MÖBIUS DISJOINTNESS FOR SKEW PRODUCTS ON THE HEISENBERG NILMANIFOLD

MATTHEW LITMAN AND ZHIREN WANG

Abstract. We prove that the Möbius function is disjoint to all Lipschitz continuous skew product dynamical systems on the 3-dimensional Heisenberg nilmanifold over a minimal rotation of the 2-dimensional torus.

1. Introduction

1.1. **Setting and statement.** The Möbius function $\mu : \mathbb{N} \to \{-1, 0, 1\}$ is defined as follows: $\mu(n) = (-1)^k$ if n is the product of k distinct primes, and $\mu(n) = 0$ otherwise. Sarnak's Möbius disjointness conjecture states that $\mu(n)$ is highly random, in the sense that it is orthogonal to all continuous observables from zero-entropy topological dynamical systems. In this article, we deal with a special case of this conjecture, namely Lipschitz continuous skew product maps on the 3-dimensional Heisenberg nilmanifold.

The Heisenberg group is

$$G = \{(x, y, z) : x, y, z \in \mathbb{R}\} \cong \mathbb{R}^3 \tag{1.1}$$

equipped with the group rule

$$(x, y, z)(x', y', z') = (x + x', y + y', z + z' + (xy' - x'y)). \tag{1.2}$$

Set $\Gamma = G(\mathbb{Z}) = \{(x, y, z) \in G : x, y, z \in \mathbb{Z}\}$ and $X = G/\Gamma$. Then X is a compact nilmanifold and its maximal torus factor is $\mathbb{T}^2 = \mathbb{R}^2/\mathbb{Z}^2$, parametrized by the x and y coordinates. X is a principal $\mathbb{T}^1$-bundle over $\mathbb{T}^2$. G acts on X by left translation.

For $\alpha, \beta \in \mathbb{R}$ and a continuous function $h : \mathbb{T}^2 \to \mathbb{T}^1$, define $T : X \to X$ by

$$\mathbf{x} \mapsto (\alpha, \beta, \widetilde{h}(x, y))\mathbf{x}, \tag{1.3}$$

Date: July 21, 2017.
This paper was the outcome of an undergraduate research project sponsored by the Eberly College of Science at Penn State University during the 2016-2017 academic year. M.L. thanks the ECoS for its support. Z.W., the faculty mentor of the project, was supported by the NSF grant DMS-1501295.

where $\mathbf{x} = (x, y, z)\Gamma$, $\widetilde{h}(x, y)$ is any lifting of the value $h(x, y) \in \mathbb{T}^1$ to $\mathbb{R}$, and $(\alpha, \beta, \widetilde{h}(x, y))$ stands for an element in G. Here we regard h as a $\mathbb{Z}^2$-periodic function on $\mathbb{R}^2$.

Indeed, the choice of $\widetilde{h}(x, y)$ does not matter. This is because for two different choices of $\widetilde{h}(x, y)$, the values of $(\alpha, \beta, \widetilde{h}(x, y))$ differ by translation by an element from the group $C = \{(0, 0, m) : m \in \mathbb{Z}\}$. This group is both in the center of G and in Γ, so the two different choices of $(\alpha, \beta, \widetilde{h}(x, y))\mathbf{x}$ represent the same point in $X = G/\Gamma$.

Without causing confusion, we will simply write (1.3) as

$$T : \mathbf{x} \mapsto (\alpha, \beta, h(x, y))\mathbf{x}. \tag{1.4}$$

Here $(\alpha, \beta, h(x, y))$ should be think of as an element in the quotient group G/C.

The map T is an isometric extension of the translation by (α, β) on $\mathbb{T}^2$, which we denote by T_0. Namely, T_0 is a factor of T, and T send fibers (which are circles $\mathbb{T}^1$) to fibers by isometries. In particular, (X, T) is a distal dynamical system and has zero topological entropy.

Recall that T_0 is minimal and ergodic on $\mathbb{T}^2$ if α, β, 1 are linearly independent over $\mathbb{Q}$. Otherwise, every orbit of T_0 is contained in a finite union of parallel 1-dimensional subtori in $\mathbb{T}^2$.

Our main result is:

Theorem 1.1. *If α, β, 1 are linearly independent over $\mathbb{Q}$ and $h : \mathbb{T}^2 \to \mathbb{T}^1$ is Lipschitz continuous, then*

$$\lim_{N\to\infty} \frac{1}{N} \sum_{n=1}^{N} f(T^n\mathbf{x})\mu(n) = 0, \ \forall \mathbf{x} \in X, \forall f \in C(X). \tag{1.5}$$

We remark that the assumption on α, β is in place only to guarantee minimality, and no extra Diophantine conditions are needed.

1.2. **Background and motivation.** The Möbius disjointness conjecture, proposed by Sarnak [30], is:

Conjecture 1.2. *For a topological dynamical system (X, T), if $h_{\mathrm{top}}(T) = 0$, then (1.5) holds.*

The conjecture has been the subject of many recent researches. For known cases of the conjecture, see [1–12, 15, 17–19, 22–24, 26–29, 31, 32], to list a few.

An important class of zero entropy topological dynamical systems are distal dynamical systems. By Furstenberg's structure theorem [14], minimal distal systems are inverse limits of towers of isometric extensions.

Möbius disjointness for homogeneous distal dynamical systems were known by the works of Davenport [4] for rotations of the circle, of Green-Tao [17] for nilflows, and of Liu-Sarnak [23] for all affine distal flows.

According to Furstenberg's structure theorem, the simplest non-homogeneous distal systems are 2-step isometric extensions, i.e. an isometric extension of a rotation on a compact abelian group.

For manifolds, $\mathbb{T}^2$ is the smallest on which one can create such a map, which is the skew product $T(x, y) = (x + \alpha, y + h(x))$. Möbius disjointness for such skew products is proved for generic α when h is $C^{1+\epsilon}$ by Kułaga-Pryzmus and Lemanczyk [22], as well as for all α when T is analytic by Liu and Sarnak [23] and Wang [32].

The aim of this paper is to demonstrate that the problem is easier to handle for non-homogeneous dynamical systems when the isometric extension's underlying fiber bundle structure is not trivial.

In the settings of Theorem 1.1, the Heisenberg nilmanifold is a non-trivial principal circle bundle over $\mathbb{T}^2$. The twistedness of the topology allows to show unique ergodicity of a dynamical system that is induced from T using the Kátai-Bourgain-Sarnak-Ziegler criterion [3, 21], assuming Lipschitz continuity. In contrast, for skew products on $\mathbb{T}^2$, which is a trivial circle bundle over the circle, the works [23] and [32] required either methods from harmonic analysis or Matomäki-Radziwiłł-Tao bounds [25] on short averages of multiplicative functions, in addition to the Kátai-Bourgain-Sarnak-Ziegler criterion, and needed to assume analyticity.

We remark that the proof of Theorem 1.1 can be easily extended to skew products on higher dimensional Heisenberg manifolds and other 2-step nilmanifolds. However, we are not going to pursue this direction in detail.

Notations. On $\mathbb{T}^1 = \mathbb{R}/\mathbb{Z}$, $\|\cdot\|$ denotes the distance to the origin. The function $e(\cdot)$ on $\mathbb{T}^1$ (or $\mathbb{R}$) is defined as $e(x) = e^{2\pi i x}$. For a compact nilmanifold or torus Y, m_Y denotes the unique uniform probability measure on Y, which descends from a Haar measure on the universal cover of Y.

2. Proof of Theorem 1.1

2.1. **Reduction of the joining dynamics.** We suppose $\mathbf{x}_0 \in X$ and a function $f_0 \in C(X)$ does not satisfy (1.5). By translating both the point and the function, we may assume without loss of generality that $\mathbf{x}_0$ is the identity point Γ in $X = G/\Gamma$, i.e. represented by $(0, 0, 0)$.

Definition 2.1. A continuous function $f : X \to \mathbb{C}$ has vertical oscillation of frequency $\xi \in \mathbb{Z}$ if for all $\tau \in X$ and $z \in \mathbb{R}$,

$$f((0, 0, z)\tau) = e(\xi z)f(\tau).$$

Lemma 2.2. *There exists a non-zero integer ξ and a continuous function $f : X \to \mathbb{C}$ of vertical oscillation of frequency ξ, such that*

$$\frac{1}{N}\sum_{i=1}^{N} f(T^n\mathbf{x}_0)\mu(n) \not\to 0 \text{ as } N \to \infty. \tag{2.1}$$

Proof. By our earlier hypothesis, there are $\delta \in (0,1)$ and a subsequence $\{N_i\}$ of $\mathbb{N}$, such that

$$\left|\frac{1}{N_i}\sum_{n=1}^{N_i} f_0(T^n\mathbf{x}_0)\mu(n)\right| > \delta.$$

On the other hand, by the proof of [16, Lemma 3.7], there are finitely many continuous functions f_j, $1 \le j \le J$ on X of vertical oscillation, respectively of frequecy ξ_j, such that $\|f_0 - \sum_{j=1}^{J} f_j\|_{L^\infty} < \frac{\delta}{2}$. It follows that

$$\left|\frac{1}{N_i}\sum_{n=1}^{N_i} f_0(T^n\mathbf{x}_0)\mu(n) - \sum_{j=1}^{J}\frac{1}{N_i}\sum_{n=1}^{N_i} f_j(T^{pn}\mathbf{x}_0)\mu(n)\right| < \frac{\delta}{2}.$$

and hence

$$\left|\sum_{j=1}^{J}\frac{1}{N_i}\sum_{n=1}^{N_i} f_j(T^{pn}\mathbf{x}_0)\mu(n)\right| > \delta - \frac{\delta}{2} = \frac{\delta}{2}.$$

for all i. In other words, $\sum_{j=1}^{J}\frac{1}{N}\sum_{n=1}^{N} f_j(T^{pn}\mathbf{x}_0)\mu(n) \not\to 0$ as $N \to \infty$. We deduce that for at least one j, $\frac{1}{N}\sum_{n=1}^{N} f_j(T^{pn}\mathbf{x}_0)\mu(n) \not\to 0$. Let $f = f_j$ and $\xi = \xi_j$. Then (2.1) holds.

It remains to claim that $\xi \ne 0$. Indeed, if $\xi = 0$, then f is constant under translations along the vertical subgroup $\{(0,0,z)\}$, which are fibers of $X \to \mathbb{T}^2$. Equivalently, f can be thought of as a continuous function on $\mathbb{T}^2$, and (2.1) can be rewritten as

$$\frac{1}{N}\sum_{i=1}^{N} f(T_0^n(0,0))\mu(n) \not\to 0 \text{ as } N \to \infty.$$

As T_0 is the translation by (α,β) on $\mathbb{T}^2$, this contradicts Davenport's theorem [4]. So we conclude that $\xi \ne 0$. □

The following important criterion guarantees Möbius disjointness and is due to Kátai [21] and Bourgain-Sarnak-Ziegler [3]:

Theorem 2.3. *For a dynamical system $(\mathcal{X}, T)$, a continuous function $f \in C(\mathcal{X})$, and a point $x \in \mathcal{X}$, if the equation (1.5) fails to hold, then there exist a pair of distinct primes $p > q$, such that $\frac{1}{N}\sum_{n=1}^{N} f(T^{pn}x)\overline{f(T^{qn}x)}$ does not converge to 0 as $N \to \infty$.*

By this criterion, for a pair of distinct primes $p > q$,

$$\frac{1}{N}\sum_{n=1}^{N} f(T^{pn}\mathbf{x}_0)\overline{f(T^{qn}\mathbf{x}_0)} \not\to 0 \text{ as } N \to \infty. \tag{2.2}$$

We study the dynamics of the pair $(T^{pn}\mathbf{x}_0, T^{qn}\mathbf{x}_0)$.

Lemma 2.4. *The set* $G_1 = \{(x_1, y_1, z_1, x_2, y_2, z_2) \mid q(x_1, y_1) = p(x_2, y_2)\} \subseteq G^2$ *is a closed subgroup of* G^2 *with the following properties:*

(i) G_1/Γ_1*, where* $\Gamma_1 = G_1 \cap (\Gamma \times \Gamma)$*, is compact.*
(ii) $(T^{pn}\mathbf{x}_0, T^{qn}\mathbf{x}_0) \in X_1 = G_1/\Gamma_1$ *for all* n*.*

Proof. (i) The nilpotent group G is (the real points of) an algebraic group defined over $\mathbb{Q}$ and thus so is G^2. The lattice Γ is given by $G(\mathbb{Z})$. In order to show that Γ_1 is cocompact in G_1, it suffices to prove G_1 is a subgroup defined over $\mathbb{Q}$. This is true by definition.

(ii) Notice that $(\mathbf{x}_0, \mathbf{x}_0)$ is the identity element in $X^2 = G^2/\Gamma^2$. It suffices to show that the embedded subnilmanifold $X_1 \subset X^2$ is $T^p \times T^q$-invariant. This can be verified from the definition of G_1, because T^p adds $(p\alpha, p\beta)$ to the coordinate pair (x_1, y_1) and T^q adds $(q\alpha, q\beta)$ to (x_2, y_2). □

Lemma 2.5. *For* $D = \{(0, 0, z_1, 0, 0, z_2 \mid z_1 = z_2\} \subset G_1$ *and* $G_* = G_1/D$*, the subgroup* $\Gamma_* = \Gamma_1/\Gamma_1 \cap D$ *is a cocompact lattice in* G_**, and thus* $X_* = G_*/\Gamma_* = X_1/(D/\Gamma_1 \cap D)$ *is a compact nilmanifold.*

Proof. Remark first that D is in the center of G_1, so G_* is a group. Again, it suffices to notice that D is an algebraic subgroup of the nilpotent group G_1 defined over $\mathbb{Q}$. □

We now describe the natural projection from X_1 to X_*. Because of the definition of G_1, each point in G_1 can be uniquely written as $(px, py, z_1, qx, qy, z_2) \in G^2$ for some $x, y, z_1, z_2 \in \mathbb{R}$, where G is parametrized as in (1.1). The D-orbit of this point is the set $\{(px, py, z_1 + a, qx, qy, z_2 + a) : a \in \mathbb{R}\}$. So $G_* = G_1/D$ can be parametrized by $\{(x, y, z) : x, y, z \in \mathbb{R}\}$, and the projection $\pi : G_1 \to G_*$ is given by

$$\pi(px, py, z_1, qx, qy, z_2) = (x, y, z_1 - z_2). \tag{2.3}$$

Because p, q are distinct primes, each point in $\Gamma_1 = G_1 \cap \Gamma$ can be uniquely written as $(px, py, z_1, qx, qy, z_2) \in G^2$ for some $x, y, z_1, z_2 \in \mathbb{Z}$. Combining this with (2.3), we see that $\Gamma_* = \pi(\Gamma_1)$ is just the set of integer points $\{(x, y, z) \in G_* : x, y, z \in \mathbb{Z}\}$ of G_*.

Lemma 2.6. *The group rule in* G_**, which we denote by* $*$*, is given by*

$$(x, y, z) * (x', y', z') = (x + x', y + y', z + z' + (p^2 - q^2)(xy' - x'y)).$$

Proof. The group rule in G_1 is

$$\begin{aligned}(px,&py,z_1,qx,qy,z_2)(px',py',z_1',qx',qy',z_2') =\\ &(p(x+x'),p(y+y'),z_1+z_1'+p^2(xy'-x'y),\\ &q(x+x'),q(y+y'),z_2+z_2'+q^2(xy'-x'y)).\end{aligned}$$

Applying π to both sides, we get the formula in the lemma. □

It is not hard to see that the 2-step compact nilmanifold $X_* = G_*/\Gamma_*$, similar to the Heisenberg nilmanifold $X = G/\Gamma$, is a principal $\mathbb{T}^1$-bundle over $\mathbb{T}^2$. The base $\mathbb{T}^2$ is parametrized by the first two coordinates (x, y).

We indifferently denote by π the projection from X_1 to X_*, which is induced from $\pi : G_1 \to G_*$. By Lemma 2.4, for all n we have a point $\pi(T^{pn}\mathbf{x}, T^{qn}\mathbf{x}) \in X_*$.

The group G_* acts by left translation on X_*/Γ_*. We keep the symbol $*$ to denote this action. It should be noted that, as $\pi : G_1 \to G_*$ is a group morphism, for $g \in G_1$ and $\overline{\mathbf{x}} \in X_1$, $\pi g * \pi\overline{\mathbf{x}} = \pi(g * \overline{\mathbf{x}})$.

To proceed, we will need an expression for the n-th iterate T^n for $n \in \mathbb{N}$.

Lemma 2.7. *Let $h_n(x, y) = \sum_{i=0}^{n-1} h(x+i\alpha, y+i\beta)$. Then for $\mathbf{x} = (x, y, z)\Gamma \in X$ and $n \in \mathbb{N}$,*

$$T^n\mathbf{x} = \big(x + n\alpha, y + n\beta, h_n(x, y)\big)\mathbf{x}.$$

Proof. Because T factors to T_0 on $\mathbb{T}^2$, the projection of $T^n\mathbf{x}$ to $\mathbb{T}^2$ is represented by $(x + n\alpha, y + n\beta)$. Thus $T^{n+1}\mathbf{x} = \big(\alpha, \beta, h(x + n\alpha, y + n\beta)\big) \cdot T^n\mathbf{x}$.

When $n = 0$, the equality in the lemma automatically holds as $h_0(x, y) = 0$. Suppose the lemma is true for n, then

$$\begin{aligned}&T^{n+1}\mathbf{x}\\ =&\big(\alpha, \beta, h(x + n\alpha, y + n\beta)\big)\big(n\alpha, n\beta, h_n(x, y)\big)\mathbf{x}\\ =&\big((n + 1)\alpha,(n + 1)\beta, h_n(x, y) + h(x + n\alpha, y + n\beta) + \alpha\cdot n\beta - \beta\cdot n\alpha\big)\mathbf{x}\\ =&\big((n + 1)\alpha,(n + 1)\beta, h_{n+1}(x, y)\big)\mathbf{x}\end{aligned}$$

by the group rule (1.2). This establishes the lemma by induction. □

Given the functions h_n in Lemma 2.7, we can define a piecewise continuous function $H : \mathbb{T}^2 \mapsto \mathbb{T}^1$ by

$$H(x, y) = h_p(px, py) - h_q(qx, qy) \tag{2.4}$$

on $\mathbb{T}^2$.

Corollary 2.8. $\pi \circ (T^p \times T^q) = T_* \circ \pi$*, where*

$$T_*\mathbf{x}_* = (\alpha, \beta, H(x, y)) * \mathbf{x}_*$$

if $\mathbf{x}_* \in X_*$ is the equivalence class containing $(x, y, z) \in G_*$.

We remark that here, as in (1.4), $(\alpha,\beta,H(x,y))$ should be viewed as an element of the group G_*/C_* where $C_*=\{(0,0,m)\in G_*:m\in\mathbb{Z}\}$. For different choices of $\widetilde{H}(x,y)\in\mathbb{R}$ lifting $H(x,y)\in\mathbb{T}^1$, $(\alpha,\beta,\widetilde{H}(x,y))$ differ by a defect in C_*. As C_* is both in the cetner of G_* and in Γ_*, this defect does not affect the position of $(\alpha,\beta,\widetilde{H}(x,y))*\mathbf{x}_*$. So we can write H instead of $\widetilde{H}$ in Corollary 2.8.

Proof. Suppose $\overline{\mathbf{x}}\in X_1$ is represented by $(px,py,z_1,qx,qy,z_2)\in G_1$. By Lemma 2.7 and formula (2.3)

$$\begin{aligned}\pi((T^p\times T^q)\overline{\mathbf{x}})=&\pi((p\alpha,p\beta,h_p(px,py),q\alpha,q\beta,h_q(qx,qy))\cdot\overline{\mathbf{x}})\\=&\pi((p\alpha,p\beta,h_p(px,py),q\alpha,q\beta,h_q(qx,qy)))*\pi\overline{\mathbf{x}}\\=&(\alpha,\beta,H(x,y))*\pi\overline{\mathbf{x}}.\end{aligned}$$

The corollary is proved. □

Note that T_* is a skew product map on X_*. It also descends to T_0 on $\mathbb{T}^2$, and acts by rotations along the fiber direction. Hence, T_* preserves the uniform probability measure $\mathrm{m}_{T_{X_*}}$.

We define f_1 on $X^2=G^2/\Gamma^2$ (and thus on $X_1\subseteq X^2$) by $f_1(\mathbf{x}_1,\mathbf{x}_2)=f(\mathbf{x}_1)\overline{f}(\mathbf{x}_2)$. Because f has vertical oscillation of frequency ξ, f_1 is invariant by D. Thus f_1 descends to a function f_* on X_*.

Lemma 2.9. $\int_{X_*}f_*\mathrm{dm}_{X_*}=0$.

Proof. Since $\xi\neq 0$, we have that

$$\begin{aligned}&\int_{X_1}f_1\mathrm{dm}_{X_1}\\=&\int_{x,y,z_1,z_2\in[0,1)}f_1\big((px,py,z_1,qx,qy,z_2)\Gamma^2\big)\mathrm{d}x\mathrm{d}y\mathrm{d}z_1\mathrm{d}z_2\\=&\int_{x,y\in[0,1)}\Big(\int_0^1 f\big((px,py,z_1)\Gamma\big)\mathrm{d}z_1\Big)\Big(\int_0^1\overline{f}\big((qx,qy,z_2)\Gamma\big)\mathrm{d}z_2\Big)\mathrm{d}x\mathrm{d}y\\=&\int_{x,y\in[0,1)}0\cdot 0\mathrm{d}x\mathrm{d}y=0.\end{aligned}$$

This implies the lemma, as f_1 and m_{X_1} respectively descend to f_* and m_{X_*}. □

Let $\mathbf{x}_{0*}$ be the identity point $(0,0,0)*\Gamma_*$ in X_*. By Lemma 2.4 and Corollary 2.8, the average in (2.2) can be formulated as

(2.5)
$$\frac{1}{N}\sum_{n=1}^N f(T^{pn}\mathbf{x}_0)\overline{f}(T^{qn}\mathbf{x}_0)=\frac{1}{N}\sum_{n=1}^N f_*(T_*^n\pi(\mathbf{x}_0,\mathbf{x}_0))=\frac{1}{N}\sum_{n=1}^N f_*(T_*^n\mathbf{x}_{0*}).$$

From this, we can conclude the analysis above by stating the following proposition.

Proposition 2.10. *Under the hypotheses of this section, T_* is not uniquely ergodic.*

Proof. If T_* is uniquely ergodic, its unique invariant probability measure must be m_{X_*}. Then by Birkhoff ergodic theorem and Lemma 2.9, the ergodic averages $\frac{1}{N}\sum_{n=1}^{N} f_*(T_*^n\omega_{0*})$ converges to 0. This contradicts (2.2), because of (2.5). □

2.2. **Unique ergodicity of the reduced joining dynamics.** By the proposition above, in order to prove Theorem 1.1, it suffices to show

Proposition 2.11. *T_* is uniquely ergodic.*

In [13], Furstenberg proved that the unique ergodicity for a skew product map on a circle bundle over a uniquely ergodic base that acts as rotations on the fibers is equivalent to the non-existence of invariant multi-valued graphs. He originally stated this criterion for skew products generated by a continuous cocycle. The same proof also works for measurable cocycles, which is the statement we will need (Theorem 2.12 below). For completeness' sake, we include the proof here.

Theorem 2.12. *Let (Ω_0, T_0) be a uniquely ergodic topological dynamical system, whose unique invariant probability measure is denoted by γ_0. Take $\Omega = \Omega_0 \times \mathbb{T}^1$ and define a skew product map $T : \Omega \to \Omega$ by $T(\omega_0, \zeta) = (T_0\omega_0, g(\omega_0) + \zeta)$, where $g : \Omega_0 \to \mathbb{T}^1$ is a measurable function. Then:*

(i) The product measure $\gamma = \gamma_0 \times \mathrm{m}_{\mathbb{T}^1}$ is an invariant probability measure for T;

(ii) T is uniquely ergodic if and only if for all $k \in \mathbb{N}$, the equation

$$R(T_0\omega_0) = R(\omega_0) + kg(\omega_0) \tag{2.6}$$

has no measurable solution $R : \Omega_0 \to \mathbb{T}^1$ modulo γ_0.

Proof. The proof of Part (i) is straightforward, so we only discuss the second part.

The key claim is:

T is uniquely ergodic if and only if γ is ergodic.

To see this, define the transformation $\tau_\beta : \Omega \to \Omega$ by $\tau_\beta(\omega_0, \zeta) = (\omega_0, \beta + \zeta)$. Since $\gamma = \gamma_0 \times \mathrm{m}_{\mathbb{T}^1}$, if ω_* is a generic point for (Ω, T, γ), in the sense that $\frac{1}{N}\sum_{i=0}^{N-1} \delta_{T^i\omega_*} \to \gamma$ in the weak-* topology as $N \to \infty$, then so is $\tau_\beta(\omega_*)$ for every $\beta \in \mathbb{T}^1$.

To show ergodicity implies unique ergodicity, suppose T is ergodic with respect to γ. It follows that almost all points of Ω (with respect to γ) are generic for (Ω, T, γ). So γ_0-almost every $\omega_0 \in \Omega_0$ has the property that (ω_0, ζ) is generic for (Ω, T, γ) for $m_{\mathbb{T}^1}$-almost every ζ. By applying τ_β for all $\beta \in \mathbb{T}^1$, we see that for γ_0-almost every $\omega_0 \in \Omega_0$, (ω_0, ζ) is generic for (Ω, T, γ) for all $\zeta \in \mathbb{T}^1$. Suppose T is not uniquely ergodic, then there exists an ergodic probability measure γ' other than γ for T. As any T-invariant measure on Ω projects to an invariant measure on Ω_0, and $(\Omega_0, T_0, \gamma_0)$ is uniquely ergodic, it follows that the projection of γ' on Ω_0 is γ_0. Thus for γ_0-almost all points ω_0 in the base Ω_0, there exist extended points (ω_0, ζ) that are generic for (Ω, T, γ'). This cannot happen though, since for almost every ω_0 and all ζ, (ω_0, ζ) is generic for γ, which is different from γ'. This establishes the claim.

It remains to show that the ergodicity of γ is equivalent to the condition in part (ii).

Suppose first that γ is not ergodic. Then $Tf = f$ has a non-constant solution $f \in L^2(\Omega, \gamma)$. Since $\gamma = \gamma_0 \times m_{\mathbb{T}^1}$ is a product and f is L^2 with respect to γ, we can split f into Fourier series along the $\mathbb{T}^1$ direction and write it as

$$f = \sum_{-\infty}^{\infty} c_k(\omega_0) e(k\zeta),$$

where $c_k(\omega_0) \in L^2(\Omega_0, \gamma_0)$ and $e(\xi) = e^{2\pi i \xi}$. The condition $Tf = f$ implies $\sum_{-\infty}^{\infty} c_k(T_0\omega_*) e(kg(\omega_0) + k\zeta) = \sum_{-\infty}^{\infty} c_k(\omega_0) e(k\zeta)$, or

$$c_k(T_0\omega_0) e(kg(\omega_0)) = c_k(\omega_0) \tag{2.7}$$

for every $k \in \mathbb{Z}$.

Since T_0 is ergodic, f is not reducible to a function of ω_0 alone and thus $c_k(\omega_0) \neq 0$ for at least one non-zero integer k. By the ergodicity of T_0 it follows that c_k vanishes only on a set of measure zero, which allows us to write $c_k(\omega_0)$ as $r_k(\omega_0) e(\theta_k(\omega_0))$, where $r_k(\omega_0) > 0$ and $\theta_k(\omega_0) \in \mathbb{T}^1$. From (2.7), we get that $r_k(T_0\omega_0) e(\theta_k(T_0\omega_0) + kg(\omega_0)) = r_k(\omega_0) e(\theta_k(\omega_0))$ for every k, thus $R(\omega_0) = -\theta_k(\omega_0)$ is a solution to (2.6). In addition, if $k < 0$, then we can replace k with $-k$ and R with $-R$. So one can claim $k \in \mathbb{N}$ without loss of generality.

Conversely, if (2.6) has a solution, then the non-constant measurable function $e(-k\zeta) e(R(\omega_0))$ is invariant under T modulo γ, implying that γ is not ergodic, and we are done. □

We now reparametrize X_* in a piecewise continuous way in order to identify it with $\mathbb{T}^3 = \mathbb{T}^2 \times \mathbb{T}^1$ and apply Theorem 2.12.

In the parametrization given by Lemma 2.6, the box $[0, 1)^3$ is a fundamental domain for the projection $G_* \to X_*$. Indeed, for each $(x, y, z) \in G_*$, there is a unique element of Γ_*, which we denote by $\lfloor (x, y, z) \rfloor$, such that

$(x,y,z)*\lfloor(x,y,z)\rfloor^{-1}\in[0,1)^3$. Given the group rule (1.2), it is not hard to check that

$$\lfloor(x,y,z)\rfloor=(\lfloor x\rfloor,\lfloor y\rfloor,\lfloor z-(p^2-q^2)(x\lfloor y\rfloor-\lfloor x\rfloor y)\rfloor). \tag{2.8}$$

Thus the map $\rho_0: X_*\to[0,1)^3$ given by

$$\begin{aligned}\rho_0:(x,y,z)\Gamma\mapsto&(x,y,z)*\lfloor(x,y,z)\rfloor^{-1}\\ &=(\{x\},\{y\},\{z-(p^2-q^2)(x\lfloor y\rfloor-\lfloor x\rfloor y)\})\end{aligned} \tag{2.9}$$

is bijective and provides a piecewise continuous parametrization of X_* by $[0,1)^3$. Here $\{x\}$ stands for $x-\lfloor x\rfloor$, the fractional part of x.

If $\mathbf{x}_*\in X_*$ is represented by $(x,y,z)\in F$, then $T_*\mathbf{x}$ is represented by $\big(x+\alpha,y+\beta,z+(p^2-q^2)(\alpha y-\beta x)+H(x,y)\big)\in G_*$, and thus can also be represented by the element

$$\Big(\{x+\alpha\},\{y+\beta\},\Big\{z+H(x,y)\\ +(p^2-q^2)\big((\alpha y-\beta x)-(x+\alpha)\lfloor y+\beta\rfloor+\lfloor x+\alpha\rfloor(y+\beta)\big)\Big\}\Big)$$

in $[0,1)^3$.

If we identify $[0,1)^3$ with $\mathbb{T}^3$ in the natural way, and let ρ be the composition given by $X_*\overset{\rho_0}{\to}[0,1)^3\to\mathbb{T}^3$, then ρ is bijective and piecewise continuous. Moreover, the discussion above shows that T_* is conjugate to the map

$$T'_*:(x,y,z)\mapsto(x+\alpha,y+\beta,z+H'(x,y)) \tag{2.10}$$

on $\mathbb{T}^3$ by the piecewise continuous bijection ρ, where $H':\mathbb{T}^2\to\mathbb{R}$ is defined by

$$\begin{aligned}H'(x,y)=&H(x,y)+(p^2-q^2)\big((\alpha y-\beta x)\\ &-(x+\alpha)\lfloor y+\beta\rfloor+\lfloor x+\alpha\rfloor(y+\beta)\big)\end{aligned} \tag{2.11}$$

for $(x,y)\in[0,1)^2$ and regarded as a piecewise continuous map on $\mathbb{T}^2$.

Therefore, in view of Theorem 2.12, in order to obtain Proposition 2.11, it suffices to show the following lemma:

Lemma 2.13. *For all $k\in\mathbb{N}$, the equation*

$$R(x+\alpha,y+\beta)=R(x,y)+kH'(x,y) \tag{2.12}$$

has no measurable solution $R:\mathbb{T}^2\to\mathbb{T}^1$ modulo $\mathrm{m}_{\mathbb{T}^2}$.

Our approach to Lemma 2.13 is inspired by [13, Lemma 2.2]

Notice first that, suppose $R(x,y)$ is such a solution, then the set

$$\Lambda':=\{(x,y,z)\in\mathbb{T}^3:kz=R(x,y)\},$$

which is a multi-valued graph over $\mathbb{T}^2$, is T'_* invariant except for a $m_{\mathbb{T}^2}$-null set of (x, y). Let $\Lambda = \rho^{-1}(\Lambda') \subset X_*$. Then Λ intersects every $\mathbb{T}^1$-fiber in exactly k points that form a translate of $\frac{1}{k}\mathbb{Z}/\mathbb{Z}$. Moreover, Λ is almost T_*-invariant, in the sense that there is a subset $A \subseteq \mathbb{T}^2$ with $m_{\mathbb{T}^2}(A) = 1$, such that if $\mathbf{x}_* \in \Lambda \cap \pi_{\mathbb{T}^2}^{-1}(A)$, then $T_*\mathbf{x}_* \in \Lambda$.

Lemma 2.14. *For $\mathbf{x}_* = (x, y, z)\Gamma \in X_*$ and $n \in \mathbb{N}$,*

$$T_*^n \mathbf{x}_* = \big(x + n\alpha, y + n\beta, H_n(x, y)\big)\mathbf{x}_*,$$

where $H_n(x, y) = \sum_{i=0}^{n-1} H(x + i\alpha, y + i\beta)$.

Proof. The proof is the same as that of Lemma 2.7, using the new group rule $*$ in lieu of (1.2). ∎

Given $n \in \mathbb{N}$, remark that T_*^n is conjugate by ρ to the $(T'_*)^n$. Repeating the proof of (2.10), we can show similarly that

$$(T'_*)^n(x, y, z) = (x + n\alpha, y + n\beta, z + H'_n(x, y)) \tag{2.13}$$

on $\mathbb{T}^3$, where $H'_n : \mathbb{T}^2 \to \mathbb{R}$ is defined by

$$\begin{aligned} H'_n(x, y) =& H_n(x, y) + (p^2 - q^2)\big((n\alpha y - n\beta x) \\ & - (x + n\alpha)\lfloor y + n\beta\rfloor + \lfloor x + n\alpha\rfloor(y + n\beta)\big). \end{aligned} \tag{2.14}$$

for $(x, y) \in [0, 1)^2$.

Because Λ is almost T_* invariant, it is also almost T_*^n invariant. And Λ' is almost $(T'_*)^n$ invariant in the same sense, i.e. for a subset $A \subseteq \mathbb{T}^2$ of full $m_{\mathbb{T}^2}$-measure, if $\mathbf{x}'_* \in \Lambda' \cap \pi_{\mathbb{T}^2}^{-1}(A)$, then $(T'_*)^n\mathbf{x}'_* \in \Lambda'$. This is equivalent to the statement that the equation

$$R(x + n\alpha, y + n\beta) = R(x, y) + kH'_n(x, y) \tag{2.15}$$

holds for $m_{\mathbb{T}^2}$-almost all (x, y).

Proof of Lemma 2.13. Suppose $k \in \mathbb{N}$ and $R : \mathbb{T}^2 \mapsto \mathbb{T}^1$ is a measurable solution of (2.12). Let

$$\delta_1 = \frac{|k(p^2 - q^2)d_1 - k(p^2 - q^2)\beta - \beta|}{24k(p^2 + q^2)(L + |\alpha| + |\beta|)}, \tag{2.16}$$

and

$$\nu = \frac{6}{|k(p^2 - q^2)d_1 - k(p^2 - q^2)\beta - \beta|}, \tag{2.17}$$

where d_1 is the degree of h in x and L is the Lipschitz constant of h. Note $\delta_1 > 0$ and $\nu < \infty$ because $p > q$, $k > 0$, $p, q, k, d_1 \in \mathbb{Z}$ and $\beta \notin \mathbb{Q}$. By Luzin's theorem, we can find a compact subset $\Phi \subset \mathbb{T}^2$ of measure greater than $1 - \delta_1$ such that R is continuous when restricted to Φ.

Choose $\delta_2 \in (0, \min(\frac{1}{6}, \delta_1))$ such that if $(x, y), (x', y') \in \Phi$ and $\|(x, y) - (x', y')\| < \delta_2$, then $\|R(x, y) - R(x', y')\| < \frac{1}{3}$. We then fix $n \in \mathbb{N}$ such that $\{n\alpha\}, \{n\beta\} \in (0, \delta_2)$ and $n > \nu$. Such integers n exist because T_0 is minimal on $\mathbb{T}^2$.

For $(x, y) \in (0, 1 - \delta_2)^2$, we have that $x + \{n\alpha\}, y + \{n\beta\} \in (0, 1)$ and $\lfloor x + n\alpha \rfloor = \lfloor n\alpha \rfloor$, $\lfloor y + n\beta \rfloor = \lfloor n\beta \rfloor$. Hence,

$$\begin{aligned} H'_n(x, y) =& H_n(x, y) + n(p^2 - q^2)(\alpha y - \beta x) - \lfloor n\beta \rfloor (x + n\alpha) \\ & + \lfloor n\alpha \rfloor (y + n\beta)\big), \ \forall (x, y) \in (\delta_2, 1 - \delta_2)^2. \end{aligned} \tag{2.18}$$

On the other hand, for $(x, y) \in \Phi \cap \big(\Phi - (n\alpha, n\beta)\big)$, $\|R(x + n\alpha, y + n\beta) - R(x, y)\| < \frac{1}{3}$. So by (2.15),

$$\|kH'_n(x, y)\| < \frac{1}{3}, \ \forall (x, y) \in \Phi \cap \big(\Phi - (n\alpha, n\beta)\big). \tag{2.19}$$

Because $\mathrm{m}_{\mathbb{T}^2}\big(\Phi \cap \big(\Phi - (n\alpha, n\beta)\big)\big) > 1 - 2\delta_1$ and $\mathrm{m}_{\mathbb{T}^2}\big((0, 1 - \delta_2)^2\big) > (1 - \delta_2)^2 > 1 - 2\delta_2 > 1 - 2\delta_1$, by combining (2.18) and (2.19), we know that

$$\Big\|kH_n(x, y) + nk(p^2 - q^2)(\alpha y - \beta x) - \lfloor n\beta \rfloor (x + n\alpha) + \lfloor n\alpha \rfloor (y + n\beta)\Big\| < \frac{1}{3} \tag{2.20}$$

on a subset $\Phi_1 \subset [0, 1)^2$ with $\mathrm{m}_{\mathbb{R}^2}(\Phi_1) > 1 - 4\delta_1$, where $\mathrm{m}_{\mathbb{R}^2}$ is the Lebesgue measure on $\mathbb{R}^2$.

Fix a continuous lifting $\widetilde{H}_n : \mathbb{R}^2 \to \mathbb{R}^1$ of the function $H_n : \mathbb{T}^2 \to \mathbb{T}^1$. Then (2.20) actually asserts the continuous function

$$F_n(x, y) = k\widetilde{H}_n(x, y) + nk(p^2 - q^2)(\alpha y - \beta x) - \lfloor n\beta \rfloor (x + n\alpha) + \lfloor n\alpha \rfloor (y + n\beta)$$

takes values in $\bigcup_{m \in \mathbb{Z}} (m - \frac{1}{3}, m + \frac{1}{3})$ on Φ_1.

Because h has degree d_1 in x, h_j has degree jd_1 in x. Thus $H(x, y) = h_p(px, py) - h_q(qx, qy)$ has degree $(p^2 - q^2)d_1$ in x. It in turn follows that $H_n(x, y)$ has degree $n(p^2 - q^2)d_1$ in x. In consequence, for all $y \in \mathbb{R}$, $\widetilde{H}_n(1, y) - \widetilde{H}_n(0, y) = n(p^2 - q^2)d_1$ and thus

$$F_n(1, y) - F_n(0, y) = nk(p^2 - q^2)d_1 - nk(p^2 - q^2)\beta - \lfloor n\beta \rfloor. \tag{2.21}$$

By Fubini's Theorem, there exists $y_0 \in [0, 1)$ such that

$$\mathrm{m}_{\mathbb{R}}(\{x \in [0, 1] : (x, y_0) \notin \Phi_1\}) < 4\delta_1. \tag{2.22}$$

Because F_n takes values in $\bigcup_{m \in \mathbb{Z}} (m - \frac{1}{3}, m + \frac{1}{3})$ on Φ_1, the image

$$\{F(x, y_0) : x \in [0, 1], (x, y_0) \notin \Phi_1\} \subset \mathbb{R} \tag{2.23}$$

has at least Lebesgue measure

$$
\begin{aligned}
&\frac{1}{3}(|F_n(1,y)-F_n(0,y)|-1)\\
&\geq \frac{1}{3}(|nk(p^2-q^2)d_1-nk(p^2-q^2)\beta-n\beta|-2)\\
&\geq n\cdot\frac{1}{3}|k(p^2-q^2)d_1-k(p^2-q^2)\beta-\beta|-1.
\end{aligned}
\tag{2.24}
$$

On the other hand, because h is L-Lipschitz, h_j is jL-Lipschitz and $H(x,y)$ is $(p^2+q^2)L$-Lipschitz. It in turn follows that $H_n(x,y)$ is $n(p^2+q^2)L$-Lipschitz and so is $\widetilde{H}_n$. From (2.20), the Lipschitz constant of F_n is at most

$$nk(p^2+q^2)L+nk|p^2-q^2|(|\alpha|+|\beta|)+(|\alpha|+|\beta|)\leq nk(p^2+q^2)(L+|\alpha|+|\beta|).$$

So the image (2.23) has at most Lebesgue measure

$$
\begin{aligned}
&nk(p^2+q^2)(L+|\alpha|+|\beta|)\cdot 4\delta_1\\
&\leq n\cdot\frac{1}{6}|k(p^2-q^2)d_1-k(p^2-q^2)\beta-\beta|.
\end{aligned}
\tag{2.25}
$$

Comparing (2.24) with (2.25) yields that

$$n\cdot\frac{1}{6}|k(p^2-q^2)d_1-k(p^2-q^2)\beta-\beta|\leq 1.$$

However, this contradicts the hypothesis that $n>\nu$. We arrive at a contradiction and the statement is proven. □

Proof of Theorem 1.1. Lemma 2.13 and Theorem 2.12 imply Proposition 2.11, contradicting Proposition 2.10. Thus the standing hypothesis in Section 2.1 can not be true. In other words, (1.5) must hold. □

References

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

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

[3] J. Bourgain, P. Sarnak, and T. Ziegler, *Disjointness of Möbius from horocycle flows*, From Fourier analysis and number theory to Radon transforms and geometry, 2013, pp. 67–83.

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

[5] Tomasz Downarowicz and Stanisław Kasjan, *Odometers and Toeplitz systems revisited in the context of Sarnak's conjecture*, Studia Math. **229** (2015), no. 1, 45–72.

[6] El Houcein El Abdalaoui, Stanisław Kasjan, and Mariusz Lemańczyk, *0-1 sequences of the Thue-Morse type and Sarnak's conjecture*, Proc. Amer. Math. Soc. **144** (2016), no. 1, 161–176.

[7] El Houcein El Abdalaoui, J. Kułaga-Przymus, Mariusz Lemańczyk, and Thierry de la Rue, *The Chowla and the Sarnak conjectures from ergodic theory point of view*, Discrete Contin. Dyn. Syst. (2016), to appear, arXiv:1410.1673v3.

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

[9] ———, *Automorphisms with quasi-discrete spectrum, multiplicative functions and average orthogonality along short intervals*, International Mathematics Research Notices (2016), to appear.

[10] Aihua Fan and Yunping Jiang, *Oscillating sequences, minimal mean attractability and minimal mean-Lyapunov-stability*, preprint (2015), arXiv:1511.05022v1.

[11] Sébastien Ferenzi, Joanna Kułaga-Przymus, Mariusz Lemańczyk, and C. Mauduit, *Substitutions and Möbius disjointness*, preprint (2015), arXiv:1507.01123v1.

[12] Livio Flaminio, Krzyszto Fraczek, Joanna Kułaga-Przymus, and Mariusz Lemańczyk, *Approximate orthogonality of powers for ergodic affine unipotent diffeomorphisms on nilmanifolds*, preprint (2016).

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

[14] ———, *The structure of distal flows*, Amer. J. Math. **85** (1963), 477–515.

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

[16] Ben Green and Terence Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Ann. of Math. (2) **175** (2012), no. 2, 465–540.

[17] ———, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) **175** (2012), no. 2, 541–566.

[18] Wen Huang, Zhengxing Lian, Song Shao, and Xiangdong Ye, *Sequences from zero entropy noncommutative toral automorphisms and Sarnak Conjecture*, preprint (2015), arxiv:1510.06022v1.

[19] Wen Huang, Zhiren Wang, and Guohua Zhang, *Möbius disjointness for topological model of any ergodic system with discrete spectrum*, preprint (2016), arxiv:1608.08289v2.

[20] Davit Karagulyan, *On Möbius orthogonality for interval maps of zero entropy and orientation-preserving circle homeomorphisms*, Ark. Mat. **53** (2015), no. 2, 317–327.

[21] I. Kátai, *A remark on a theorem of H. Daboussi*, Acta Math. Hungar. **47** (1986), no. 1-2, 223–225.

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

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

[24] Bruno Martin, Christian Mauduit, and Joël Rivat, *Théoréme des nombres premiers pour les fonctions digitales*, Acta Arith. **165** (2014), no. 1, 11–45 (French).

[25] Kaisa Matomäki, Maksym Radziwiłł, and Terence Tao, *An averaged form of Chowla's conjecture*, Algebra Number Theory **9** (2015), no. 9, 2167–2196.

[26] Christian Mauduit and Joël 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).

[27] ———, *Prime numbers along Rudin-Shapiro sequences*, J. Eur. Math. Soc. (JEMS) **17** (2015), no. 10, 2595–2642.

[28] Clemens Müllner, *Automatic sequences fulfill the Sarnak conjecture*, Duke Math J. (2016), to appear.

[29] Ryan Peckner, *Möbius disjointness for homogeneous dynamics*, preprint (2015), arXiv:1506.07778v1.

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

[31] William Veech, *Moebius orthogonality for generalized Morse-Kakutani flows*, Amer. J. Math. (2016), to appear.

[32] Zhiren Wang, *Möbius disjointness for analytic skew products*, Invent. Math. **209** (2017), no. 1, 175–196.

University of California - Davis, Davis, CA 95616, USA
E-mail address: mclitman@ucdavis.edu

Pennsylvania State University, University Park, PA 16801, USA
E-mail address: zhirenw@psu.edu