&B_0\cdot\sum_{\xi=-A}^{A}\mathop{\mathbb{E}}_{L=0}^{N-1}\left|\mathop{\mathbb{E}}_{n=0}^{q_{k_+}-1}\mu(L+n)e\big((\zeta_1+\xi)n\alpha\big)\right|+\frac{\delta}{4}+O\Big(\frac{q_{k_+ +1}}{N}+\frac{1}{a_{k_-}}\Big).
\end{aligned}
$$

The proposition follows by setting $B = B_0 \cdot (2A + 1)$. □

The proof of Theorem 1.1 relies on the recent theorem of Matomäki, Radziwiłł, and Tao [MRT15] on averages of non-pretentious multiplicative functions in short intervals, or more precisely the following corollary of it:

Proposition 9.3. [MRT15] *For some constant c, for all $N \geq R \geq 10$ and $\beta \in \mathbb{R}$,*

$$\mathop{\mathbb{E}}_{L=0}^{N-1}\left|\mathop{\mathbb{E}}_{n=1}^{R}\mu(L+n)e(\beta n)\right| \ll (\log N)^{-c}+\frac{\log\log R}{\log R}.$$

The implied constant does not depend on β.

This is [MRT15, Theorem 1.7], applied to μ in light of the paragraph before the theorem in that paper, which asserts that μ is sufficiently non-pretentious.

Proof of Theorem 1.1. We now give the proof of our main result. Recall that it suffices to work under Hypothesis 2.4 and assume that $f(x,y) = e_{(\zeta_1,\zeta_2)}(x,y)$. In this case Proposition 9.2 applies.

Using Proposition 9.3, the inequality in Proposition 9.2 becomes

$$
\begin{aligned}
&\left|\mathop{\mathbb{E}}_{n=0}^{N-1}\mu(n)(f\circ T^n)(x,y)\right|\\
\leq&O\Big(B(\log N)^{-c}+B\cdot\frac{\log\log q_{k_+ +1}}{\log q_{k_+ +1}}\Big)+\frac{\delta}{4}+O\Big(\frac{q_{k_+ +1}}{N}+\frac{1}{a_{k_-}}\Big).
\end{aligned}
\tag{9.5}
$$

Given α, δ, ζ_1, ζ_2, we can first choose a large $k_- \geq k_3(\alpha,\delta,\zeta_1,\zeta_2)$ which satisfies the extra condition that $q_{k_- +1} > q_{k_-}^{\frac{\tau}{2}}$. Recall that there are infinitely many such k_- under Hypothesis 2.4. Then by (7.10), one may fix a sufficiently large k_- for which the $O(\frac{1}{a_{k_-}})$ term is bounded by $\frac{\delta}{8}$. Note B is determined once k_- is fixed. One can then pick a large $k_+ \geq k_4(\alpha,\delta,k_-)$ such that the $O\big(B\cdot\frac{\log\log q_{k_+ +1}}{\log q_{k_+ +1}}\big)$ part is bounded by $\frac{\delta}{8}$. Finally, in order to make the entire sum in (9.5) less than δ, it suffices to make N much larger compared with q_{k_+} and B to make the two other terms arbitrarily small. Since δ is arbitrary, the proof is completed. $\square$

References

[Ber01] V. Berthé, *Autour du système de numération d'Ostrowski*, Bull. Belg. Math. Soc. Simon Stevin **8** (2001), no. 2, 209–239 (French, with French summary).

[Bou13a] J. Bourgain, *On the correlation of the Moebius function with rank-one systems*, J. Anal. Math. **120** (2013), 105–130.

[Bou13b] J. Bourgain, *Möbius-Walsh correlation bounds and an estimate of Mauduit and Rivat*, J. Anal. Math. **119** (2013), 147–163.

[BSZ13] J. Bourgain, P. Sarnak, and T. Ziegler, *Disjointness of Möbius from horocycle flows*, From Fourier analysis and number theory to Radon transforms and geometry, Dev. Math., vol. 28, Springer, New York, 2013, pp. 67–83.

[Dav37] H. Davenport, *On some infinite series involving arithmetical functions II*, Quat. J. Math. **8** (1937), 313–320.

[EALdlR14] E. H. El Abdalaoui, M. Lemańczyk, and T. de la Rue, *On spectral disjointness of powers for rank-one transformations and Möbius orthogonality*, J. Funct. Anal. **266** (2014), no. 1, 284–317.

[FKPLM15] S Ferenzi, J. Kułaga-Przymus, M. Lemańczyk, and C Mauduit, *Substitutions and Mbius disjointness*, preprint (2015).

[Fur61] H. Furstenberg, *Strict ergodicity and transformation of the torus*, Amer. J. Math. **83** (1961), 573–601.

[Gre12] B. Green, *On (not) computing the Möbius function using bounded depth circuits*, Combin. Probab. Comput. **21** (2012), no. 6, 942–951.

[GT12] B. Green and T. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) **175** (2012), no. 2, 541–566.

[KPL15] J. Kułaga-Przymus and M. Lemańczyk, *The Möbius function and continuous extensions of rotations*, Monatsh. Math. **178** (2015), no. 4, 553–582.

[Khi97] A. Ya. Khinchin, *Continued fractions*, Translated from the third (1961) Russian edition, Dover Publications, Inc., Mineola, NY, 1997. With a preface by B. V. Gnedenko; Reprint of the 1964 translation.

[LS15] J. Liu and P. Sarnak, *The Möbius function and distal flows*, Duke Math. J. **164** (2015), no. 7, 1353–1399.

[MMR14] B. Martin, C. Mauduit, and J. Rivat, *Théoréme des nombres premiers pour les fonctions digitales*, Acta Arith. **165** (2014), no. 1, 11–45 (French).

[MRT15] K. Matomäki, M. Radziwiłł, and T. Tao, *An averaged form of Chowla's conjecture*, preprint (2015).

[MR10] C. Mauduit and J. Rivat, *Sur un problème de Gelfond: la somme des chiffres des nombres premiers*, Ann. of Math. (2) **171** (2010), no. 3, 1591–1646 (French, with English and French summaries).

[Ost22] A. Ostrowski, *Bemerkungen zur Theorie der Diophantischen Approximationen*, Abh. Math. Sem. Univ. Hamburg **1** (1922), no. 1, 77–98 (German).

[Pec15] R. Peckner, *Möbius disjointness for homogeneous dynamics*, preprint (2015).

[Sar09] P. Sarnak, *Three lectures on the Möbius function, randomness and dynamics*, lecture notes, IAS (2009).

[Vau77] R.-C. Vaughan, *Sommes trigonométriques sur les nombres premiers*, C. R. Acad. Sci. Paris Sér. A-B **285** (1977), no. 16 (French, with English summary).

[Wan15] Z. Wang, *Möbius disjointness for analytic skew products*, preprint (2015).

Pennsylvania State University, University Park, PA 16802, USA
E-mail address: `zhirenw@psu.edu`