A NOTE ON MÔBIUS DISJOINTNESS FOR SKEW PRODUCTS ON A CIRCLE AND A NILMANIFOLD

XIAOGUANG HE and KE WANG

(Received 20 June 2022; accepted 21 August 2022; first published online 4 October 2022)

Abstract

Let \( T \) be the unit circle and \( \Gamma \setminus G \) the 3-dimensional Heisenberg nilmanifold. We consider the skew products on \( T \times \Gamma \setminus G \) and prove that the Möbius function is linearly disjoint from these skew products which improves the recent result of Huang, Liu and Wang ['Möbius disjointness for skew products on a circle and a nilmanifold', *Discrete Contin. Dyn. Syst.* 41(8) (2021), 3531–3553].

2020 *Mathematics subject classification*: primary 37A44; secondary 11L03, 11N37.

*Keywords and phrases*: Möbius function, distal flow, skew product, nilmanifold, measure complexity.

1. Introduction

Let \((X, T)\) be a topological dynamic system, where \(X\) is a compact metric space and \(T : X \to X\) a continuous map. The Möbius function \(\mu : \mathbb{N} \to \{-1, 0, 1\}\) is defined by \(\mu(1) = 1, \mu(n) = (-1)^k\) when \(n\) is the product of \(k\) distinct primes and \(\mu(n) = 0\) otherwise. The Sarnak conjecture \([12, 13]\) (also called the Möbius disjointness conjecture) is the following statement.

**Conjecture 1.1.** Let \((X, T)\) be a topological dynamical system with zero topological entropy. Then

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} f(T^n x) \mu(n) = 0 \quad \text{for all } f \in C(X) \text{ and all } x \in X.
\]

In recent years, there have been many results supporting the Möbius disjointness conjecture (see the comprehensive survey [2]). Here we discuss only the historical developments that are more relevant to this paper.

We consider the skew product \((T^2, T)\), where \(T^2 = (\mathbb{R}/\mathbb{Z})^2\), \(T\) is the transformation \(T : (x, y) \mapsto (x + \alpha, y + h(x))\) and \(h : T \to T\) is a continuous function. Since \(T\) is distal...
and distal systems have zero topological entropy [11], \(T\) should satisfy the Möbius disjointness conjecture. In fact, skew products are building blocks of distal flows according to Furstenberg’s structure theorem of minimal distal flows [3].

The skew product was first considered by Liu and Sarnak [10]. They proved Conjecture 1.1 for \(h\) analytic and \(|\hat{h}(m)| \gg e^{-\tau|m|}\) for some \(\tau > 0\), where \(\hat{h}(m)\) is the \(m\)th Fourier coefficient of \(h\). After that, Wang [16] removed the additional condition and hence obtained the Möbius disjointness conjecture for all analytic skew products on \((\mathbb{T}^2, T)\). Huang et al. [7] improved the result, assuming that \(h\) is \(C^\infty\)-smooth. Recently, Kanigowski et al. [8] proved it when \(h\) is \(C^{2+\varepsilon}\)-smooth and \(\hat{h}(0) = 0\), and de Faveri [1] proved it just assuming that \(h\) is \(C^{1+\varepsilon}\)-smooth. Kułaga-Przymus and Lemańczyk [9] proved that, if \(h\) is \(C^{1+\varepsilon}\)-smooth and \(\alpha\) is topological generic, then the Möbius disjointness conjecture is true. In 2020, Wang and Yao [15] proved strong orthogonality between the Möbius function and the skew products when \(h\) is \(C^{3+\varepsilon}\)-smooth and \(\alpha\) is measure-theoretically generic. A nilsystem is also a distal flow and the Möbius disjointness conjecture for nilsystems was proved by Green and Tao [5].

Recently, Huang et al. [6] considered the Möbius disjointness conjecture for skew products on \(\mathbb{T} \times \Gamma \backslash G\), where \(G\) is a 3-dimensional Heisenberg group with the cocompact discrete subgroup \(\Gamma\), namely

\[
G = \begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}.
\]

Then \(\Gamma \backslash G\) is the 3-dimensional Heisenberg nilmanifold. Their main result is as follows.

**Theorem 1.2 (Huang et al., [6]).** Let \(\mathbb{T}\) be the unit circle and \(\Gamma \backslash G\) the 3-dimensional Heisenberg nilmanifold. Let \(\alpha \in [0, 1)\) and let \(\varphi, \psi\) be \(C^\infty\)-smooth periodic functions from \(\mathbb{R}\) to \(\mathbb{R}\) with period 1 such that

\[
\int_0^1 \varphi(t) \, dt = 0.
\]

Let the skew product \(T\) on \(\mathbb{T} \times \Gamma \backslash G\) be given by

\[
T: (t, \Gamma g) \mapsto \left( t + \alpha, \Gamma g \begin{pmatrix} 1 & \varphi(t) & \psi(t) \\ 0 & 1 & \varphi(t) \\ 0 & 0 & 1 \end{pmatrix} \right).
\]

Then, for any \((t_0, \Gamma g_0) \in \mathbb{T} \times \Gamma \backslash G\) and any \(f \in C(\mathbb{T} \times \Gamma \backslash G)\),

\[
\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mu(n) f(T^n(t_0, \Gamma g_0)) = 0.
\]

**Remark 1.3.** In [6], it is also proved that the skew product (1.1) on \(\mathbb{T} \times \Gamma \backslash G\) is a distal flow.
As mentioned by Huang et al. [6], we can combine the ideas from [1, 6, 8] to give the following new theorem, which improves the result in Theorem 1.2.

**Theorem 1.4.** Let $\mathbb{T}$ be the unit circle and $\Gamma \backslash G$ the 3-dimensional Heisenberg nilmanifold. Let $\alpha \in [0, 1)$. Let $\phi : \mathbb{T} \to \mathbb{R}$ be a $C^{2+\varepsilon}$-smooth function and let $\psi : \mathbb{T} \to \mathbb{R}$ be a $C^{1+\varepsilon}$-smooth function. Moreover, assume that $\hat{\psi}(0) = \hat{\phi}(0) = 0$. Let the skew product $T$ on $\mathbb{T} \times \Gamma \backslash G$ be given by

$$T : (t, \Gamma g) \mapsto \left( t + \alpha, \Gamma g \begin{pmatrix} 1 & \phi(t) \\ 0 & 1 \\ 0 & 0 \end{pmatrix} \right). \quad (1.2)$$

Then, for any $(t_0, \Gamma g_0) \in \mathbb{T} \times \Gamma \backslash G$ and any $f \in C(\mathbb{T} \times \Gamma \backslash G)$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mu(n) f(T^n(t_0, \Gamma g_0)) = 0.

**Remark 1.5.** From the definition of $T$ in (1.2), we can easily compute

$$T^n : (t, \Gamma g) \mapsto \left( t + n\alpha, \Gamma g \begin{pmatrix} 1 & \Phi(n, t) & \Psi(n, t) + \frac{1}{2} \Phi^2(n, t) - \frac{1}{2} H(n, t) \\ 0 & 1 & \Phi(n, t) \\ 0 & 0 & 1 \end{pmatrix} \right),$$

where $\Phi(n, t) = \sum_{l=0}^{n-1} \phi(l\alpha + t)$, $\Psi(n, t) = \sum_{l=0}^{n-1} \psi(l\alpha + t)$ and $H(n, t) = \sum_{l=0}^{n-1} \phi^2(l\alpha + t)$.

**Notation.** For a topological dynamical system, let $M(X, T)$ be the set of $T$-invariant Borel probability measures on $X$. We write $e(x)$ for $e^{2\pi i x}$ and $|x| = \min_{n \in \mathbb{Z}} |x - n|$ for the distance between $x$ and the nearest integer. For positive $A$, the notation $B = O(A)$ or $B \ll A$ means that there exists a positive constant $c$ such that $|B| \leq cA$. If the constant $c$ depends on a parameter $b$, we write $B = O_b(A)$ or $B \ll_b A$. The notation $A \asymp B$ means that $A \ll B$ and $B \ll A$. For a topological space $X$, we use $C(X)$ to denote the set of all continuous complex-valued functions on $X$.

### 2. Theorem 1.4 for rational $\alpha$

Let $G$ be the 3-dimensional Heisenberg group, with the cocompact discrete subgroup $\Gamma$, and $\Gamma \backslash G$ the 3-dimensional Heisenberg nilmanifold. We follow the papers of Huang et al. [6, Section 2] and Tolimieri [14] for the following lemmas.

For $0 \leq j \leq m - 1$, $j, m \in \mathbb{Z}$, we define the functions $\phi_{m_j}$ and $\phi^*_{m_j}$ on $G$ by

$$\phi_{m_j} \begin{pmatrix} 1 & y & z \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} = e(mz + jx) \sum_{k \in \mathbb{Z}} e^{-\pi(y+k+j/m)^2} e(mx),$$

and

$$\phi^*_{m_j} \begin{pmatrix} 1 & y & z \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} = ie(mz + jx) \sum_{k \in \mathbb{Z}} e^{-\pi(y+k+j/m+1/2)^2} e\left(\frac{1}{2}(y + k + \frac{j}{m}) + mx\right).$$
We can check that $\phi_{mj}$ and $\phi_{mj}^*$ are $\Gamma$-invariant, that is,

$$\phi_{mj}(\gamma g) = \phi_{mj}(g), \quad \phi_{mj}^*(\gamma g) = \phi_{mj}^*(g)$$

for any $g \in G$ and for any $\gamma \in \Gamma$. Thus, $\phi_{mj}$ and $\phi_{mj}^*$ can be regarded as functions on the nilmanifold $\Gamma \backslash G$.

**Lemma 2.1** [6, Proposition 2.3]. Let $\mathcal{A}$ be the subset of $f \in C(\mathbb{T} \times \Gamma \backslash G)$ such that

$$f : \left( t, \Gamma \begin{pmatrix} 1 & y & z \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} \right) \mapsto e(\xi_1 t + \xi_2 x + \xi_3 y) \phi \left( \begin{pmatrix} 1 & y & z \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} \right),$$

where $\xi_1, \xi_2, \xi_3 \in \mathbb{Z}$ and $\phi = \phi_{mj}, \bar{\phi}_{mj}, \phi_{mj}^*$ or $\bar{\phi}_{mj}^*$ for some $0 \leq j \leq m - 1$. Here, $\bar{\phi}_{mj}$ and $\bar{\phi}_{mj}^*$ stand for the complex conjugates of $\phi_{mj}$ and $\phi_{mj}^*$, respectively. Let $\mathcal{B}$ be the subset of $f \in C(\mathbb{T} \times \Gamma \backslash G)$ satisfying $f : (t, \Gamma g) \mapsto f_1(t)f_2(\Gamma g)$ with $f_1 \in C(\mathbb{T})$ and $f_2 \in C_0(\Gamma \backslash G)$. Then the $\mathbb{C}$-linear subspace spanned by $\mathcal{A} \cup \mathcal{B}$ is dense in $C(\mathbb{T} \times \Gamma \backslash G)$.

To prove Theorem 1.4 for rational $\alpha$, we need to consider two cases: $f \in \mathcal{A}$ and $f \in \mathcal{B}$. We use the argument given by Huang et al. [6, Section 3].

**Lemma 2.2** [1, Theorem 1]. Let $\alpha \in \mathbb{R}$ and let $h : \mathbb{T} \to \mathbb{T}$ be $C^{1+\varepsilon}$-smooth. Define the skew product $T : \mathbb{T}^2 \to \mathbb{T}^2$ by

$$T : (x, y) \mapsto (x + \alpha, y + h(x)).$$

Then the Möbius disjointness conjecture holds for this $(\mathbb{T}^2, T)$.

**Corollary 2.3.** Let $\alpha \in \mathbb{R}$ and let $h_1, h_2 : \mathbb{T} \to \mathbb{T}$ be $C^{1+\varepsilon}$-smooth functions. Let $T : \mathbb{T}^3 \to \mathbb{T}^3$ be given by

$$T : (x, y, z) \mapsto (x + \alpha, y + h_1(x), z + h_2(x)).$$

Then the Möbius disjointness conjecture holds for $(\mathbb{T}^3, T)$.

**Proof.** The proof is similar to the proof of [6, Corollary 3.2]; at the last step, we use Lemma 2.2. □

**Lemma 2.4.** Let $\mathcal{B} \subset C(\mathbb{T} \times \Gamma \backslash G)$ be as in Lemma 2.1. Let $T$ be as in Theorem 1.4 and let $f \in \mathcal{B}$. Then, for any $(t_0, \Gamma g_0) \in \mathbb{T} \times \Gamma \backslash G$,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mu(n) f(T^n(t_0, \Gamma g_0)) = 0.$$

**Proof.** The proof is similar to the proof of [6, Corollary 3.2] but using Corollary 2.3. □
LEMMA 2.5 [6, Proposition 3.5]. Let \((\mathbb{T} \times \Gamma \backslash G, T)\) be as in Theorem 1.4 and assume \(\alpha \in \mathbb{Q} \cap [0, 1)\). Let \(\mathcal{A}\) be as in Proposition 2.1. Then, for any \((t_0, \Gamma g_0) \in \mathbb{T} \times \Gamma \backslash G\), any \(f \in \mathcal{A}\) and any \(A > 0\),

\[
\sum_{n \leq N} \mu(n) f(T^n(t_0, \Gamma g_0)) \ll_A \frac{N}{\log^A N},
\]

where the implied constant depends on \(A\) and \(\alpha\) only.

PROPOSITION 2.6. Theorem 1.4 holds for rational \(\alpha\).

PROOF. The desired result follows from Lemmas 2.1, 2.4 and 2.5. \(\square\)

3. Theorem 1.4 for irrational \(\alpha\)

Before proving Theorem 1.4 for irrational \(\alpha\), we choose a proper metric on \(\mathbb{T} \times \Gamma \backslash G\). This can be found in [4, Sections 2 and 5] and [6, Section 5]. Let the closed connected subgroups \(G = G_1 \supseteq G_2 \supseteq G_3 = \{\text{id}_G\}\) be the lower central series filtration \(G_\bullet\) on \(G\). Let

\[
X_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad X_2 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad X_3 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.
\]

Then \(X = \{X_1, X_2, X_3\}\) is a Mal’cev basis adapted to \(G_\bullet\). The corresponding Mal’cev coordinate map \(\kappa : G \to \mathbb{R}^3\) is given by

\[
\kappa \begin{pmatrix} 1 & y & z \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} = (x, y, z - xy).
\] (3.1)

The metric \(d_G : G \times G \to \mathbb{R}_{\geq 0}\) on \(G\) is defined to be the largest metric such that \(d_G(g_1, g_2) \leq |\kappa(g_1^{-1}g_2)|\), where \(|\cdot|\) is the \(l^\infty\)-norm on \(\mathbb{R}^3\). This metric can be more explicitly expressed as

\[
d_G(g_1, g_2) = \inf \left\{ \sum_{i=0}^{n-1} \min(|\kappa(h_i^{-1}h_{i-1})|, |\kappa(h_i^{-1}h_{i-1})|) : h_0, \ldots, h_n \in G; h_0 = g_1, h_n = g_2 \right\},
\]

from which we can see that \(d_G\) is left-invariant. By (3.1),

\[
\left| \begin{pmatrix} 1 & y & z \\ 0 & 1 & x \\ 0 & 0 & 1 \end{pmatrix} \right| \leq |x| + |y| + |z|
\]

provided that \(x, y \in [0, 1)\). The above metric on \(G\) descends to a metric on \(\Gamma \backslash G\) given by

\[
d_{\Gamma \backslash G}(\Gamma g_1, \Gamma g_2) := \inf\{d_G(g_1', g_2') : g_1, g_2' \in G, \Gamma g_1 = \Gamma g_1', \Gamma g_2 = \Gamma g_2'\}.
\]
It can be proved that $d_{\Gamma \setminus G}$ is indeed a metric on $\Gamma \setminus G$. Since $d_G$ is left-invariant, we also have

$$d_{\Gamma \setminus G}(\Gamma g_1, \Gamma g_2) = \inf_{\gamma \in \Gamma} d_G(g_1, \gamma g_2).$$

Finally, we take $d_T$ to be the canonical Euclidean metric on $T$, and $d = d_{T \times \Gamma \setminus G}$ the $l^2$-product metric of $d_T$ and $d_{\Gamma \setminus G}$ given by

$$d((t_1, \Gamma g_1), (t_2, \Gamma g_2)) = (d_T(t_1, t_2)^2 + d_{\Gamma \setminus G}(\Gamma g_1, \Gamma g_2)^2)^{1/2}.$$

To prove Theorem 1.4 for irrational $\alpha$, we use the recent new ideas of Kanigowski et al. [8, Theorem 1] and de Faveri [1] to give the following lemmas.

**Lemma 3.1.** Let $0 < \varepsilon < 1/100$ and $\alpha$ be an irrational number. Let $\phi : T \to \mathbb{R}$ be a $C^{2+\varepsilon}$-smooth function and let $\psi : T \to \mathbb{R}$ be a $C^{1+\varepsilon}$-smooth function. Assume that $\hat{\phi}(0) = \hat{\psi}(0) = 0$. Then there exists an unbounded sequence $\{r_n\}_{n \geq 1}$ of $\mathbb{N}$ such that

$$\int_{T \times \Gamma \setminus G} d(T^{r_n}(t, \Gamma g), (t, \Gamma g))^2 \, d\nu(t, \Gamma g) \ll r_n^{-\varepsilon/100}$$

for any $\nu \in M(T \times \Gamma \setminus G, T)$.

**Lemma 3.2 [8].** Let $(X, T)$ be a topological dynamical system and assume that for every $\nu \in M(X, T)$, $(X, \mathcal{B}, \nu, T)$ satisfies the PR rigidity condition: there exists a linearly dense set $\mathcal{F} \subset C(X)$ such that for each $f \in \mathcal{F}$, we can find $\delta > 0$ and a sequence $\{q_n\}_{n \geq 1}$ satisfying

$$\sum_{j=\lceil -q_n \rceil}^{q_n} \| f \circ T^{jq_n} - f \|^2_{L^2(\nu)} \to 0.$$

Then, $(X, T)$ is Möbius disjoint, that is,

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} \mu(n) f(T^n x) = 0.$$

**Proof of Theorem 1.4 Assuming Lemma 3.1.** The case that $\alpha$ is rational has been proved in Proposition 2.6. So we assume that $\alpha$ is irrational. Let $\{r_n\}_{n \geq 1}$ be the sequence from Lemma 3.1. For any $\nu \in M(T \times \Gamma \setminus G, T)$,

$$\| f \circ T^{kr_n} - f \|^2_{L^2(\nu)} \leq |k| \sum_{j=1}^{k} \| f \circ T^{jr_n} - f \circ T^{(j-1)r_n} \|^2_{L^2(\nu)} = k^2 \cdot \| f \circ T^{r_n} - f \|^2_{L^2(\nu)}. \quad (3.2)$$

Here we use the triangle inequality and the $T$-invariance of $\nu$. If $f$ is also Lipschitz continuous, then, by Lemma 3.1,

$$\| f \circ T^{r_n} - f \|^2_{L^2(\nu)} \ll \int_{T \times \Gamma \setminus G} d(T^{r_n}(t, \Gamma g), (t, \Gamma g))^2 \, d\nu(t, \Gamma g) \ll r_n^{-\varepsilon/100}. \quad (3.3)$$
From (3.2) and (3.3),
\[
\lim_{n \to \infty} \sum_{|k| \leq T^{kn}} \| f \circ T^{krn} - f \|_{L^2(\nu)}^2 = 0
\]
for every \( \nu \in M(\T \times \Gamma \setminus G, T) \), which satisfies the condition of Lemma 3.2 and Theorem 1.4 is proved.

4. Proof of Lemma 3.1

In this section, we prove Lemma 3.1 and hence finish the proof of our main result. We assume that \( \alpha \) is irrational. Let
\[
\alpha = [0; a_1, a_2, \ldots, a_k, \ldots] = \frac{1}{a_1 + \frac{1}{a_2 + \frac{1}{a_3 + \ldots}}}
\]
be the continued fraction expansion of \( \alpha \). Let \( p_k/q_k = [0; a_1, a_2, \ldots, a_k] \) be the \( k \)th convergent of \( \alpha \). Then we have the following well-known properties of \( p_k/q_k \).

**Lemma 4.1.** Let \( \alpha \in [0, 1) \) be an irrational number and \( p_k/q_k \) the \( k \)th convergent of \( \alpha \). Then:

(i) \( p_0 = 0, p_1 = 1 \) and \( p_{k+2} = a_{k+2}p_{k+1} + p_k \) for any \( k \geq 0 \); \( q_0 = 1, q_1 = a_1 \) and \( q_{k+2} = a_{k+2}q_{k+1} + q_k \) for all \( k \geq 0 \);

(ii) \( 1/(q_k + q_{k+1}) < \| q_k \alpha \| < 1/q_k \) for any \( k \geq 1 \);

(iii) if \( 0 < q < q_{k+1} \), then \( \| q_k \alpha \| \leq \| q \alpha \| \).

We also need the following two lemmas.

**Lemma 4.2.** For \( \alpha \in \R \setminus Q \) and \( k \geq 2 \),
\[
\sum_{0 < |q| < q_k} \frac{1}{\| q \alpha \|^2} \leq q_k^2 \quad \text{and} \quad \sum_{0 < |q| < q_k} \frac{1}{\| q \alpha \|} \leq q_k \log q_k.
\]

**Proof.** The first part is from [1, Lemma 2] and we can similarly get the second.

**Lemma 4.3.** For \( \alpha \in \R \setminus Q \), \( k \geq 1 \) and \( 1 \leq c \leq q_k \),
\[
\sum_{q_k \leq |q| < q_{k+1}} \frac{1}{q^2} \min \left\{ \frac{1}{\| q \alpha \|^2}, c^2 \right\} \ll \frac{c}{q_k} \quad \text{and} \quad \sum_{q_k \leq |q| < q_{k+1}} \frac{1}{q^{2+\varepsilon}} \min \left\{ \frac{1}{\| q \alpha \|}, c \right\} \ll \frac{c}{q_k^{1+\varepsilon}}.
\]

**Proof.** The first part is from [1, Lemma 3] and we can similarly get the second.

Now we are ready to give the proof of Lemma 3.1.
PROOF OF LEMMA 3.1. Let \( \Phi(n, t) = \sum_{l=0}^{n-1} \phi(l \alpha + t), \ \Psi(n, t) = \sum_{l=0}^{n-1} \psi(l \alpha + t) \) and \( H(n, t) = \sum_{l=0}^{n-1} \phi^2(l \alpha + t) \), then
\[
T^n(t, \Gamma g) = \begin{pmatrix} 1 & \Phi(n, t) & \Psi(n, t) + \frac{1}{2} \Phi^2(n, t) - \frac{1}{2} H(n, t) \\ 0 & 1 & \Phi(n, t) \\ 0 & 0 & 1 \end{pmatrix}.
\]

We consider
\[
d((t, \Gamma g), T^n(t, \Gamma g))^2 = ||n\alpha||^2 + d_G^2 \begin{pmatrix} \Gamma g, \Gamma g \begin{pmatrix} 1 & \Phi(n, t) & \Psi(n, t) + \frac{1}{2} \Phi^2(n, t) - \frac{1}{2} H(n, t) \\ 0 & 1 & \Phi(n, t) \\ 0 & 0 & 1 \end{pmatrix} \\ \end{pmatrix} \]
\[
\leq ||n\alpha||^2 + d_G^2 \begin{pmatrix} 1 & \Phi(n, t) & \Psi(n, t) + \frac{1}{2} \Phi^2(n, t) - \frac{1}{2} H(n, t) \\ 0 & 1 & \Phi(n, t) \\ 0 & 0 & 1 \end{pmatrix} \]
\[
\leq ||n\alpha||^2 + \left| \begin{pmatrix} 1 & \Phi(n, t) & \Psi(n, t) + \frac{1}{2} \Phi^2(n, t) - \frac{1}{2} H(n, t) \\ 0 & 1 & \Phi(n, t) \\ 0 & 0 & 1 \end{pmatrix} \right|^2 \]
\[
\ll ||n\alpha||^2 + |\Phi(n, t)|^2 + |\Psi(n, t)|^2 + |\Phi(n, t)|^4 + |H(n, t)|^2.
\]

Therefore,
\[
\int_{T \times \Gamma G} d(T^n(t, \Gamma g), (t, \Gamma g))^2 dv(t, \Gamma g)
\ll ||n\alpha||^2 + \int_{T \times \Gamma G} (|\Phi(n, t)|^2 + |\Psi(n, t)|^2 + |\Phi(n, t)|^4 + |H(n, t)|^2) dv(t, \Gamma g).
\]

So we should choose a sequence \( \{r_n\}_{n \geq 1} \) of \( \mathbb{N} \) such that \( ||r_n \alpha||^2, \int_{T \times \Gamma G} |\Phi(r_n, t)|^2 dv(t, \Gamma g), \int_{T \times \Gamma G} |\Psi(r_n, t)|^2 dv(t, \Gamma g), \int_{T \times \Gamma G} |\Phi(r_n, t)|^4 dv(t, \Gamma g) \) and \( \int_{T \times \Gamma G} |H(r_n, t)|^2 dv(t, \Gamma g) \) are bounded. The following argument is similar to [1, 8].

First, we consider \( \int_{T \times \Gamma G} |\Phi(n, t)|^2 dv(t, \Gamma g) \). This integral is only dependent on the first coordinate \( t \). With the projection map \( \pi(t, \Gamma g) = t \), we can rewrite the integral as
\[
\int_{T \times \Gamma G} |\Phi(n, \pi(t, \Gamma g))|^2 dv(t, \Gamma g) = \int_T |\Phi(n, t)|^2 d(\pi_\nu)(t),
\]
where the Borel probability measure \( \pi_\nu \) is the Lebesgue measure on \( T \) since \( \alpha \) is irrational. Since \( \phi(t) = \sum_{m \neq 0} \hat{\phi}(m)e(mt) \), we have
\[
\Phi(n, t) = \sum_{l=0}^{n-1} \phi(l \alpha + t) = \sum_{l=0}^{n-1} \sum_{m \neq 0} \hat{\phi}(m)e(m(l \alpha + t)) = \sum_{m \neq 0} \hat{\phi}(m)e(mt) \frac{1 - e(mn \alpha)}{1 - e(m \alpha)}.
\]
By substituting this into (4.1), we get
\[ \int_{\mathbb{T}} \left| \Phi(n, t) \right|^2 dt = \int_{\mathbb{T}} \left| \sum_{m \neq 0} \phi(m) e(mt) \frac{1 - e(m \alpha)}{1 - e(m \alpha)} \right|^2 dt = \sum_{m \neq 0} \left| \phi(m) \right|^2 \left| \frac{1 - e(m \alpha)}{1 - e(m \alpha)} \right|^2. \]  
(4.2)

Now we can choose the desired sequence \( \{r_n\}_{n \geq 1} \) of \( \mathbb{N} \). Temporarily choose \( r_n = q_n \), where \( q_n \) is defined in Lemma 4.1. Then we break the sum in (4.2) into two sums according to \( 0 < |m| < q_n \) and \( |m| \geq q_n \).

For the second sum, we use the fact that \( \hat{\phi}(m) \ll m^{-1-\varepsilon}(m \neq 0) \). (We only need \( \phi \) to be \( C^{1+\varepsilon} \)-smooth here, but the stronger assumption that \( \phi \) is \( C^{2+\varepsilon} \)-smooth is needed later.) We also have \( |1 - e(m \alpha)| \asymp ||m\alpha||, |1 - e(m \alpha)| \leq 2 \) and the trivial bound \( |(1 - e(m \alpha))/(1 - e(m \alpha))| \leq n \).

Then
\[ \sum_{|m| \geq q_n} \left| \phi(m) \right|^2 \left| \frac{1 - e(m \alpha)}{1 - e(m \alpha)} \right|^2 \ll \sum_{|m| \geq q_n} \frac{1}{|m|^{2+2\varepsilon}} \min\left( \frac{1}{||m\alpha||^2}, r_n^2 \right) \]
\[ < q_n^{-2\varepsilon} \sum_{k=n}^{\infty} \sum_{q_k \leq |m| < q_{k+1}} \frac{1}{m^2} \min\left( \frac{1}{||m\alpha||^2}, q_n^2 \right) \]
\[ \ll q_n^{-2\varepsilon} \sum_{k=n}^{\infty} \frac{q_n}{q_k} \ll r_n^{-2\varepsilon} \]  
(by Lemma 4.3).  
(4.3)

For the first sum, we use the fact that \( |1 - e(m \alpha)| \asymp ||m\alpha|| \) and \( |1 - e(m \alpha)| \asymp ||mq_n \alpha|| \leq m^2 ||q_n \alpha||^2 < m^2 q_{n+1}^{-2} \). Then
\[ \sum_{0 < |m| < q_n} \left| \phi(m) \right|^2 \left| \frac{1 - e(m \alpha)}{1 - e(m \alpha)} \right|^2 \ll \frac{1}{q_n^{2\varepsilon}} \sum_{0 < |m| < q_n} \frac{1}{|m|^{2\varepsilon} ||m\alpha||^2}. \]  
(4.4)

For this sum, we consider the following two cases:

**Case A.** There is a subsequence \( \{q_{b_n}\}_{n \geq 1} \) of \( \{q_n\}_{n \geq 1} \) such that \( q_{b_n+1} \geq q_{b_n}^2 \) for all \( n \geq 1 \).

In this case, we replace \( \{r_n\}_{n \geq 1} \) by \( \{q_{b_n}\}_{n \geq 1} \) and notice that the estimate (4.3) still holds for this new \( \{r_n\}_{n \geq 1} \). By Lemma 4.2,
\[ \frac{1}{q_n^{2\varepsilon}} \sum_{0 < |m| < q_n} \frac{1}{|m|^{2\varepsilon} ||m\alpha||^2} \leq \frac{1}{q_n^{2\varepsilon}} \sum_{0 < |m| < q_{b_n}} \frac{1}{|m|^{2\varepsilon} ||m\alpha||^2} \ll q_n^{-2\varepsilon} \ll r_n^{-2\varepsilon}. \]

**Case B.** For all sufficiently large \( n \), we have \( q_{n+1} < q_n^2 \).

In this case, for any \( 0 < k < n \), Lemma 4.2 gives
\[ \sum_{0 < |m| < q_n} \frac{1}{|m|^{2\varepsilon} ||m\alpha||^2} \ll \sum_{0 < |m| < q_k} \frac{1}{|m|^{2\varepsilon} ||m\alpha||^2} + \sum_{q_k \leq |m| < q_n} \frac{1}{|m|^{2\varepsilon} ||m\alpha||^2} \ll q_k^{-2\varepsilon} q_n^{-2\varepsilon}. \]
(4.5)
We choose $k$ with $0 < k < n$ such that $q_k \in [q_n^{1/4}, q_n^{1/2}]$, which is possible since for all sufficiently large $n$, we have $q_{n+1} < q_n^2$. Then (4.4) and (4.5) give

$$\frac{1}{q_{n+1}^2} \sum_{0<|m|<q_n} \frac{1}{|m|^{2\epsilon}} \frac{1}{\|m\alpha\|^2} = \frac{1}{q_{n+1}^2} (q_k^2 + q_k^{2\epsilon} q_n^2) \leq \frac{1}{q_{n+1}^2} (q_n + q_n^{2-\epsilon/2}) \ll q_n^{-\epsilon/2} \leq r_n^{-\epsilon/2}.$$

This proves that $\int_{G \setminus \Gamma} \Phi(r_n, t)^2 \, dv(t, \Gamma g) \ll r_n^{-\lambda}$ for some $\lambda > \epsilon/100$. By the same argument, we see that with this choice of $\{r_n\}_{n \geq 1}$, the same estimate also holds for the integrals of $|\Psi(r_n, t)|^2$.

For the first term $\|na\|$, we have

$$\|r_n a\|^2 < q_{b_{n+1}}^2 < q_{b_n}^2 \leq r_n^{−\lambda},$$

where $b_n$ is the sequence of indices with $q_{b_n+1} \geq q_{b_n}^7$ in Case A and $b_n = n$ in Case B.

The estimates for integrals of $|H(r_n, t)|^2$ and $|\Phi(r_n, t)|^4$ are similar, so we only state the differences in the proof.

For the integral of $|H(r_n, t)|^2$, we need to slightly modify the choice of $r_n$ since now the zeroth Fourier coefficient of $\eta(t) := \varphi^2(t)$ does not vanish in general. Therefore, we should consider the extra term $\|r_n \hat{\eta}(0)\|$. To overcome this difficulty, we follow the idea of de Faveri [1]. By Dirichlet’s approximation theorem, for any $n \geq 1$, we can find some $s_n \leq q_n^0$ such that $\|s_n q_n \hat{\eta}(0)\| < q_n^{-\delta}$ where $0 < \delta < \epsilon/10$. Then we can slightly modify the definition of $r_n$ by setting $r_n = s_n q_n$. Since $s_n$ is quite small compared with $q_n$, the above argument for this new choice of $r_n$ is still valid.

For the integral of $|\Phi(r_n, t)|^4$, we cannot use the Parseval identity, so we estimate it pointwise. At this point, we need the stronger smoothness for $\phi$ to obtain the desired upper bound for the Fourier coefficients. We consider the same two cases.

**Case A.** In this case, we have $q_{n+1} < q_n^2$ for all sufficiently large $n$ and $r_n = s_n q_n$. Since $\phi$ is $C^{2+\epsilon}$-smooth, we have $\hat{\phi}(m) \ll |m|^{−2−\epsilon}$ for $m \neq 0$. Thus,

$$\Phi(r_n, t) = \sum_{m \neq 0} |\hat{\phi}(m)| \frac{1−e(m r_n a)}{1−e(m a)} \ll \left( \sum_{0<|m|<q_n} + \sum_{|m|\geq q_n} \right) \frac{1}{|m|^{2+\epsilon}} \min \left\{ r_n, \frac{1}{\|ma\|} \right\}.$$ 

For $|m| \geq q_n$, by Lemma 4.3,

$$\sum_{|m|\geq q_n} \frac{1}{|m|^{2+\epsilon}} \min \left\{ r_n, \frac{1}{\|ma\|} \right\} = \sum_{k=n}^{\infty} \frac{r_n}{q_k^{1+\delta}} \leq \sum_{k=n}^{\infty} \frac{q_n^{1+\delta}}{q_k^{1+\epsilon}} \leq q_n^{\delta-\epsilon} \sum_{k=n}^{\infty} \frac{q_n}{q_k} \leq r_n^{-(1-\gamma)(\epsilon-\delta)} \leq r_n^{-\lambda}.$$
For \(|m| < q_n|\),
\[
\sum_{0 < |m| < q_n} |\phi(m)| \frac{1 - e(mr_n \alpha)}{1 - e(m \alpha)} \ll \sum_{0 < |m| < q_n} \frac{1}{|m|^{2+\varepsilon}} \frac{s_n|m||q_n \alpha|}{||m\alpha||}
\]
\[
\ll \frac{q_n^\delta}{q_{n+1}} \sum_{0 < |m| < q_n} \frac{1}{|m|^{1+\varepsilon}} \frac{1}{||m\alpha||}
\]
Now since \(q_{n+1} < q_n^2\), there exists some \(q_k \in [q_n^{1/4}, q_n^{1/2}]\). So
\[
\frac{q_n^\delta}{q_{n+1}} \sum_{0 < |m| < q_n} \frac{1}{|m|^{1+\varepsilon}} \frac{1}{||m\alpha||} = \frac{q_n^\delta}{q_{n+1}} \left( \sum_{0 < |m| < 5q_k} + \sum_{q_k \leq |m| < q_n} \right) \frac{1}{|m|^{1+\varepsilon}} \frac{1}{||m\alpha||}
\]
\[
\ll \frac{q_n^\delta}{q_n}(q_k \log q_k + q_k^{-1-\varepsilon}q_n \log q_n) \quad \text{(by Lemma 4.2)}
\]
\[
\ll q_n^{\delta-1/4-\varepsilon/4+1/100} \ll r_n^{-\lambda/4}
\]
provided that \(\lambda\) is sufficiently small.

Case B. In this case, there is a subsequence \(\{q_{b_n}\}\) of \(\{q_n\}\) such that \(q_{b_n+1} \geq q_{b_n}^2\) and we choose \(r_n = s_{b_n} q_{b_n}\). Then, by the same argument as in Case A,
\[
\sum_{|m| \geq q_{b_n}} |\phi(m)|^2 \left| \frac{1 - e(mr_n \alpha)}{1 - e(m \alpha)} \right|^2 \ll q_{b_n}^{\delta - \varepsilon} \ll r_n^{-\lambda}
\]
if \(\lambda\) is sufficiently small.

Finally, for \(0 < |m| < q_{b_n}\),
\[
\sum_{0 < |m| < q_{b_n}} |\phi(m)| \frac{1 - e(mr_n \alpha)}{1 - e(m \alpha)} \ll \sum_{0 < |m| < q_{b_n}} \frac{1}{|m|^{2+\varepsilon}} \frac{s_{b_n}|m||q_{b_n} \alpha|}{||m\alpha||}
\]
\[
\ll \frac{q_{b_n}^\delta}{q_{b_n+1}} \sum_{0 < |m| < q_{b_n}} \frac{1}{|m|^{1+\varepsilon}} \frac{1}{||m\alpha||}
\]
\[
\ll \frac{q_{b_n}^\delta}{q_{b_n}^2} q_{b_n} \log q_{b_n} \ll r_n^{-\lambda} \quad \text{(by Lemma 4.2)}
\]
provided that \(\lambda\) is sufficiently small. This proves Lemma 3.1.

\[\square\]

Acknowledgement

We are grateful to the anonymous referees for the helpful corrections and suggestions.

References

[1] A. de Faveri, ‘Möbius disjointness for \(C^{1+\varepsilon}\) skew products’, \textit{Int. Math. Res. Not. IMRN} 2022(4) (2022), 2513–2531.
[2] S. Ferenczi, J. Kulaga-Przymus and M. Lemańczyk, ‘Sarnak’s conjecture: what’s new’, in: *Ergodic Theory and Dynamical Systems in Their Interactions with Arithmetics and Combinatorics*, Lecture Notes in Mathematics, 2213 (eds. S. Ferenczi, J. Kulaga-Przymus and M. Lemańczyk) (Springer, Cham, 2018), 163–235.

[3] H. Furstenberg, ‘The structure of distal flows’, *Amer. J. Math.* 85 (1963), 477–515.

[4] B. Green and T. Tao, ‘The quantitative behaviour of polynomial orbits on nilmanifolds’, *Ann. of Math. (2)* 175 (2012), 465–540.

[5] B. Green and T. Tao, ‘The Möbius function is strongly orthogonal to nilsequences’, *Ann. of Math. (2)* 175 (2012), 541–566.

[6] W. Huang, J. Liu and K. Wang, ‘Möbius disjointness for skew products on a circle and a nilmanifold’, *Discrete Contin. Dyn. Syst.* 41(8) (2021), 3531–3553.

[7] W. Huang, Z. Wang and X. Ye, ‘Measure complexity and Möbius disjointness’, *Adv. Math.* 347 (2019), 827–858.

[8] A. Kanigowski, M. Lemańczyk and M. Radziwiłł, ‘Rigidity in dynamics and Möbius disjointness’, *Fund. Math.* 255(3) (2021), 309–336.

[9] J. Kulaga-Przymus and M. Lemańczyk, ‘The Möbius function and continuous extensions of rotations’, *Monatsh. Math.* 178(4) (2015), 553–582.

[10] J. Liu and P. Sarnak, ‘The Möbius function and distal flows’, *Duke Math. J.* 164 (2015), 1353–1399.

[11] W. Parry, ‘Zero entropy of distal and related transformations’, in: *1968 Topological Dynamics Symposium* (eds. J. Auslander and W. H. Gottschalk) (Colorado State University, Fort Collins, CO, 1967), 383–389.

[12] P. Sarnak, Three Lectures on the Möbius Function, Randomness and Dynamics, IAS Lecture Notes, 2009; available online at https://www.math.ias.edu/files/wam/2011/PSMobius.pdf.

[13] P. Sarnak, ‘Möbius randomness and dynamics’, *Not. S. Afr. Math. Soc.* 43 (2012), 89–97.

[14] R. Tolimieri, ‘Analysis on the Heisenberg manifold’, *Trans. Amer. Math. Soc.* 288 (1977), 329–343.

[15] Y. Wang and W. Yao, ‘Strong orthogonality between the Möbius function and skew products’, *Monatsh. Math.* 192(3) (2020), 745–759.

[16] Z. Wang, ‘Möbius disjointness for analytic skew products’, *Invent. Math.* 209 (2017), 175–196.

XIAOGUANG HE, College of Mathematical Sciences, Sichuan University, Chengdu, Sichuan 610016, PR China
e-mail: hexiaoguangsdu@gmail.com

KE WANG, Data Science Institute, Shandong University, Jinan, Shandong 250100, PR China
e-mail: ke_wang@sdu.edu.cn