

# Möbius disjointness for analytic skew products

Zhiren Wang<sup>1</sup>

Received: 29 September 2015 / Accepted: 12 November 2016 /

Published online: 28 November 2016

© Springer-Verlag Berlin Heidelberg 2016

**Abstract** We show that the Möbius function is disjoint to every analytic skew product dynamical system on  $\mathbb{T}^2$  over a rotation of the circle.

## Contents

1 Introduction . . . . .	175
2 Existence of a coboundary . . . . .	178
3 Eliminating non-resonant frequencies . . . . .	182
4 Almost linearity of orbit along arithmetic progressions . . . . .	185
5 Almost periodicity of orbit . . . . .	186
6 Averages in short intervals . . . . .	189
References . . . . .	195

## 1 Introduction

Let  $h$  be a continuous map from the circle  $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$  to itself. For  $\alpha \in [0, 1)$  we consider the skew transform on  $\mathbb{T}^2$  given by

$$T(x, y) = (x + \alpha, y + h(x)). \quad (1.1)$$

We show the following Möbius disjointness statement:

---

✉ Zhiren Wang  
zhirenw@psu.edu

<sup>1</sup> Pennsylvania State University, University Park, PA 16802, USA

**Theorem 1.1** *Suppose  $h : \mathbb{T}^1 \mapsto \mathbb{T}^1$  is analytic and  $T$  is as above. Then*

$$\frac{1}{N} \sum_{n \leq N} \mu(n) f(T^n(x_0, y_0)) \rightarrow 0, \quad \forall (x_0, y_0) \in \mathbb{T}^2, \quad \forall f \in C^0(\mathbb{T}^2). \quad (1.2)$$

We emphasize that there is no assumption on  $\alpha$ .

The Möbius function is defined by  $\mu(n) = (-1)^k$  if  $n$  is the product of  $k$  distinct primes and  $\mu(n) = 0$  otherwise. An important theme in dynamical system and number theory during the recent years is the randomness of  $\mu$ , which is characterized by Sarnak's Möbius Disjointness Conjecture:

**Conjecture 1.2** [18] *If  $X$  is a compact metric space,  $T : X \mapsto X$  is a continuous map with zero topological entropy, then  $\frac{1}{N} \sum_{n \leq N} \mu(n) f(T^n(x)) \rightarrow 0$  for all continuous functions  $f$  on  $X$  and all  $x \in X$ .*

An important feature of the conjecture is that disjointness is expected to hold for every  $x$  instead of for almost every  $x$  with respect to some  $T$ -invariant probability measure.

The case where  $X$  is finite is equivalent to the prime number theorem in arithmetic progressions. And the case where  $T$  is a rotation on the circle is Davenport's theorem [4] that  $\sum_{n \leq N} \mu(n) e(\alpha n) = o(N)$  uniformly for all  $\alpha$ .

Many special cases of Sarnak's Conjecture have been established. To list a few: [1–3, 5, 7, 8, 11, 14, 16, 17]. The majority of these results rely on Vinogradov's bilinear method, or one of its newer variants: either Vaughan's identity [19] or the Bourgain–Sarnak–Ziegler criterion [3].

It should be remarked that in all the above cases, the dynamical system is regular in the sense that for every point  $x$ ,  $\frac{1}{N} \sum_{n \leq N} \delta_{T^n(x)}$  converges to some  $T$ -invariant probability measure on  $X$  in weak- $*$  topology.

Skew products on  $\mathbb{T}^2$  provide the simplest examples of irregular dynamics. In the case of  $C^{1+\epsilon}$  skew products, for Diophantine rotation numbers  $\alpha$ , which have full Lebesgue measure, the dynamics is topologically conjugate to an affine transformation, hence is regular and Conjecture 1.2 holds. Kułaga-Przymus and Lemańczyk showed in [11] that this indeed holds for a much larger topologically generic set of  $\alpha$ . However, Furstenberg constructed in [6] a counterexample showing that for some  $\alpha$  and some analytic function  $h$ , the dynamics of  $T$  is not regular.

In [13], Liu and Sarnak established Conjecture 1.2 for a class of analytic skew products of the form (1.1). This was a remarkable achievement because their result covers a topological conjugate of Furstenberg's counterexample, and therefore it was the first time that Conjecture 1.2 was proved for an irregular dynamical system. The hypothesis in [13] was that the Fourier coefficients  $h$  decay with exponential bounds both from above and from below, i.e.,  $e^{-\tau_1|m|} < |\hat{h}(m)| < e^{-\tau_2|m|}$  for some  $\tau_1, \tau_2 > 0$ . The upper bound cor-

responds to analyticity, while the lower bound was posed in order to get extra cancellation using van der Corput’s method of exponential sums.

The main result in this paper, Theorem 1.1, proves Conjecture 1.2 for every analytic skew product on  $\mathbb{T}^2$  without assuming a lower bound for  $|\hat{h}(m)|$ .

Our proof is a mixture of dynamical and number-theoretical arguments. It is based on a dynamical dichotomy due to Furstenberg [6], and uses the Bourgain–Sarnak–Ziegler criterion from [3]. The main new ingredient is the recent estimate by Matomäki–Radziwiłł–Tao [15] on averages of multiplicative functions in short intervals.

In [6], it was shown that  $T$  either is uniquely ergodic or preserves an invariant measure that is a finite extension of the Lebesgue measure in the  $x$  coordinate. We apply this to a two-parameter family of maps of the form

$$(x, y) \mapsto \left( x + \alpha, y + \sum_{l=0}^{p_1-1} h(x_0 + p_1x + l\alpha) - \sum_{l=0}^{p_2-1} h(x_0 + p_2x + l\alpha) \right)$$

where  $(p_1, p_2)$  are pairs of distinct primes.

If unique ergodicity holds for all pairs of primes, it will follow that there is no correlation between the sequences  $f(T^{p_1^n}(x_0, y_0))$  and  $f(T^{p_2^n}(x_0, y_0))$  and the Bourgain-Sarnak-Ziegler criterion allows to conclude (1.2).

In the case where unique ergodicity fails for at least one pair  $(p_1, p_2)$ , using Fourier series, we are able to show, after several steps of dynamical reductions, that in order for (1.2) to fail, the sequence  $n \mapsto f(T^n(x_0, y_0))$  must be well approximated by a periodic sequence of period  $Q$  in every interval of length  $l$ . The analyticity assumption allows to make  $l$  exponentially large in terms of  $Q$ , while  $Q$  can be arbitrarily large itself. Then we decompose, in each interval of length  $l$ , the periodic sequence into a linear combination of Dirichlet characters  $\chi$  of modulus  $Q$ . This reduces the problem to controlling the average of the multiplicative function  $\mu(n)\chi(n)$  on a typical interval of length  $l$ .

Matomäki, Radziwiłł and Tao proved in [15] that for a non-pretentious multiplicative function  $\nu$  and  $X \gg l \gg 1$ , on most intervals  $I \subset [X, 2X)$  of length  $l$  the average of  $\nu$  is very small. The function  $\mu\chi$  is known to be non-pretentious, so this theorem applies in our setting. Moreover, the condition that  $l$  is exponentially large compared to  $Q$  assures that the bound does not explode when different  $\mu\chi$ ’s are combined and gives the desired disjointness.

In fact, our method easily extends to all  $h$  satisfying  $|\hat{h}(m)| \ll e^{-\tau|m|^{\frac{1}{2}+\epsilon}}$ . But we will focus on the analytic case in this paper.

**Notations.** We will write  $m_{\mathbb{T}^d}$  for the Lebesgue probability measure on  $\mathbb{T}^d$ , and  $e(\theta)$  for the function  $e^{2\pi i\theta}$ . The norm  $\|\theta\|$  of  $\theta \in \mathbb{T}^1$  is the distance from  $\theta$  to 0. This norm extends naturally to  $\theta \in \mathbb{R}$ , that is,  $\|\theta\| = \text{dist}(\theta, \mathbb{Z})$ . A basic fact that will be used repeatedly is  $\|\theta\| \ll e(\theta) - 1 \ll \|\theta\|$ .

## 2 Existence of a coboundary

Remark that  $T^n(x, y) = (x + n\alpha, y + H(n, x))$  where

$$H(n, x) = \sum_{l=0}^{n-1} h(x + l\alpha). \tag{2.1}$$

One useful property is:

$$H(n_1n_2, x) = \sum_{l=0}^{n_1-1} H(n_2, x + ln_2\alpha). \tag{2.2}$$

*Remark 2.1* By density of trigonometric polynomials in  $C^0(\mathbb{T}^2)$ , it suffices to check for every  $f(x, y) = e(\xi_1x + \xi_2y)$ ,  $\xi_1, \xi_2 \in \mathbb{Z}$  in order to establish (1.2).

In this section, we make the following reduction.

**Lemma 2.2** *For a continuous map  $h : \mathbb{T}^1 \mapsto \mathbb{T}^1$ , if (1.2) fails, then  $\alpha$  is irrational, and there exist:*

- $s \in \mathbb{N}$ ;
- a pair of distinct primes  $p_1, p_2$ ;
- a measurable map  $g : \mathbb{T}^1 \mapsto \mathbb{T}^1$ ,

such that

$$s(H(p_1, x_0 + p_1x) - H(p_2, x_0 + p_2x)) = g(x + \alpha) - g(x), \text{ for } m_{\mathbb{T}^1}\text{-a.e. } x. \tag{2.3}$$

Here the map  $H$  is defined by (2.1).

Two tools are needed to prove this observation. The first is the Bourgain–Sarnak–Ziegler criterion for Möbius disjointness, which is an extension of Vinogradov’s bilinear method and Vaughan’s identity, and eliminates the need of a saving of degree  $\log^{-C} N$  in the hypothesis.

*Remark 2.3* As noted in [3], the Bourgain-Sarnak-Ziegler criterion is a dynamical interpretation of the bilinear method, which turns the study of type II sums into that of joinings between  $T^{p_1}$  and  $T^{p_2}$  where  $p_1, p_2$  are different primes. In the case of skew products (1.1), Kułaga-Przymus and Lemańczyk studied such joinings in [11] as dynamical systems on  $\mathbb{T}^3$ . A simple observation is that to study the correlation in Proposition 2.4 below, it suffices to study a certain  $\mathbb{T}^2$ -factor of the dynamics on  $\mathbb{T}^3$ , which would deduce Lemma 2.2 from from [11, Theorem 2.5.2] and Furstenberg’s dichotomy (Proposition [6] below). We include the proof of Lemma 2.2 for completeness.

**Proposition 2.4** [3] *Suppose two functions  $F, \nu : \mathbb{N} \mapsto \mathbb{C}$  satisfy  $|F| \leq 1, |\nu| \leq 1$  and  $\nu$  is multiplicative. If for any pair of distinct primes  $p_1, p_2,$*

$$\sum_{n \leq N} F(p_1 n) \overline{F(p_2 n)} = o(N),$$

then

$$\sum_{n \leq N} \nu(n) F(n) = o(N).$$

The second tool we need is the following dynamical dichotomy proved by Furstenberg.

**Proposition 2.5** [6] *Suppose  $\{\Omega_0, T_0, \mu_0\}$  is uniquely ergodic topological dynamical system with  $\mu_0$  being the unique ergodic measure, and  $h : \Omega_0 \mapsto \mathbb{T}^1$  is a continuous function. Let  $T : \Omega_0 \times \mathbb{T}^1 \mapsto \Omega_0 \times \mathbb{T}^1$  be defined by  $T(x, y) = (T_0(x), y + h(x))$ . Then exactly one of the following is true:*

- (1)  *$T$  is ergodic with  $\mu_0 \times m_{\mathbb{T}^1}$  being the unique ergodic measure.*
- (2) *There exists a measurable map  $g : \Omega \mapsto \mathbb{T}^1$  and a non-zero integer  $s$  such that  $sh(x) = g(T_0(x)) - g(x)$  for  $\mu$ -almost every  $x$ .*

Notice that in any case,  $\mu_0 \times m_{\mathbb{T}^1}$  is  $T$ -invariant. And in case (2), for any  $y_0 \in \mathbb{T}^1$ , the closed set  $\{(x, y) : k(y - y_0) = f(x)\}$  is  $T$  invariant. As this set is a measurable  $k$ -fold cover of  $\Omega_0$ , it naturally supports a  $T$ -invariant probability measure that projects to  $\mu_0$  and assigns equal mass to  $k$  distinct points in each fiber along  $y$ -direction.

*Proof of Lemma 2.2* Suppose that (1.2) fails for some  $f$ . The  $\alpha$  must be irrational (see e.g. [13]). By Remark 2.1, we can assume  $f(x, y) = e(\xi_1 x + \xi_2 y)$ . Furthermore,  $\xi_2$  cannot be 0, as otherwise

$$\sum_{n \leq N} \mu(n) f(T^n(x_0, y_0)) = \sum_{n \leq N} \mu(n) e(\xi_1(x_0 + n\alpha)) = e(\xi_1 x_0) \sum_{n \leq N} \mu(n) e(n\xi_1 \alpha),$$

which is uniformly  $o(N)$  for all  $\alpha$  by Davenport’s estimate [4].

By Proposition 2.4 there are distinct primes  $p_1, p_2$  such that

$$\frac{1}{N} \sum_{n \leq N} f(T^{p_1 n})(x_0, y_0) \overline{f(T^{p_2 n})(x_0, y_0)} \rightarrow 0. \tag{2.4}$$

By (2.1), the left hand side is equal to:

$$\begin{aligned}
 & \frac{1}{N} \sum_{n \leq N} e(\xi_1(x_0 + p_1 n \alpha) + \xi_2(y_0 + H(p_1 n, x_0))) \\
 & \quad e(-\xi_1(x_0 + p_2 n \alpha) - \xi_2(y_0 + H(p_2 n, x_0))) \\
 &= \frac{1}{N} \sum_{n \leq N} e(\xi_1(p_1 - p_2)n\alpha + \xi_2(H(p_1 n, x_0) - H(p_2 n, x_0))) \tag{2.5} \\
 &= \frac{1}{N} \sum_{n \leq N} f((p_1 - p_2)n\alpha, (H(p_1 n, x_0) - H(p_2 n, x_0))).
 \end{aligned}$$

Define a new dynamical system of the form (1.1) by

$$\tilde{T}(x, y) = (x + \alpha, y + H(p_1, x_0 + p_1 x) - H(p_2, x_0 + p_2 x)).$$

We claim that

$$\tilde{T}^n(0, 0) = (n\alpha, H(p_1 n, x_0) - H(p_2 n, x_0)) \tag{2.6}$$

for all  $n \geq 0$ . In fact, this is true for  $n = 0$ . Assume for induction that (2.6) holds for  $n$ , then  $\tilde{T}^{n+1}(0, 0)$  equals

$$(n\alpha + \alpha, H(p_1 n, x_0) - H(p_2 n, x_0) + H(p_1, x_0 + p_1 n \alpha) - H(p_2, x_0 + p_2 n \alpha)).$$

Because by (2.2), for all  $p \in \mathbb{N}$ ,

$$\begin{aligned}
 H(pn, x_0) + H(p, x_0 + pn\alpha) &= \sum_{l=0}^{n-1} H(p, x_0 + ln\alpha) + H(p, x_0 + pn\alpha) \\
 &= \sum_{l=0}^n H(p, x_0 + ln\alpha) = H(p(n+1), x_0),
 \end{aligned}$$

(2.6) holds for  $n + 1$  as well.

It follows that, for  $\tilde{f}(x, y) = f((p_1 - p_2)x, y)$ ,

$$(2.5) = \frac{1}{N} \sum_{n \leq N} \tilde{f}(\tilde{T}^n(0, 0)), \tag{2.7}$$

This implies that  $m_{\mathbb{T}^2}$  cannot be the unique ergodic probability measure for  $\tilde{T}$ . Suppose otherwise, then any weak\* limit of a subsequence of  $\{\frac{1}{N} \sum_{n \leq N} \delta_{\tilde{T}^n(0,0)}\}_{N=1}^\infty$ , which is a  $\tilde{T}$ -invariant probability measure, must

equal  $m_{\mathbb{T}^2}$ . Because the space of probability measures on  $\mathbb{T}^2$  is weak\* compact,  $\frac{1}{N} \sum_{n \leq N} \delta_{\tilde{T}^n(0,0)} \rightarrow m_{\mathbb{T}^2}$ . By Birkhoff’s ergodic theorem, (2.5) converges to

$$\begin{aligned} \iint \tilde{f}(x, y) dx dy &= \iint f((p_1 - p_2)x, y) dx dy = \iint f(x, y) dx dy \\ &= \int e(\xi_1 x) dx \cdot \int e(\xi_2 y) dy, \end{aligned}$$

which equals 0 because  $\xi_2 \neq 0$ . This contradicts (2.4).

Since  $\alpha$  is irrational,  $x \mapsto x + \alpha$  is uniquely ergodic on  $\mathbb{T}^1$ . By applying Proposition 2.5 to the map  $\tilde{T}$ , we know that there are  $s \in \mathbb{N}$  and a measurable map  $g : \mathbb{T}^1 \mapsto \mathbb{T}^1$  for which (2.3) holds for  $m_{\mathbb{T}^1}$ -almost every  $x$ .  $\square$

Recall that the degree of a continuous map  $h : \mathbb{T}^1 \mapsto \mathbb{T}^1$  is the image  $d = h_*1 \in \pi_1(\mathbb{T}^1) = \mathbb{Z}$ , where 1 is the identity in  $\pi_1(\mathbb{T}^1)$ .

**Corollary 2.6** *If a Lipschitz continuous map  $h : \mathbb{T}^1 \mapsto \mathbb{T}^1$  makes (1.2) fail, then  $h$  is homotopically trivial.*

The corollary follows from [11, Remark 2.5.7], we again give the proof for completeness.

*Proof* Recall first a few simple facts:

- (1) If  $h$  has degree  $d$ , then so does  $x \mapsto h(x + a)$  for all  $a \in \mathbb{R}$ .
- (2) If  $h$  has degree  $d$ , then  $x \mapsto h(ax)$  has degree  $ad$  for all  $a \in \mathbb{Z}$ .
- (3) If  $h_1$  and  $h_2$  respectively have degrees  $d_1, d_2$ , then the degree of  $h_1 + h_2$  is  $d_1 + d_2$ .

It follows that if  $h$  has degree  $d$ , then the map

$$\begin{aligned} &H(p_1, x_0 + p_1x) - H(p_2, x_0 + p_2x) \\ &= \sum_{l=0}^{p_1-1} h(x_0 + p_1x + l\alpha) - \sum_{l=0}^{p_2-1} h(x_0 + p_2x + l\alpha) \end{aligned}$$

has degree  $(p_1^2 - p_2^2)d$ .

Assume, in order to get a contradiction, that  $d \neq 0$ , then for any two different primes  $p_1, p_2$ ,  $H(p_1, x_0 + p_1x) - H(p_2, x_0 + p_2x)$  is homotopically non-trivial. Since  $h$  is assumed to be Lipschitz, by [6, Lemma 2.2], there is no solution to (2.3). So, by Lemma 2.2, (1.2) must hold. This contradicts the hypothesis and concludes the proof.  $\square$

*Remark 2.7* If  $h$  is as in Corollary 2.6, then it can be realized as a Lipschitz function from  $\mathbb{R}/\mathbb{Z}$  to  $\mathbb{R}$ . In particular, one can talk about its Fourier series,

which pointwise converges to  $h$  because of Lipschitz continuity, i.e., for all  $x \in \mathbb{T}^1$ ,

$$h(x) = \sum_{m \in \mathbb{Z}} \hat{h}(m)e(mx). \tag{2.8}$$

### 3 Eliminating non-resonant frequencies

From now on, we suppose that (1.2) is not true. In addition, assume that  $h$  is analytic, and hence (2.8) holds and there exists  $\tau > 0$  for which

$$|\hat{h}(m)| \ll e^{-\tau|m|}, \tag{3.1}$$

where the implied constant may depend on  $h$  and  $\tau$ .

Recall that  $\alpha$  must be irrational in this case. Let  $\frac{p_k}{q_k}$  be the convergents given by  $\alpha$ 's continued fraction expansion  $[0; a_1, a_2, \dots]$ .

*Remark 3.1* We list below a few fundamental properties of approximation of  $\alpha$  by  $p_k, q_k$  (see [12]):

- (1)  $p_{k+1} = a_k p_k + p_{k-1}, q_{k+1} = a_k q_k + q_{k-1}$  and  $(p_k, q_k) = 1$ ;
- (2)  $\frac{1}{q_{k+1} + q_k} < \|q_k \alpha\| < \frac{1}{q_{k+1}}$ ;
- (3) If for integers  $m$  and  $n, |\alpha - \frac{n}{m}| < \frac{1}{2m^2}$ , then  $\frac{n}{m} = \frac{p_j}{q_j}$  for some  $j$ .

For all  $b_1, b_2 \in \mathbb{N}$ , define

$$M_{b_1, b_2} = \bigcup_{\substack{k: q_{k+1} > e^{\frac{\tau}{2} q_k} \\ q_k \geq b_1}} \{m : q_k \leq |m| < b_2 q_{k+1} \text{ and } q_k |m|\} \tag{3.2}$$

The following lemma is similar to [13, Lemma 4.1].

**Lemma 3.2** *Suppose  $h : \mathbb{T}^1 \mapsto \mathbb{R}$  satisfies (3.1). Then for all  $b_1, b_2 \in \mathbb{N}$ , the series*

$$\sum_{m \notin M_{b_1, b_2} \cup \{0\}} \hat{h}(m) \frac{1}{e(m\alpha) - 1} e(mx)$$

*converges uniformly and hence defines a continuous function  $\phi(x)$ .*

*Proof of Lemma 2.2* Since  $M_{b_1, 1} \subset M_{b_1, b_2}$ , it suffices to prove for  $b_2 = 1$ . If  $m \notin M_{b_1, 1} \cup \{0\}$ , then at least one of the following three cases holds:

- (1)  $|m| < \min_{q_k \geq b_1} q_k$ . Since only finitely many frequencies are involved in this case, they can be ignored without affecting the convergence.



(2)  $q_k \leq |m| < q_{k+1}$ ,  $q_k \nmid |m|$ . In this case, we claim that  $\|m\alpha\| \geq \frac{1}{2|m|}$ . By Remark 3.1,  $\|m\alpha\| \geq \frac{1}{2m}$  unless:  $|m| = lq_j$  and  $\|m\alpha\| = |m\alpha - lp_j|$  for some  $j \leq k$  and  $l \in \mathbb{Z} \setminus \{0\}$ . In the later case, by the assumption on  $m$ ,  $j$  must be less than  $k$ , and we still have

$$\|m\alpha\| = |l| \cdot |q_j\alpha - p_j| = |l| \cdot \|q_j\alpha\| > \frac{|l|}{q_{j+1} + q_j} \geq \frac{1}{q_k + q_{k-1}} > \frac{1}{2q_k} \geq \frac{1}{2m}.$$

It follows that

$$\begin{aligned} & \sum_{\substack{q_k \leq |m| < q_{k+1} \\ q_k \nmid m}} \left| \hat{h}(m) \frac{1}{e(m\alpha) - 1} e(mx) \right| \\ & \ll \sum_{\substack{q_k \leq |m| < q_{k+1} \\ q_k \nmid m}} (e^{-\tau|m|} \cdot |m| \cdot 1) \leq 2 \sum_{m=q_k}^{\infty} e^{-\tau m} m \tag{3.3} \\ & \ll e^{-\frac{\tau}{2}q_k}. \end{aligned}$$

(3)  $q_k \leq |m| < q_{k+1} \leq e^{\frac{\tau}{2}q_k}$  and  $q_k |m$ . In this case,  $|m| = aq_k$  with  $1 \leq a \leq a_k$ . By Remark 3.1,  $\frac{a}{q_{k+1} + q_k} < a\|q_k\alpha\| < \frac{a}{q_{k+1}} \leq \frac{a_k}{q_{k+1}} < \frac{1}{q_k}$ , and thus  $a\|q_k\alpha\| = \|m\alpha\|$ . Therefore, because  $q_{k+1} \leq e^{\frac{\tau}{2}q_k}$

$$\begin{aligned} & \sum_{\substack{q_k \leq |m| < q_{k+1} \\ q_k | m}} \left| \hat{h}(m) \frac{1}{e(m\alpha) - 1} e(mx) \right| \\ & \ll \sum_{\substack{|m|=aq_k \\ 1 \leq a \leq a_k}} (e^{-\tau|m|} \cdot \frac{q_{k+1} + q_k}{a} \cdot 1) \ll 2 \sum_{a=1}^{\infty} e^{-\tau a q_k} e^{\frac{\tau}{2}q_k} \tag{3.4} \\ & \ll e^{-\tau q_k} e^{\frac{\tau}{2}q_k} = e^{-\frac{\tau}{2}q_k}. \end{aligned}$$

Since both (3.3) and (3.4) are convergent when summed over all the  $q_k$ 's. The lemma follows.

In the next reduction, we eliminate the non-resonant frequencies, i.e., those not in  $M_{b_1, b_2}$ .

**Corollary 3.3** *Suppose  $h$  is as in Lemma 3.2. For arbitrary  $b_1, b_2 \in \mathbb{N}$  let  $h_1(x) = \sum_{m \in M_{b_1, b_2} \cup \{0\}} \hat{h}(m)e(mx)$ . Then:*

- (1)  $T$  satisfies (1.2) if and only if  $T_1(x, y) = (x + \alpha, y + h_1(x))$  does;

(2) For distinct primes  $p_1, p_2$ ,  $h$  satisfies (2.3) for a measurable map  $g : \mathbb{T}^1 \mapsto \mathbb{T}^1$  if and only if  $h_1$  satisfies the same condition with respect to another measurable map  $g_2$ .

*Proof* Write  $h(x) = h_1(x) + h_2(x)$ , where  $h_1$  and  $h_2$  are partial sums from the Fourier expansion of  $h$ , respectively over  $M_{b_1, b_2} \cup \{0\}$  and its complement.

Let  $\phi(x)$  be given by Lemma 3.2. Then

$$\begin{aligned} \phi(x + \alpha) - \phi(x) &= \sum_{m \notin M_{b_1, b_2} \cup \{0\}} \hat{h}(m) \frac{1}{e(m\alpha) - 1} (e(mx + m\alpha) - e(mx)) \\ &= \sum_{m \notin M_{b_1, b_2} \cup \{0\}} \hat{h}(m) \frac{1}{e(m\alpha) - 1} (e(m\alpha) - 1)e(mx) \\ &= h_2(x). \end{aligned}$$

Denote  $\Phi(x, y) = (x, y - \phi(x))$ . Then  $\Phi^{-1}(x, y) = (x, y + \phi(x))$ , and

$$\begin{aligned} (\Phi^{-1} \circ T_1 \circ \Phi)(x, y) &= (\Phi^{-1} \circ T_1)(x, y - \phi(x)) \\ &= \Phi^{-1}(x + \alpha, y - \phi(x) + h_1(x)) \\ &= (x + \alpha, y - \phi(x) + h_1(x) + \phi(x + \alpha)) \\ &= (x + \alpha, y + h_1(x) + h_2(x)) = (x + \alpha, y + \phi(x)) \\ &= T(x, y). \end{aligned}$$

Therefore,  $f(T^n(x_0, y_0)) = f_1(T_1^n(x_1, y_1))$ , where  $f_1 = f \circ \Phi^{-1}$  and  $(x_1, y_1) = \Phi(x_0, y_0)$ . This shows claim (1).

As for claim (2), define  $H_1(p, x), H_2(p, x)$  as in (2.1) respectively using  $h_1$  and  $h_2$ . Then for any  $p, H(p, x_0 + px) = H_1(p, x_0 + px) + H_2(p, x_0 + px)$ . However,  $H_2(p, x) = \sum_{l=0}^{p-1} (\phi(x + (l + 1)\alpha) - \phi(x + l\alpha)) = \phi(x + p\alpha) - \phi(x)$ .

Therefore,

$$\begin{aligned} &s(H(p_1, x_0 + p_1x) - H(p_2, x_0 + p_2x)) \\ &\quad - s(H_1(p_1, x_0 + p_1x) - H_1(p_2, x_0 + p_2x)) \\ &= s(H_2(p_1, x_0 + p_1x) - H_2(p_2, x_0 + p_2x)) \\ &= \phi(x_0 + p_1x + p_1\alpha) - \phi(x_0 + p_1x) = g_2(x + \alpha) - g_2(x). \end{aligned}$$

where  $g_2(x) = \phi(x_0 + px)$ . So (2.3) holds for  $H$  and  $g$  if and only if it holds for  $H_1$  and  $g_1$  with  $g_1 = g - g_2$ . □

### 4 Almost linearity of orbit along arithmetic progressions

The aim of this section is to show the orbit point  $T^n(x_0, y_0) \in \mathbb{T}^2$  varies in an almost linear way when  $n$  assumes value in an arithmetic progression of step length  $q_k$ .

**Lemma 4.1** *Suppose an analytic function  $h : \mathbb{T}^1 \mapsto \mathbb{R}$  satisfies (3.1), and there are  $b_1 \geq 2b_2 \geq 1$  such that  $\hat{h}$  is supported on  $M_{b_1, b_2} \cup \{0\}$ . Then the estimate*

$$\left| H(q_k, x) - q_k \hat{h}(0) \right| \ll e^{-\frac{\tau}{4}q_k}$$

holds uniformly for all  $x \in \mathbb{T}^1$  and for all  $k$  satisfying  $q_{k+1} > e^{\frac{\tau}{2}q_k}$ , where the implied constant depends on  $h, \tau$  and  $b_1, b_2$ .

*Proof* Using geometric series, one easily finds out that for all  $n \in \mathbb{N}$ ,

$$\begin{aligned} H(n, x) &= \sum_{l=0}^{n-1} h(x + l\alpha) = \sum_{l=0}^{n-1} \sum_{m \in \mathbb{Z}} \hat{h}(m) e(mx + lm\alpha) \\ &= n\hat{h}(0) + \sum_{m \neq 0} \hat{h}(m) \frac{e(nm\alpha) - 1}{e(m\alpha) - 1} e(mx). \end{aligned} \tag{4.1}$$

Because  $\hat{h}$  is supported on  $M_{b_1, b_2} \cup \{0\}$ ,

$$\begin{aligned} \left| H(q_k, x) - q_k \hat{h}(0) \right| &= \left| \sum_{m \in M_{b_1, b_2}} \hat{h}(m) \frac{e(q_k m \alpha) - 1}{e(m \alpha) - 1} e(mx) \right| \\ &\leq \sum_{\substack{j: q_{j+1} > e^{\frac{\tau}{2}q_j} \\ q_j \geq b_1}} \sum_{\substack{q_j \leq m < b_2 q_{j+1} \\ q_j | m}} \left| \hat{h}(m) \frac{e(q_k m \alpha) - 1}{e(m \alpha) - 1} e(mx) \right| \\ &= \sum_{\substack{j: q_{j+1} > e^{\frac{\tau}{2}q_j} \\ q_j \geq b_1}} \sum_{a=1}^{b_2 a_j} \left| \hat{h}(aq_j) \frac{e(aq_j q_k \alpha) - 1}{e(aq_j \alpha) - 1} e(aq_j x) \right| \\ &=: \sum_{\substack{j: q_{j+1} > e^{\frac{\tau}{2}q_j} \\ q_j \geq b_1}} |H_{j,k}(x)|. \end{aligned} \tag{4.2}$$

We estimate  $|H_{j,k}(x)|$  separately according to  $j < k$ , or  $j \geq k$ .

(1)  $j < k$ . Remark  $a\|q_j\alpha\| \leq aq_{j+1}^{-1} \leq b_2a_jq_{j+1}^{-1} < b_2q_j^{-1} \leq b_2b_1^{-1} \leq \frac{1}{2}$ . This implies  $\|aq_j\alpha\| = a\|q_j\alpha\| \geq a(q_{j+1} + q_j)^{-1}$ .

On the other hand,  $\|aq_jq_k\alpha\| \leq aq_j\|q_k\alpha\| \leq aq_jq_{k+1}^{-1}$ . By these estimates and the analyticity assumption (3.1),

$$|H_{j,k}(x)| \ll \sum_{a=1}^{b_2a_j} e^{-\tau aq_j} \cdot \frac{aq_jq_{k+1}^{-1}}{a(q_{j+1} + q_j)^{-1}} \ll \sum_{a=1}^{\infty} e^{-\tau aq_j} q_jq_{j+1}q_{k+1}^{-1} \ll q_{j+1}q_{k+1}^{-1}.$$

Thus by the assumption  $q_{k+1} > e^{\frac{\tau}{2}q_k}$ ,

$$\sum_{\substack{j < k: q_{j+1} > e^{\frac{\tau}{2}q_j} \\ q_j \geq b_1}} |H_{j,k}(x)| \ll \sum_{j=1}^{k-1} q_{j+1}q_{k+1}^{-1} \ll q_kq_{k+1}^{-1} < q_k e^{-\frac{\tau}{2}q_k} \ll e^{-\frac{\tau}{4}q_k}. \tag{4.3}$$

Where we used that  $\sum_{j=1}^J q_j \leq 4q_J$ , which is due to the basic fact that  $q_{j+2} \geq q_{j+1} + q_j > 2q_j$  for all  $j$ .

(2)  $j \geq k$ . In this case, we use the trivial estimate that  $\left| \frac{e(aq_jq_k\alpha) - 1}{e(aq_j\alpha) - 1} \right| \leq q_k$ . This implies:

$$|H_{j,k}(x)| \ll \sum_{a=1}^{b_2a_j} e^{-\frac{\tau}{2}aq_j} \cdot q_k \ll e^{-\frac{\tau}{2}q_j} q_k,$$

and thus,

$$\sum_{\substack{j \geq k: q_{j+1} > e^{\frac{\tau}{2}q_j} \\ q_j \geq b_1}} |H_{j,k}(x)| \ll \sum_{q=q_k}^{\infty} e^{-\frac{\tau}{2}q_k} q_k \ll e^{-\frac{\tau}{2}q_k} q_k \ll e^{-\frac{\tau}{4}q_k}. \tag{4.4}$$

The lemma is established by combining (4.2), (4.3) and (4.4). □

### 5 Almost periodicity of orbit

We know from the previous lemma that when  $n$  moves along arithmetic progressions of step length  $q_k$ , the  $y$ -coordinate of  $T^n(x, y)$  is close to a linear sequence in  $\mathbb{T}^1$  of slope  $\hat{h}(0)$ , which is the average value of the function  $h$ . Our next goal is to control this average. Unlike the degree  $d$ ,  $\hat{h}(0)$  cannot be fully eliminated. However, it exhibits similar Diophantine patterns to those of  $\alpha$ .

**Lemma 5.1** *Suppose  $h$  is as in Lemma 4.1, and (1.2) is not true, then there exists  $\eta \in (0, \frac{\tau}{4})$ ,  $S \in \mathbb{N}$ , such that for any given  $\delta > 0$ :*

For all sufficiently large  $k$  with  $q_{k+1} > e^{\frac{\tau}{2}q_k}$  and any positive integer  $a \leq e^{\eta q_k}$ ,

$$\|aSq_k\hat{h}(0)\| < \delta.$$

*Proof* By Lemma 2.2, there exists  $s \in \mathbb{N}$  and distinct primes  $p_1 < p_2$ , and a measurable map  $g : \mathbb{T}^1 \mapsto \mathbb{T}^1$  such that (2.3) holds.

By Lemma 3.3, after replacing  $h$  by  $\sum_{m \in M_{2p_2,1} \cup \{0\}} \hat{h}(m)e(mx)$ , (2.3) still holds for the same  $p_1, p_2$  and a different  $g$ . In addition,  $\hat{h}(0)$  does not change by this modification. So we can assume  $\hat{h}$  is supported on  $M_{p_2,1} \cup \{0\}$ .

By Luzin’s theorem, there is a compact set  $\Omega \subset \mathbb{T}^1$  of Lebesgue measure at least 0.9 on which  $g$  is continuous. There is  $\epsilon$  such that if  $\|x - x'\| < \epsilon$  and  $x, x' \in \Omega$  then  $\|g(x') - g(x)\| < \frac{\delta}{2}$ . Without loss of generality, we will assume  $\epsilon < \frac{\delta}{2}$ .

For  $\eta < \frac{\tau}{4}$ , under the assumptions on  $k$  and  $a$ ,  $a < e^{-\frac{\tau}{4}q_k}q_{k+1}$ , and thus,

$$\|aq_k\alpha\| < a\|q_k\alpha\| < e^{-\frac{\tau}{4}q_k}q_{k+1} \cdot q_{k+1}^{-1} = e^{-\frac{\tau}{4}q_k},$$

and is less than  $\epsilon$  when  $k$  is sufficiently large. Consider the pair of points  $x$  and  $x + aq_k\alpha$ , they both belong to  $\Omega$  if  $x$  lies in  $\Omega' = \Omega \cap (\Omega - aq_k\alpha)$ . As  $m_{\mathbb{T}^1}(\Omega - aq_k\alpha) = m_{\mathbb{T}^1}(\Omega) \geq 0.9$ , we know  $\Omega'$  is non-empty.

Write  $\psi(x)$  for  $g(x + \alpha) - g(x)$ , which by (2.3) equals  $s(H(p_1, x_0 + p_1x) - H(p_2, x_0 + p_2x))$  for almost every  $x \in \Omega'$ .

Fix hereafter such a generic point  $x \in \Omega'$ , then

$$\|g(x + aq_k\alpha) - g(x)\| < \frac{\delta}{2}. \tag{5.1}$$

We apply (4.1) to get the Fourier expansion of  $\psi$ :

$$\psi(x) = s \sum_m \hat{h}(m) \left( \sum_{l=0}^{p_1} e(lm\alpha + mx_0 + p_1mx) - \sum_{l=0}^{p_2} e(lm\alpha + mx_0 + p_2mx) \right).$$

In other words,

$$\hat{\psi}(m) = s \left( \hat{h} \left( \frac{m}{p_1} \right) \sum_{l=0}^{p_1} e(lm\alpha + mx_0) - \hat{h} \left( \frac{m}{p_2} \right) \sum_{l=0}^{p_2} e(lm\alpha + mx_0) \right),$$

with the convention that  $\hat{h}(\theta) = 0$  if  $\theta \notin \mathbb{Z}$ .

This implies several properties of  $\psi$ . First, as  $h$  satisfies (3.1),

$$|\hat{\psi}(m)| \leq s \left( p_1 \left| \hat{h} \left( \frac{m}{p_1} \right) \right| - p_1 \left| \hat{h} \left( \frac{m}{p_1} \right) \right| \right) \ll e^{-\frac{\tau}{p_2}|m|}, \tag{5.2}$$

with an implied constant depending on  $h, s$  and  $p_2$ .

Second,

$$\text{supp } \hat{\psi} \subset p_1 M_{2p_2,1} \cup p_2 M_{2p_2,1} \subset M_{2p_2,p_1} \cup M_{2p_2,p_2} = M_{2p_2,p_2}. \tag{5.3}$$

Furthermore,

$$\hat{\psi}(0) = s(p_1 - p_2)\hat{h}(0). \tag{5.4}$$

Construct  $\Psi(n, x)$  as in (2.1) with  $h$  replaced by  $\psi$ , i.e.  $\Psi(n, x) = \sum_{l=1}^{n-1} \psi(x + l\alpha)$ .

Thanks to (5.2) and (5.3), we can apply Lemma 4.1 to  $\Psi$  to assert that, uniformly for all  $x \in \mathbb{T}^1$ ,

$$|\Psi(q_k, x) - q_k \hat{\psi}(0)| \ll e^{-\frac{\tau}{4p_2}q_k}. \tag{5.5}$$

On the other hand, notice that

$$g(x + aq_k\alpha) - g(x) = \Psi(aq_k, x) = \sum_{l=0}^{a-1} \Psi(q_k, x + lq_k\alpha). \tag{5.6}$$

From (5.1), (5.5) and (5.6), we deduce

$$\begin{aligned} & \|aq_k \hat{\psi}(0)\| \\ & \leq \left\| \sum_{l=0}^{a-1} \Psi(q_k, x + lq_k\alpha) - aq_k \hat{\psi}(0) \right\| + \left\| \sum_{l=0}^{a-1} \Psi(q_k, x + lq_k\alpha) \right\| \\ & \leq \sum_{l=0}^{a-1} |\Psi(q_k, x + lq_k\alpha) - aq_k \hat{\psi}(0)| + \|g(x + aq_k\alpha) - g(x)\| \\ & < Cae^{-\frac{\tau}{4p_2}q_k} + \frac{\delta}{2}, \end{aligned}$$

where  $C$  is a constant that depend only on  $h, s$  and  $p_2$ .

Choose  $\eta < \frac{\tau}{8p_2}$ . Then  $\|aq_k \hat{\psi}(0)\| < C \cdot e^{\frac{\tau}{8p_2}q_k} \cdot e^{-\frac{\tau}{4p_2}q_k} + \frac{\delta}{2}$  is less than  $\delta$  for sufficiently large  $k$ .

However, by (5.4),  $\|aq_k \hat{\psi}(0)\| = \|aq_k s(p_1 - p_2)\hat{h}(0)\|$ . So the lemma is true with  $S = s(p_1 - p_2)$ . □

This, together with Lemma 4.1, shows that an orbit of  $T$  is, within intervals of length up to  $Sq_k e^{\eta q_k}$ , almost periodic with period  $Sq_k$ .

**Corollary 5.2** *Let  $\eta$  and  $S$  be as in Lemma 5.1. Then given any  $\delta > 0$ , for all sufficiently large  $k$  with  $q_{k+1} > e^{\frac{\tau}{2}q_k}$  and any positive integer  $a \leq e^{\eta q_k}$ ,*

$$d(T^{aS q_k}(x, y), (x, y)) < \delta, \quad \forall (x, y) \in \mathbb{T}^2.$$

*Proof*  $T^{aS q_k}(x, y) - (x, y) = (aS q_k \alpha, H(aS q_k, x))$ . So it suffices to bound both  $\|aS q_k \alpha\|$  and  $\|H(aS q_k, x)\|$  by  $\frac{\delta}{2}$ .

First,  $\|aS q_k \alpha\| \leq aS \|q_k \alpha\| < e^{\frac{\tau}{4}q_k} S q_{k+1}^{-1} < e^{\frac{\tau}{4}q_k} S e^{-\frac{\tau}{2}q_k} = S e^{-\frac{\tau}{4}q_k}$ . As  $S$  is independent of  $k$ , this is less than  $\frac{\delta}{2}$  when  $q_k$  is large enough.

On the other hand, by Lemma 4.1,

$$\begin{aligned} & \|H(aS q_k, x)\| \\ &= \left\| \sum_{l=1}^{aS-1} H(q_k, x + lq_k) \right\| = \left\| \sum_{l=1}^{aS-1} (H(q_k, x + lq_k) - q_k \hat{h}(0)) + aS q_k \hat{h}(0) \right\| \\ &\leq \sum_{l=1}^{aS-1} |H(q_k, x + lq_k) - q_k \hat{h}(0)| + \|aS q_k \hat{h}(0)\| \\ &\leq C a S e^{-\frac{\tau}{4}q_k} + \|aS q_k \hat{h}(0)\|, \end{aligned}$$

where  $C$  is a constant independent of  $k$ . By Lemma 5.1,  $\|aS q_k \hat{h}(0)\| < \frac{\delta}{4}$  when  $k$  is sufficiently large. As for the first term, as  $a < e^{-\eta q_k}$ ,  $\eta < \frac{\tau}{4}$  and  $C, S$  do not depend on  $k$ ,  $C a S e^{-\frac{\tau}{4}q_k} \ll e^{-(\frac{\tau}{4}-\eta)q_k}$  and also becomes less than  $\frac{\delta}{4}$  for large  $k$ . Therefore  $\|H(aS q_k, x)\| < \frac{\delta}{2}$ . This completes the proof.  $\square$

### 6 Averages in short intervals

We continue to assume that (3.1) is true but (1.2) is not, and try to obtain a contradiction.

Disjointness from  $\mu(n)$  is known for periodic sequences. And in light of the previous section, we know  $T^n(x_0, y_0)$  is not far from a periodic sequence. In this part we show that the error obtained there does not accumulate to destroy the disjointness.

To begin with, remark that we can always assume there are infinitely many  $k$ 's with  $q_{k+1} > e^{\frac{\tau}{2}q_k}$ . Otherwise, Corollary 3.3 allows to assume  $\hat{h}$  is supported at  $\{0\}$ . Then  $T(x, y) = (x + \alpha, y + \hat{h}(0))$  is a rotation of  $\mathbb{T}^2$ , for which (1.2) is known to hold by Davenport's theorem [4].

From now on, we fix a function  $f(x, y) = e(\xi_1 x + \xi_2 y)$ , which is allowed by Remark 2.1. It suffices to show for all  $\delta > 0$  that, for sufficiently large  $N$ ,

$$\mathbb{E}_{n < N} f(T^n(x_0, y_0))\mu(n) \ll \delta, \tag{6.1}$$

with an implied constant that is allowed to depend on  $h$  and  $f$ .

Let  $\eta$  and  $S$  are given by Lemma 5.2. Suppose  $k$  is sufficiently large and satisfies  $q_{k+1} > e^{\frac{\tau}{4}q_k}$ . And set  $A = \lfloor e^{\eta q_k} \rfloor$ . Furthermore, let  $N_0 < \frac{N}{2}$  be an integer.

The average (1.2) decomposes into averages in shorter intervals of length  $ASq_k$ :

$$\begin{aligned} & \mathbb{E}_{n < N} \mu(n)f(T^n(x_0, y_0)) \\ &= \mathbb{E}_{L=N_0}^{N-1} \mu(n)f(T^n(x_0, y_0)) + O\left(\frac{N_0}{N}\right) \\ &= \mathbb{E}_{L=N_0}^{N-1} \mathbb{E}_{n=L}^{L+ASq_k-1} \mu(n)f(T^n(x_0, y_0)) + O\left(\frac{ASq_k}{N}\right) + O\left(\frac{N_0}{N}\right) \\ &= \mathbb{E}_{L=N_0}^{N-1} \mathbb{E}_{l=L}^{L+Sq_k-1} \mathbb{E}_{a=0}^{A-1} \mu(l+aSq_k)f(T^{l+aSq_k}(x_0, y_0)) + O\left(\frac{ASq_k}{N} + \frac{N_0}{N}\right). \end{aligned}$$

As  $a \leq A < e^{\eta q_k}$ , by Lemma 5.2, for sufficiently large  $k$ ,

$$|f(T^{l+aSq_k}(x_0, y_0)) - f(T^l(x_0, y_0))| < \|f\|_{C^1} \delta.$$

Therefore,

$$\begin{aligned} & \mathbb{E}_{n < N} \mu(n)f(T^n(x_0, y_0)) \\ &= \mathbb{E}_{L=N_0}^{N-1} \mathbb{E}_{l=L}^{L+Sq_k-1} \mathbb{E}_{a=0}^{A-1} \mu(l+aSq_k)f(T^l(x_0, y_0)) + O(\delta) + O\left(\frac{ASq_k}{N} + \frac{N_0}{N}\right). \end{aligned}$$

For each  $L$ , construct a function  $F_L : \mathbb{N} \mapsto \mathbb{C}$  by  $F_L(n) = f(T^l(x_0, y_0))$  where  $l$  is the unique integer in  $[L, L + Sq_k)$  such that  $l \equiv n \pmod{Sq_k}$ . Then  $F_L$  is periodic with period  $Sq_k$  and  $|F_L| = 1$ , and

$$\begin{aligned} & \mathbb{E}_{n < N} \mu(n)f(T^n(x_0, y_0)) \\ &= \mathbb{E}_{L=N_0}^{N-1} \mathbb{E}_{n=L}^{L+ASq_k-1} \mu(n)F_L(n) + O\left(\delta + \frac{ASq_k}{N} + \frac{N_0}{N}\right). \end{aligned} \tag{6.2}$$



We next decompose  $\mathbb{E}_{n=L}^{L+Asq_k-1} \mu(n)F_L(n)$  into short averages of Dirichlet characters.

**Lemma 6.1** *For all  $L, Q, A \in \mathbb{N}$  and any periodic function  $F : \mathbb{N} \mapsto \mathbb{C}$  of period  $Q$  with  $|F| \leq 1$ ,*

$$\left| \mathbb{E}_{L \leq n < L+AQ} \mu(n)F(n) \right|^2 \leq Q \mathbb{E}_{\substack{d|Q \\ \chi \bmod^* \frac{Q}{d}}} \left| \mathbb{E}_{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d}} \mu(r)\chi(r) \right|^2,$$

where the first average on the right hand side is taken over all pairs  $(d, \chi)$  such that  $d|Q$  and  $\chi$  is a Dirichlet character of conductor  $\frac{Q}{d}$ .

*Proof* We group  $n$  according to  $(n, Q)$ :

$$\begin{aligned} & \sum_{L \leq n < L+AQ} \mu(n)F(n) \\ &= \sum_{d|Q} \sum_{\substack{L \leq n < L+AQ \\ (n, Q)=d}} \mu(n)F(n) = \sum_{d|Q} \sum_{\substack{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d} \\ (r, \frac{Q}{d})=1}} \mu(rd)F(rd) \\ &= \sum_{d|Q} \sum_{\substack{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d} \\ (r, \frac{Q}{d})=(r, d)=1}} \mu(rd)F(rd) = \sum_{d|Q} \mu(d) \sum_{\substack{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d} \\ (r, \frac{Q}{d})=(r, d)=1}} \mu(r)F(rd). \end{aligned} \tag{6.3}$$

Here the requirement that  $(r, d) = 1$  is because otherwise  $n = rd$  is not square-free and  $\mu(rd) = 0$ .

Identify  $\{r : \frac{L}{d} \leq r < \frac{L}{d} + \frac{Q}{d}, (r, \frac{Q}{d}) = 1\}$  with the finite abelian group  $(\mathbb{Z}/(\frac{Q}{d})\mathbb{Z})^\times$ . The Dirichlet characters of conductor  $\frac{Q}{d}$  form an orthonormal basis of the  $l^2$ -space on this group with respect to the uniform probability measure. Hence  $F(rd)\mathbf{1}_{(r, d)=1}$  can be decomposed as  $\sum_{\chi \bmod^* \frac{Q}{d}} w_{F, \chi} \chi$  on this group. Then,

$$\sum_{\chi \bmod^* \frac{Q}{d}} |w_{F, \chi}|^2 = \|F(rd)\mathbf{1}_{(r, d)=1}\|_{l^2}^2 \leq \|F(rd)\|_{l^\infty} \leq 1. \tag{6.4}$$

It follows from this and (6.3) that,

$$\begin{aligned} \left| \mathbb{E}_{L \leq n < L + A Q} \mu(n) F(n) \right|^2 &= \frac{1}{(A Q)^2} \left| \sum_{L \leq n < L + A Q} \mu(n) F(n) \right|^2 \\ &\leq \frac{1}{(A Q)^2} \left| \sum_{d|Q} \sum_{\chi \bmod^* \frac{Q}{d}} w_{F, \chi} \sum_{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d}} \mu(r) \chi(r) \right|^2 \\ &= \frac{1}{(A Q)^2} \left| \sum_{d|Q} \sum_{\chi \bmod^* \frac{Q}{d}} \frac{A Q}{d} \omega_{F, \chi} \mathbb{E}_{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d}} \mu(r) \chi(r) \right|^2 \\ &= \left| \sum_{d|Q} \sum_{\chi \bmod^* \frac{Q}{d}} \frac{\omega_{F, \chi}}{d} \mathbb{E}_{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d}} \mu(r) \chi(r) \right|^2. \end{aligned}$$

We apply Cauchy-Schwarz inequality and get

$$\begin{aligned} &\left| \mathbb{E}_{L \leq n < L + A Q} \mu(n) F(n) \right|^2 \\ &\leq \left( \sum_{d|Q} \sum_{\chi \bmod^* \frac{Q}{d}} \frac{|\omega_{F, \chi}|^2}{d^2} \right) \left( \sum_{d|Q} \sum_{\chi \bmod^* \frac{Q}{d}} \left| \mathbb{E}_{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d}} \mu(r) \chi(r) \right|^2 \right) \\ &\stackrel{(6.4)}{\leq} \left( \sum_{d|Q} \frac{1}{d^2} \right) \left( \sum_{d|Q} \sum_{\chi \bmod^* \frac{Q}{d}} \left| \mathbb{E}_{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d}} \mu(r) \chi(r) \right|^2 \right) \\ &\ll Q \sum_{\chi \bmod^* \frac{Q}{d}} \mathbb{E}_{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d}} \left| \mathbb{E}_{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d}} \mu(r) \chi(r) \right|^2 \end{aligned}$$

where the last inequality is because  $\sum_{d=1}^{\infty} \frac{1}{d^2}$  converges and there are  $Q$  dirichlet characters  $\chi$  of modulus  $Q$ .

In order to control  $\left| \mathbb{E}_{\frac{L}{d} \leq r < \frac{L}{d} + A \frac{Q}{d}} \mu(r) \chi(r) \right|$ , we make use of a recent theorem from [15] that allows to bound averages of a non-pretentious multiplicative function in random short intervals. For this purpose, we recall the definition of pretentiousness first.

For multiplicative functions  $\nu, \nu' : \mathbb{N} \mapsto \mathbb{C}$  whose absolute values are bounded by 1, define

$$\mathbb{D}(\nu, \nu', X) = \left( \sum_{p \leq X} \frac{1 - \operatorname{Re}(\nu(p)\overline{\nu'(p)})}{p} \right)^{\frac{1}{2}}.$$

The function below measures how closely  $\nu$  pretends to be  $n^{it}$ :

$$M(\nu, X) = \inf_{|t| \leq X} \mathbb{D}(\nu, n^{it}, X)^2.$$

One direction of Halász’s Theorem [9] asserts that if  $\sum_{n \leq X} \nu(n) = o(X)$ , then for all  $t \in \mathbb{R}$ ,

$$\lim_{X \rightarrow \infty} D(\nu, n^{it}, X) = \infty.$$

It is well known that (see e.g. [10, (5.80)]) for a Dirichlet character of conductor  $q$  and any  $C > 0$ ,

$$\sum_{n \leq X} \mu(n)\chi(n) \ll_C q^{\frac{1}{2}}(\log X)^{-C}, \tag{6.5}$$

we see that  $\lim_{X \rightarrow \infty} D(\mu\chi, 1, X) = \infty$  for any given  $\chi$ .

It follows from [15, Lemma C.1], which is based on an argument of Granville and Soundararajan, that

$$\inf_{|t| \leq X} \mathbb{D}(\mu, \bar{\chi}n^{it}, X) \geq \frac{1}{4} \min(\sqrt{\log \log X}, D(\mu, \bar{\chi}, X)) + O(1),$$

or equivalently,

$$\inf_{|t| \leq X} \mathbb{D}(\mu\chi, n^{it}, X) \geq \frac{1}{4} \min(\sqrt{\log \log X}, D(\mu\chi, 1, X)) + O(1).$$

It follows that:

$$\lim_{X \rightarrow \infty} M(\mu\chi, X) = \infty. \tag{6.6}$$

The following proposition, due to Matomäki, Radziwiłł and Tao, is the key input in our method and will be applied to  $\nu(n) = \mu(n)\chi(n)$ .

**Proposition 6.2** [15, Theorem A.1] *Let  $v$  be a multiplicative function with  $|v| \leq 1$  and  $X \geq l \geq 10$ . Then*

$$\mathbb{E}_{X \leq L < 2X} \left| \mathbb{E}_{L \leq n < L+l} v(n) \right|^2 \ll e^{-M(v,X)} M(v, X) + (\log X)^{-\frac{1}{50}} + \left( \frac{\log \log l}{\log l} \right)^2.$$

*Proof of Theorem 1.1* Let  $J = \lceil \log_2 \frac{1}{\delta} \rceil$ . We choose  $N_0 = \lfloor 2^{-J} N \rfloor$ , then  $\frac{N_0}{N} < \delta$ . Cut  $[N_0, N]$  into  $J$  dyadic intervals:  $[2^{-j} N, 2^{-j+1} N)$  for  $j = 1, \dots, J$ .

For each pair  $(d, \chi)$  and every  $j$ , by Proposition 6.2,

$$\begin{aligned} & \left| \mathbb{E}_{2^{-j} N \leq L < 2^{-j+1} N} \left| \mathbb{E}_{\frac{L}{d} \leq r < \frac{L}{d} + \frac{ASq_k}{d}} \mu(r) \chi(r) \right| \right|^2 \\ & \ll \rho \left( \frac{2^{-j+1} N}{d} \right) + \left( \frac{\log \log \frac{ASq_k}{d}}{\log \frac{ASq_k}{d}} \right)^2, \end{aligned} \tag{6.7}$$

where  $\rho(X) = \rho_\chi(X) := e^{-M(\mu\chi, X)} M(\mu\chi, X) + (\log X)^{-\frac{1}{50}}$  is a positive function that converges to 0 as  $X \rightarrow \infty$ , thanks to (6.6). However, we may make  $\rho$  independent of  $\chi$  (but dependent of  $S$  and  $q_k$ ), by taking  $\rho(X) = \rho_{Sq_k}(X) := \max_{\chi \bmod Sq_k} \rho_\chi(X)$ . Furthermore, by replacing  $\rho_{Sq_k}(X)$  with  $\sup_{X' > X} \rho_{Sq_k}(X')$  we may also assume  $\rho_{Sq_k}$  is decreasing.

Recall that  $S$  and  $\eta$  are fixed constants determined by the dynamical system  $T$ . Notice  $d \leq Sq_k, 2^{-j+1} \geq \delta$ , and  $A \leq e^{\eta q_k}$ . Hence,

$$(6.7) \ll \rho_{Sq_k} \left( \frac{\delta N}{Sq_k} \right) + \left( \frac{\log \log A}{\log A} \right)^2 \ll \rho_{Sq_k} \left( \frac{\delta N}{Sq_k} \right) + q_k^{-2} \log^2 q_k.$$

After combining the dyadic intervals, we have

$$\mathbb{E}_{L=N_0}^{N-1} \left| \mathbb{E}_{\frac{L}{d} \leq r < \frac{L}{d} + \frac{ASq_k}{d}} \mu(r) \chi(r) \right|^2 \ll \rho_{Sq_k} \left( \frac{\delta N}{Sq_k} \right) + q_k^{-2} \log^2 q_k. \tag{6.8}$$

It now follows that

$$\begin{aligned} & \mathbb{E}_{n < N} \mu(n) f(T^n(x_0, y_0)) \\ & \stackrel{(6.2)}{=} \mathbb{E}_{L=N_0}^{N-1} \mathbb{E}_{n=L}^{L+ASq_k-1} \mu(n) F_L(n) + O \left( \delta + \frac{ASq_k}{N} + \frac{N_0}{N} \right) \end{aligned}$$

$$\begin{aligned}
 &\stackrel{\text{(Cauchy-Schwarz)}}{\ll} \left( \mathbb{E}_{L=N_0}^{N-1} \left| \mathbb{E}_{n=L}^{L+ASq_k-1} \mu(n) F_L(n) \right|^2 \right)^{\frac{1}{2}} + \delta + \frac{ASq_k}{N} \\
 &\stackrel{\text{(Lemma 6.1)}}{\ll} \left( Sq_k \mathbb{E}_{L=N_0}^{N-1} \left| \mathbb{E}_{\substack{n=L \\ \chi \bmod^* \frac{Sq_k}{d}}}^{L+ASq_k-1} \mu(n) \chi(n) \right|^2 \right)^{\frac{1}{2}} + \delta + \frac{ASq_k}{N} \\
 &\stackrel{\text{(6.8)}}{\ll} \left( Sq_k \cdot \rho_{Sq_k} \left( \frac{\delta N}{Sq_k} \right) + Sq_k^{-1} \log^2 q_k \right)^{\frac{1}{2}} + \delta + \frac{e^{\eta q_k} Sq_k}{N}. \tag{6.9}
 \end{aligned}$$

Once  $\delta$  is fixed, for sufficiently large  $q_k$ ,

$$Sq_k^{-1} \log^2 q_k \ll \delta^2.$$

Fix such a  $q_k$  that, in addition, verifies  $q_{k+1} > e^{\frac{\eta}{2} q_k}$ , then

$$\mathbb{E}_{n < N} \mu(n) f(T^n(x_0, y_0)) \ll \left( Sq_k \cdot \rho_{Sq_k} \left( \frac{\delta N}{Sq_k} \right) + \delta^2 \right)^{\frac{1}{2}} + \delta + \frac{e^{\eta q_k} Sq_k}{N}.$$

Because  $\delta, \eta, S$  and  $q_k$  are now all fixed and  $\rho_{Sq_k}$  is a function that decays to 0, for sufficiently large  $N$ ,  $Sq_k \cdot \rho_{Sq_k} \left( \frac{\delta N}{Sq_k} \right) \ll \delta^2$  and  $\frac{e^{\eta q_k} Sq_k}{N} \ll \delta$ . In consequence, (6.1) is verified for  $N$  large enough. As  $\delta$  is arbitrary, (1.2) holds, contradicting the hypothesis so far. This completes the proof of Theorem 1.1. □

**Acknowledgements** I am deeply grateful to Peter Sarnak for encouragement and insightful discussions. Thanks are also due to Joanna Kułaga-Przymus and Mariusz Lemańczyk for valuable comments on an earlier draft. This work was carried out during a research stay as a member at the Institute for Advanced Study, and I would like to thank the IAS for its support. This work was also supported by NSF Grants DMS-1451247 and DMS-1501295.

**References**

1. Bourgain, J.: On the correlation of the Moebius function with rank-one systems. *J. Anal. Math.* **120**, 105–130 (2013)
2. Bourgain, J.: Möbius–Walsh correlation bounds and an estimate of Mauduit and Rivat. *J. Anal. Math.* **119**, 147–163 (2013)
3. Bourgain, J., Sarnak, P., Ziegler, T.: Disjointness of Möbius from horocycle flows, From Fourier analysis and number theory to Radon transforms and geometry, *Dev. Math.*, vol 28, Springer, New York, pp. 67–83 (2013)
4. Davenport, H.: On some infinite series involving arithmetical functions II. *Q. J. Math.* **8**, 313–320 (1937)

5. El Abdalaoui, E.H., Lemańczyk, M.: On spectral disjointness of powers for rank-one transformations and Möbius orthogonality. *J. Funct. Anal.* **266**(1), 284–317 (2014)
6. Furstenberg, H.: Strict ergodicity and transformation of the torus. *Am. J. Math.* **83**, 573–601 (1961)
7. Green, B.: On (not) computing the Möbius function using bounded depth circuits. *Combin. Probab. Comput.* **21**(6), 942–951 (2012)
8. Green, B., Tao, T.: The Möbius function is strongly orthogonal to nilsequences. *Ann. Math.* (2) **175**(2), 541–566 (2012)
9. Halász, G.: Über die Mittelwerte multiplikativer zahlentheoretischer Funktionen. *Acta Math. Acad. Sci. Hungar.* **19**, 365–403 (1968). (German)
10. Iwaniec, H., Kowalski, E.: *Analytic Number Theory*. American Mathematical Society Colloquium Publications, vol. 53. American Mathematical Society, Providence, RI (2004)
11. Kułaga-Przymus, J., Lemańczyk, M.: The Möbius function and continuous extensions of rotations. *Monatsh. Math.* **178**(4), 553–582 (2015)
12. Khinchin, A.Y.: *Continued Fractions*. Translated from the third (1961) Russian edition, With a preface by B. V. Gnedenko; Reprint of the: translation, p. 1997. Dover Publications Inc, Mineola, NY (1964)
13. Liu, J., Sarnak, P.: The Möbius function and distal flows. *Duke Math. J.* **164**(7), 1353–1399 (2015)
14. Martin, B., Mauduit, C., Rivat, J.: Théorème des nombres premiers pour les fonctions digitales. *Acta Arith.* **165**(1), 11–45 (2014). (French)
15. Matomäki, K., Radziwiłł, M., Tao, T.: An averaged form of Chowla’s conjecture. *Algebra Number Theory* **9**(9), 2167–2196 (2015)
16. Mauduit, C., Rivat, J.: Sur un problème de Gelfond: la somme des chiffres des nombres premiers. *Ann. Math.* (2) **171**(3), 1591–1646 (2010). (French, with English and French summaries)
17. Peckner, R.: Möbius disjointness for homogeneous dynamics. Preprint (2015). [arXiv:1506.07778](https://arxiv.org/abs/1506.07778)
18. Sarnak, P.: Three lectures on the Möbius function randomness and dynamics. Lecture notes. IAS (2012). <https://publications.ias.edu/node/512>
19. Vaughan, R.-C.: Sommes trigonométriques sur les nombres premiers. *C. R. Acad. Sci. Paris Sér. A-B* **285**(16), A981–A983 (1977). (French, with English summary)