THE MÖBIUS FUNCTION AND DISTAL FLOWS

JIANYA LIU & PETER SARNAK

Abstract. We prove that the Möbius function is linearly disjoint from an analytic skew product on the 2-torus. These flows are distal and can be irregular in that their ergodic averages need not exist for all points. We also establish the linear disjointness of Möbius from various distal homogeneous flows.

1. Introduction

Let $\mathscr{X} = (T, X)$ be a flow, namely X is a compact topological space and $T : X \to X$ a continuous map. The sequence $\xi(n)$ is observed in $\mathscr{X}$ if there is an $f \in C(X)$ and an $x \in X$, such that $\xi(n) = f(T^n x)$. Let $\mu(n)$ be the Möbius function, that is $\mu(n)$ is 0 if n is not square-free, and is $(-1)^t$ if n is a product of t distinct primes. We say that μ is linearly disjoint from $\mathscr{X}$ if

$$\frac{1}{N}\sum_{n\le N}\mu(n)\xi(n) \to 0, \quad \text{as } N \to \infty, \tag{1.1}$$

for every observable ξ of $\mathscr{X}$. The Möbius Disjointness Conjecture of the second author asserts that μ is linearly disjoint from every $\mathscr{X}$ whose entropy is 0 [16], [17]. The results for $\mu(n)$ in this paper can be proved in the same way for similar multiplicative functions such as $\lambda(n) = (-1)^{\tau(n)}$ where $\tau(n)$ is the number of prime factors of n. This Conjecture has been established for many flows $\mathscr{X}$ (see [5], [14], [9], [3], [2]) however all of these flows are quasi-regular (or rigid) in the sense that the Bikrhoff averages

$$\frac{1}{N}\sum_{n\le N}\xi(n) \tag{1.2}$$

exists for every ξ observed in $\mathscr{X}$. In this paper we establish some new cases of the Disjointness Conjecture and in particular for irregular flows $\mathscr{X}$, that is ones for which (1.2) fails. These flows are complicated in terms of the behavior of their individual orbits but they are distal and of zero entropy, so that the disjointness is still expected to hold.

Date: March 21, 2013.

2000 *Mathematics Subject Classification.* 37A45, 11L03, 11N37.

Key words and phrases. The Möbius function, distal flow, affine linear map, skew product, nilmanifold.

Our first result is concerned with certain regular flows, namely affine linear maps of a compact abelian group X. Such a flow (T, X) is given by

$$T(x) = Ax + b \tag{1.3}$$

where A is an automorphism of X and $b \in X$ (see [10], [11]).

Theorem 1.1. *Let $\mathscr{X} = (T, X)$ be an affine linear flow on a compact abelian group which is of zero entropy. Then μ is linearly disjoint from $\mathscr{X}$.*

The flows in Theorem 1.1 are distal, and our main result is concerned with nonlinear distal flows on such spaces. We restrict to $X = \mathbb{T}^2$ the two dimensional torus $\mathbb{R}^2/\mathbb{Z}^2$ and consider nonlinear smooth (or even analytic) skew products as discussed in Furstenberg [6]. $T : \mathbb{T}^2 \to \mathbb{T}^2$ is given by

$$T(x, y) = (ax + \alpha, cx + dy + h(x)) \tag{1.4}$$

where $a, c, d \in \mathbb{Z}, ad = \pm 1, \alpha \in \mathbb{R}$ and h is a smooth periodic function of period 1. The affine linear part is in the form

$$\begin{bmatrix} a & 0 \\ c & d \end{bmatrix} \in GL_2(\mathbb{Z}),$$

ensuring that S has zero entropy (and it can always be brought into this form). The flow $(T, \mathbb{T}^2)$ is distal and this skew product is a basic building block (with $e(h(x))$ continuous) in Furstenberg's classification theory of minimal distal flows [7]. If α is diophantine, that is

$$\left|\alpha - \frac{a}{q}\right| \geq \frac{c}{q^m}$$

for some $c > 0, m < \infty$ and all a/q rational, then T can be conjugated by a smooth map of $\mathbb{T}^2$ to its affine linear part

$$(x, y) \mapsto (ax + \alpha, cx + dy + \beta) \tag{1.5}$$

where

$$\beta = \int_0^1 h(x)dx$$

(see [15]). Hence the disjointness of μ from $\mathscr{X} = (T, \mathbb{T}^2)$ for a T with a diophantine α, follows from Theorem 1.1. However if α is not diophantine the dynamics of the flow $(T, \mathbb{T}^2)$ can be very different from an affine linear flow. For example, as Furstenberg shows it may be irregular (i.e. the limits in (1.2) fail to exist for certain observables). Our main result is a proof that these nonlinear skew products are linearly disjoint from μ, at least if h satisfies some further technical hypothesis. Firstly we assume that h is analytic, namely that if

$$h(x) = \sum_{m \in \mathbb{Z}} \hat{h}(m) e(mx) \tag{1.6}$$

then

$$\hat{h}(m) \ll e^{-\tau|m|} \tag{1.7}$$

for some $\tau > 0$. Secondly we assume that there is $\tau_2 < \infty$ such that

$$|\hat{h}(m)| \gg e^{-\tau_2|m|}. \tag{1.8}$$

This is not a very natural condition being an artifact of our proof. However it is not too restrictive and the following applies rather generally (and most importantly there is no condition on α).

Theorem 1.2. *Let $\mathscr{X} = (T, \mathbb{T}^2)$ be of the form (1.4), with h satisfying (1.7) and (1.8). Then μ is linearly disjoint from $\mathscr{X}$.*

Theorem 1.1 deals with the affine linear distal flows on the n-torus. A different source of homogeneous distal flows are the affine linear flows on nilmanifold $X = G/\Gamma$ where G is a nilpotent Lie group and Γ a lattice in G. For $\mathscr{X} = (T, G/\Gamma)$ where $T(x) = \alpha x \Gamma$ with $\alpha \in G$, i.e. translation on G/Γ, the linear disjointness of μ and $\mathscr{X}$ is proven in [8] and [9]. Using the classification of zero entropy (equivalently distal) affine linear flows on nilmanifolds [4], and Green and Tao's results we prove

Theorem 1.3. *Let $\mathscr{X} = (T, G/\Gamma)$ where T is an affine linear map of the nilmanifold G/Γ of zero entropy. Then μ is linearly disjoint from $\mathscr{X}$.*

We end the introduction with brief outline of the paper and proofs. Theorem 1.1 with a rate of convergence is proved in §2. We first reduce to the torus case and then handle the torus case by Fourier analysis and classical results of Davenport and Hua on exponential sums concerning the Möbius function, which is stated as Lemma 2.1 in the present paper. The proof of Theorem 1.2 occupies §§3-6. The assertion of Thereom 1.2 holds for all α, and so we have to consider all diophantine possibilities of α. The case when α is rational is easy and this is done in §3. When α is irrational we have to distinguish three cases (A), (B), and (C), and the first two cases with rates of convergence are handled in §4 and §5 respectively via different analytic techniques. The most complicated case (C) is studied in §6, and the tool for this is the Bourgain-Sarnak-Ziegler finite version of the Vinogradov method (see Lemma 6.2), incorporated with various analytic methods such as Poisson's summation and stationary phase. Thus in case (C) we offer no rate. Furstenberg [6] gives examples of skew product transformations of the form (1.4) which are not regular in the sense of (1.2). Many of the flows $\mathscr{X}$ in Theorem 1.2 have this property and we show in §7 that Furstenberg's examples are smoothly conjugate to such $\mathscr{X}$'s. In particular his examples are linearly disjoint from μ. By analyzing the structure of affine linear maps of nilmanifolds, Theorem 1.3 is reduced in §8 to a recent result of Green-Tao of polynomial obits on nilmanifolds (see Lemma 8.1).

Throughout the paper there are various double exponential functions like $e(e(f(n)))$ against the Möbius function $\mu(n)$ where $e(x) = e^{2\pi i x}$ as usual, and so we have to keep track of the dependence of each parameter very carefully.

2. Theorem 1.1

2.1. **Reduction to the toral case.** We first reduce to the case that X is a torus (not necessarily connected), that is $X = \mathbb{T}^r \times C = \mathbb{R}^r/\mathbb{Z}^r \times C$ for some integer $r \geq 0$ and C is a finite (abelian) group. Since the linear combinations of characters $\psi \in \Gamma := \widehat{X}$, the (discrete) dual group of X, are dense in $C(X)$ it suffices to show that

$$\frac{1}{N}\sum_{n\leq N} \mu(n)\psi(T^n x) \to 0, \quad \text{as } N \to \infty \tag{2.1}$$

for every fixed $x \in X$ and $\psi \in \Gamma$. So fix $\psi \in \Gamma$ and let C_ψ be the smallest closed subgroup of Γ containing ψ and invariant by A. Here we are denoting by T the affine linear map $Tx = Ax + b$ of X and A acts on Γ by $A\phi(x) = \phi(Ax)$ for $\phi \in \Gamma, x \in X$. If C is a subgroup of Γ let $C^\perp$ the annihilator of C be the closed subgroup of X given by $C^\perp = \{x \in X : c(x) = 1 \text{ for all } c \in C\}$. Set X_ψ to be the compact quotient group $X/C_\psi^\perp$. By definition $\widehat{X/C_\psi} = C_\psi$. For $x \in X$ and $y \in C_\psi^\perp$,

$$T(x + y) = A(x + y) + b = Ax + b + Ay \equiv Ax + b \bmod C_\psi^\perp$$

since C_ψ is A-invariant. Hence $T(x + y) = Tx \bmod C_\psi^\perp$, that is T induces an affine linear map T_ψ of X_ψ. Put another way the flow $\mathscr{X}_\psi = (T_\psi, X_\psi)$ is a factor of $\mathscr{X} = (T, X)$. Since we are assuming that $\mathscr{X}$ has zero entropy it follows that so does $\mathscr{X}_\psi$. This in turn implies that C_ψ is finitely generated, as shown by Aoki (see [1] page 13). Being the dual of X_ψ it follows that X_ψ is isomorphic to $\mathbb{T}^r \times C$ for some $r \geq 0$ and finite C. Moreover for $n \geq 0$

$$\psi(T^n x) = \widetilde{\psi}(T_\psi^n \dot{x})$$

where in the last $\dot{x}$ is the projection of x in X_ψ and $\widetilde{\psi}$ is the character on X_ψ induced by ψ. In particular the observable $\psi(T^n x)$ on $\mathscr{X}$ is equal to $\widetilde{\psi}(T_\psi^n \dot{x})$ on $\mathscr{X}_\psi$. Thus (2.1) will follow from the linear disjointness of the Möbius function from $\mathscr{X}_\psi$. This completes the reduction to the toral case.

2.2. **Affine linear maps on a torus.** We have reduced Theorem 1.1 to the case that $\mathscr{X} = (T, X)$ with $X = \mathbb{R}^r/\mathbb{Z}^r \times C$ with C finite and T in the form (1.3) and of zero entropy. For our purpose of examining observables $\xi(n)$ in this flow, we can "linearize" the flow by doubling the number of variables. That is consider $Y = X \times X$ and the linear automorphism W given by

$$W(x_1, x_2) = (Ax_1 + x_2, x_2). \tag{2.2}$$

$\mathscr{Y} = (W, Y)$ is clearly of zero entropy since $\mathscr{X}$ is so, and the orbit $W^n(x_1, b)$ is equal to $(T^n x_1, b), n \geq 1$. Hence it suffices to prove Theorem 1.1 for such $\mathscr{Y}$'s. That is we can assume that $\mathscr{X} = (W, X)$ with $X = \mathbb{R}^m/\mathbb{Z}^m \times F$, F finite and W is a linear automorphism of X of zero entropy. Either by noting that the induced action of W on $\widehat{X}$ must preserve $1 \times F$ (since these are precisely the elements of finite order in $\widehat{X}$) or using the continuity of W to conclude that it preserves the connected component of 0 in X (i.e. $\mathbb{R}^m/\mathbb{Z}^m \times \{0\}$), we see that W takes the block triangular form

$$W(\theta, f) = (B\theta + Cf, Df) \tag{2.3}$$

where $B : \mathbb{R}^m/\mathbb{Z}^m \to \mathbb{R}^m/\mathbb{Z}^m$ is an automorphism of this (connected) torus, $C : F \to \mathbb{R}^m/\mathbb{Z}^m$ is a homomorphism and $D : F \to F$ is an automorphism of F. The automorphism B lifts to a linear automorphism $\widetilde{B}$ of $\mathbb{R}^m$ which preserves $\mathbb{Z}^m$, so that $\widetilde{B} \in GL_m(\mathbb{Z})$. Since W has zero entropy so does B and it is known that this implies that $\widetilde{B}$ is quasi-unipotent [4]. That is, for some $\nu_1 \geq 1, \widetilde{B}^{\nu_1} = U$ is unipotent, or $U = I + N_1$ with N_1 nilpotent and I the identity matrix. Also since F is finite it is clear that $D^{\nu_2} = I$ for some $\nu_2 \geq 1$. Let $\nu = \mathrm{lcm}(\nu_1, \nu_2)$. Then we have that

$$W^\nu(\theta, f) = ((I + N_1)\theta + C_1 f, f) \tag{2.4}$$

where C_1 is a morphism from F to $\mathbb{R}^m/\mathbb{Z}^m$. In particular

$$\Phi := W^\nu = I + N \tag{2.5}$$

where $N : X \to X$ satisfies $N^{k+1} \equiv 0$ for some $k \geq 0$. Thus for $q \geq 0$ an integer

$$\Phi^q = \sum_{t=0}^{q} \binom{q}{t} N^t = \sum_{t=0}^{\min(k,q)} \binom{q}{t} N^t. \tag{2.6}$$

Writing $n \geq 0$ as $n = q\nu + l$ with $0 \leq l < \nu$ we have

$$W^n = W^{q\nu + l} = \Phi^{q+l}$$

and hence if $x \in X$ and $n = q\nu + l$ then

$$W^n x = \sum_{t=0}^{\min(k,q)} \binom{q}{t} N^t W^l x = \sum_{t=0}^{\min(k,q)} \binom{q}{t} \xi_{l,t} \tag{2.7}$$

where

$$\xi_{l,t} = W^l N^t x. \tag{2.8}$$

For q varying, $q \geq k$ and $\psi \in \widehat{X}$ fixed we have

$$\begin{aligned} \psi(W^{q\nu+l}x) &= \psi\left(\sum_{t=0}^{k}\binom{q}{t}\xi_{l,t}\right) \\ &= \psi(\xi_{l,0})\psi(\xi_{l,1})^{\binom{q}{1}}\psi(\xi_{l,2})^{\binom{q}{2}}\cdots\psi(\xi_{l,k})^{\binom{q}{k}}. \end{aligned} \tag{2.9}$$

The character $\psi \in \widehat{X}$ has the form $\psi : x \mapsto e(\langle v, x\rangle)$ for some $v = (v_1, \ldots, v_m) \in \mathbb{Z}^m$ where $\langle v, x\rangle$ means the dot product in $\mathbb{R}^m$, and hence the right-hand side of (2.9) is $e(Y(q))$ where $Y(q)$ is a polynomial in q with degree $\leq k$ and coefficients depending on v and the ξ's. Changing variables from q to n by $n = \nu q + l$ with $0 \leq l \leq \nu - 1$, we see that $Y(q) = \phi(n)$ a polynomial in n with degree $\leq k$ and coefficients depending on v, ν, l and the ξ's. It follows that

$$\begin{aligned} \sum_{n\leq N}\mu(n)\psi(W^n x) &= \sum_{l=0}^{\nu-1}\sum_{\substack{n\leq N \\ n\equiv l(\bmod \nu)}}\mu(n)\psi(W^n x) \\ &= \sum_{l=0}^{\nu-1}\sum_{\substack{n\leq N \\ n\equiv l(\bmod \nu)}}\mu(n)e(\phi(n)). \end{aligned} \tag{2.10}$$

Theorem 1.1 for (W, X) now follows from the following classical result proved by Davenport [5] for ϕ linear and by Hua [12] for ϕ nonlinear. This lemma will also be used in later sections.

Lemma 2.1. *Let ν be a positive integer and $0 \leq l < \nu$. Let*

$$\phi(u) = \alpha_d u^d + \alpha_{d-1}u^{d-1} + \cdots + \alpha_1 u + \alpha_0$$

be a real polynomial of degree $d > 0$. Then, for arbitrary $A > 0$,

$$\sum_{\substack{n\leq N \\ n\equiv l(\bmod \nu)}}\mu(n)e(\phi(n)) \ll \frac{N}{\log^A N} \tag{2.11}$$

where the implied constant may depend on A and ν, but is independent of any of the coefficients $\alpha_d, \ldots, \alpha_0$.

This can be established by Vinogradov's method or its modern variants, such as Vaughan's identity or Heath-Brown's identity. The estimate (2.11) with the congruence condition $n \equiv l(\mathrm{mod}\nu)$, but with μ replaced by Λ the von Mangoldt function, was established in Hua [12], Theorem 10.

3. Theorem 1.2 with α rational

3.1. **Reduction.** Without loss of generality we may assume that $a = d = 1$ in (1.4). Thus

$$T: \ (x_1, x_2) \mapsto (x_1 + \alpha, cx_1 + x_2 + h(x_1)), \tag{3.1}$$

where $c \in \mathbb{Z}, \alpha \in \mathbb{R}$ and h is a smooth periodic function of period 1. Since the linear combinations of characters $\psi \in \widehat{\mathbb{T}}^2$ are dense in $C(\mathbb{T}^2)$, it is sufficient to show that

$$\sum_{n \le N} \mu(n)\psi(T^n x) = o(N), \quad \text{as } N \to \infty$$

for any fixed $x \in X$ and any fixed $\psi \in \widehat{\mathbb{T}}^2$. Note that any $\psi \in \widehat{\mathbb{T}}^2$ has the form $\psi : x \mapsto e(\langle b, x\rangle)$ for some $b = (b_1, b_2) \in \mathbb{Z}^2$ where $\langle b, x\rangle$ means the dot product in $\mathbb{R}^2$. Applying (3.1) repeatedly, we have $T^n : (x_1, x_2) \mapsto (y_1(n), y_2(n))$ with

$$y_1(n) = x_1 + n\alpha, \tag{3.2}$$

$$y_2(n) = c\frac{n(n-1)}{2}\alpha + cnx_1 + x_2 + \sum_{j=0}^{n-1} h(x_1 + j\alpha). \tag{3.3}$$

It follows that

$$\langle b, y(n)\rangle = b_1 y_1(n) + b_2 y_2(n) = P(n) + b_2 \sum_{j=0}^{n-1} h(x_1 + j\alpha),$$

where

$$P(n) = b_1(x_1 + n\alpha) + b_2\left(c\frac{n(n-1)}{2}\alpha + cnx_1 + x_2\right), \tag{3.4}$$

a polynomial of n with degree at most 2 and with coefficients depending on α, x_1, c, and b. Put

$$\begin{aligned} S(N) &= \sum_{n \le N} \mu(n) e(\langle b, y(n)\rangle) \\ &= \sum_{n \le N} \mu(n) e\left(P(n) + b_2 \sum_{j=0}^{n-1} h(x_1 + j\alpha)\right). \end{aligned} \tag{3.5}$$

Then the aim is to prove that

$$S(N) = o(N) \tag{3.6}$$

for any fixed $x = (x_1, x_2) \in \mathbb{T}^2$ and any fixed $b = (b_1, b_2) \in \mathbb{Z}^2$, which will be done in §§3-6. We may suppose that $b_2 \neq 0$ since otherwise (3.6) follows from Lemma 2.1 with $\nu = l = 1$ immediately.

Some of our results in §§3-6 actually hold for any smooth periodic h, not necessarily analytic. Suppose that $h : \mathbb{R} \to \mathbb{R}$ is a smooth periodic function with period 1. Then it has the Fourier expansion

$$h(x) = \sum_{m \in \mathbb{Z}} \hat{h}(m) e(mx), \tag{3.7}$$

which converges absolutely and uniformly on $\mathbb{R}$, and its coefficients $\hat{h}(m)$ satisfy

$$\hat{h}(m) \ll_A (|m| + 2)^{-A} \tag{3.8}$$

for arbitrary $A > 0$. We can transform $S(N)$ by inserting the Fourier expansion (3.7) of h. Thus,

$$\begin{aligned} \sum_{j=0}^{n-1} h(x_1 + j\alpha) &= \sum_{m \in \mathbb{Z}} \hat{h}(m) e(mx_1) \sum_{j=0}^{n-1} e(jm\alpha) \\ &= \sum_{m \in \mathbb{Z}} \hat{h}(m) e(mx_1) \frac{e(nm\alpha) - 1}{e(m\alpha) - 1} \end{aligned}$$

where we understand that

$$\frac{e(nm\alpha) - 1}{e(m\alpha) - 1} = n \quad \text{for } m\alpha \in \mathbb{Z}. \tag{3.9}$$

This can happen only when α is rational. It follows that

$$S(N) = \sum_{n \le N} \mu(n) e\left(P(n) + b_2 \sum_{m \in \mathbb{Z}} \hat{h}(m) e(mx_1) \frac{e(nm\alpha) - 1}{e(m\alpha) - 1} \right). \tag{3.10}$$

3.2. **The case of rational α.** In this section we establish (3.6) for rational α.

Proposition 3.1. *Let $S(N)$ be as in (3.5), and $h : \mathbb{R} \to \mathbb{R}$ a smooth periodic function with period 1. If $\alpha = a/q \in \mathbb{Q}$, then*

$$S(N) \ll N \log^{-A} N, \tag{3.11}$$

where $A > 0$ is arbitrary, and the implied constant depends on A only.

Proof. By (3.10) and (3.9),

$$S(N) = \sum_{n \le N} \mu(n) \left(P(n) + b_2 n \sum_{m \in \mathbb{Z}} \hat{h}(qm) e(qmx_1) \right).$$

The last series over m is absolutely convergent, and its sum is a constant β depending on x_1 as well as $\alpha = a/q$. Hence

$$S(N) = \sum_{n \le N} \mu(n) e(P(n) + b_2 \beta n),$$

from which and Lemma 2.1 the proposition follows. □

4. The continued fraction expansion of α

4.1. **The continued fraction expansion of α.** From now on we assume that α is irrational, and our argument will depend on the continued fraction expansion of α. Every real number α has its continued fraction representation

$$\alpha = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots}} \tag{4.1}$$

where $a_0 = [\alpha]$ is the integral part of α, and $a_1, a_2, \ldots$ are positive integers. The expression (4.1) is infinite since $\alpha \notin \mathbb{Q}$. We write $[a_0; a_1, a_2, \ldots]$ for the expression on the right-hand side of (4.1), which is the limit of the finite continued expressions

$$[a_0; a_1, a_2, \ldots, a_k] = a_0 + \cfrac{1}{a_1 + \frac{1}{a_2 + \cdots + \frac{1}{a_k}}} \tag{4.2}$$

as $k \to \infty$. Writing

$$\frac{l_k}{q_k} = [a_0; a_1, a_2, \ldots, a_k],$$

we have $l_0 = a_0, l_1 = a_0 a_1 + 1, q_0 = 1, q_1 = a_1$, and for $k \geq 2$,

$$l_k = a_k l_{k-1} + l_{k-2}, \quad q_k = a_k q_{k-1} + q_{k-2}.$$

Since α is irrational we have $q_{k+1} \geq q_k + 1$ for all $k \geq 1$. An induction argument gives the stronger assertion that $q_k \geq 2^{(k-1)/2}$ for all $k \geq 2$, and thus q_k increases at least like an exponential function of k. The irrationality of α also implies that, for all $k \geq 2$,

$$\frac{1}{2q_k q_{k+1}} < \left| \alpha - \frac{l_k}{q_k} \right| < \frac{1}{q_k q_{k+1}}, \tag{4.3}$$

which will be used in our later argument.

Let $\mathcal{Q}$ be the set of all q_k with $k = 0, 1, 2, \ldots$; note that $q_0 = 1$. Sometimes it is convenient to abbreviate q_k to q, and q_{k+1} to q^+. Let B be a large positive constant to be decided later. The set $\mathcal{Q}$ can be partitioned as $\mathcal{Q}^\flat \cup \mathcal{Q}^\sharp$ where

$$\mathcal{Q}^\flat = \{q \in \mathcal{Q} : q^+ \leq q^B\}, \quad \mathcal{Q}^\sharp = \{q \in \mathcal{Q} : q^+ > q^B\}. \tag{4.4}$$

Lemma 4.1. *Let $h : \mathbb{R} \to \mathbb{R}$ be a smooth periodic function of period 1. Then the following two series*

$$\sum_{q \in \mathcal{Q}} \sum_{\substack{q \leq m < q^+ \\ q \nmid m}} \frac{|\hat{h}(m)|}{\|m\alpha\|}, \quad \sum_{q \in \mathcal{Q}^\flat} \sum_{\substack{q \leq m < q^+ \\ q | m}} \frac{|\hat{h}(m)|}{\|m\alpha\|} \tag{4.5}$$

are convergent.

Proof. We first establish the convergence of the first series in (4.5). By (4.3), α can be written in the form

$$\alpha = \frac{l}{q} + \frac{\gamma}{q(q+1)}, \quad q \in \mathcal{Q},\ (l,q)=1,\ |\gamma|<1.$$

Therefore, for $m = 1, 2, \ldots, q-1$,

$$m\alpha = \frac{ml}{q} + \frac{m\gamma}{q(q+1)} = \frac{ml}{q} + \frac{\gamma'}{q+1}, \quad |\gamma'|<1,$$

and hence

$$\sum_{m=1}^{q-1} \frac{1}{\|m\alpha\|} = \sum_{m=1}^{q-1} \frac{1}{\|\frac{ml}{q} + \frac{\gamma'}{q+1}\|}.$$

We write $ml \equiv r (\mathrm{mod} q)$ with $1 \le |r| \le q/2$, so that the last denominator is

$$\ge \frac{|r|}{q} - \frac{|\gamma'|}{q+1} = \frac{(q+1)|r| - q|\gamma'|}{q(q+1)} \ge \frac{|r|}{q(q+1)},$$

and consequently

$$\sum_{m=1}^{q-1} \frac{1}{\|m\alpha\|} \ll \sum_{1 \le r \le q/2} \frac{q(q+1)}{r} \ll q(q+1)\log q.$$

It follows that, for any positive t,

$$\sum_{\substack{m \le t \\ q \nmid m}} \frac{1}{\|m\alpha\|} \ll \left(\frac{t}{q} + 1\right) q^2 \log q. \tag{4.6}$$

By (3.8) and partial integration,

$$\sum_{\substack{q \le m < q^+ \\ q \nmid m}} \frac{|\hat{h}(m)|}{\|m\alpha\|} \ll \sum_{\substack{q \le m < q^+ \\ q \nmid m}} \frac{m^{-A}}{\|m\alpha\|} \ll \int_q^\infty t^{-A} d\Bigg\{ \sum_{\substack{m \le t \\ q \nmid m}} \frac{1}{\|m\alpha\|} \Bigg\}$$
$$\ll q^2 \log q \int_q^\infty t^{-A}\left(\frac{t}{q}+1\right) dt \ll q^{-A+3} \log q,$$

and hence the first series in (4.5) is convergent.

Next we consider the second series in (4.5). Again by (4.3),

$$\frac{m}{2qq^+} < \left| m\alpha - m\frac{l}{q} \right| < \frac{m}{qq^+}.$$

Since $q|m$, we may write $m = m'q$, and hence the above becomes

$$\frac{m'}{2q^+} < \|m\alpha\| < \frac{m'}{q^+}.$$

It follows that

$$\sum_{\substack{q \le m < q^+ \\ q|m}} \frac{1}{\|m\alpha\|} \le \sum_{m' \le q^+/q} \frac{2q^+}{m'} \ll q^+ \log q^+, \tag{4.7}$$

and the last term is $\ll q^B \log(q^B)$ since $q \in \mathcal{Q}^\flat$. From this and (3.8) we deduce that

$$\sum_{\substack{q \le m < q^+ \\ q|m}} \frac{|\hat{h}(m)|}{\|m\alpha\|} \ll q^{-A+B} \log(q^B),$$

which proves that second series in (4.5) is also convergent. The lemma is proved. $\square$

4.2. **Transformation of the sum** $S(N)$**.** Lemma 4.1 can be used to understand the sum over m in (3.10); it implies that the following two series

$$\sum_{q \in \mathcal{Q}} \sum_{\substack{q \le m < q^+ \\ q \nmid m}} \hat{h}(m) e(mx_1) \frac{e(nm\alpha)}{e(m\alpha) - 1} \tag{4.8}$$

and

$$\sum_{q \in \mathcal{Q}^\flat} \sum_{\substack{q \le m < q^+ \\ q | m}} \hat{h}(m) e(mx_1) \frac{e(nm\alpha)}{e(m\alpha) - 1} \tag{4.9}$$

are absolutely convergent. Denote by $g(n\alpha + x_1)$ the sum of these two series, that is

$$g(n\alpha + x_1) = \left\{ \sum_{q \in \mathcal{Q}} \sum_{\substack{q \le m < q^+ \\ q \nmid m}} + \sum_{q \in \mathcal{Q}^\flat} \sum_{\substack{q \le m < q^+ \\ q | m}} \right\} \hat{h}(m) e(mx_1) \frac{e(nm\alpha)}{e(m\alpha) - 1},$$

where $g : \mathbb{R} \to \mathbb{R}$ is a smooth periodic functions of period 1. It follows that

$$\left\{ \sum_{q \in \mathcal{Q}} \sum_{\substack{q \le m < q^+ \\ q \nmid m}} + \sum_{q \in \mathcal{Q}^\flat} \sum_{\substack{q \le m < q^+ \\ q | m}} \right\} \hat{h}(m) e(mx_1) \frac{e(nm\alpha) - 1}{e(m\alpha) - 1} = g(n\alpha + x_1) - g(x_1).$$

Therefore the sum over m in (3.10) can be written as

$$g(x_1 + n\alpha) - g(x_1) + H(x) \tag{4.10}$$

with

$$H(x) = \sum_{q\in\mathcal{Q}^\sharp} \sum_{\substack{q\le m<q^+ \\ q|m}} \hat{h}(m)e(mx_1)\frac{e(nm\alpha)-1}{e(m\alpha)-1}. \tag{4.11}$$

Inserting these into (3.10), we have

$$S(N) = e(-b_2g(x_1)) \sum_{n\le N} \mu(n)e\{b_2g(n\alpha+x_1) + P(n) + b_2H(n)\}$$

with P as in (3.4).

In the following we shall prove that the factor $e\{b_2g(n\alpha+x_1)\}$ can be removed by Fourier analysis, and is hence harmless. Since $g:\mathbb{R}\to\mathbb{R}$ is a smooth periodic function of period 1, we have the Fourier expansion

$$e(b_2g(u)) = \sum_{m\in\mathbb{Z}} a(m)e(mu), \tag{4.12}$$

where

$$a(m) = \int_0^1 e(b_2g(u))e(-mu)du.$$

Since $g(u)$ depends on the constant B in (4.4), we see that $a(m)$ depends on b_2 and B only. The series (4.12) converges absolutely and uniformly in $u\in\mathbb{R}$, and hence

$$\begin{aligned} S(N) &\le \left|\sum_{n\le N}\mu(n)e\{P(n)+b_2H(n)\}\sum_{m\in\mathbb{Z}}a(m)e(mx_1+mn\alpha)\right| \\ &\le \sum_{m\in\mathbb{Z}}|a(m)|\left|\sum_{n\le N}\mu(n)e\{b_2H(n)+P(n)+mn\alpha\}\right| \\ &\ll \sup_{\alpha,m}\left|\sum_{n\le N}\mu(n)e\{b_2H(n)+P(n)+mn\alpha\}\right|, \end{aligned} \tag{4.13}$$

where the implied constant depends only on b_2 and the constant B in (4.4). The polynomial $P(n)+mn\alpha$ is harmless, but the complexity comes from $H(n)$ which we deal with in the following subsections.

4.3. **Theorem 1.2 with α irrational.** To estimate the right-hand side of (4.13), we rewrite the function H in (4.11) as

$$H(n) = \sum_{q\in\mathcal{Q}^\sharp} F(n;q)$$

where

$$F(n;q)=\sum_{\substack{q\le m<q^+\\ q|m}}\hat{h}(m)e(mx_1)\frac{e(nm\alpha)-1}{e(m\alpha)-1}. \tag{4.14}$$

We want to truncate $H(n)$ at Y, where Y is to be decided a little later. Application of (1.7) gives, for $|m|>Y$,

$$\hat{h}(m)e(mx_1)\frac{e(nm\alpha)-1}{e(m\alpha)-1}\ll e^{-\tau Y}N,$$

and therefore

$$H(n)=\sum_{\substack{q\in\mathcal{Q}^\sharp\\ q\le Y}}F(n;q)+O(e^{-\tau Y}N)=:F(n)+O(e^{-\tau Y}N). \tag{4.15}$$

If we set

$$Y=\frac{8}{\tau}\log N, \tag{4.16}$$

then the last O-term in (4.15) is $\ll N^{-7}$, and hence (4.13) becomes

$$S(N)\ll 1+\sup_{\alpha,m}|T(N)|, \tag{4.17}$$

where we should remember the implied constant depends only on b_2 and B, and where we have written

$$T(N)=\sum_{n\le N}\mu(n)e\{b_2F(n)+P(n)+mn\alpha\}. \tag{4.18}$$

Thus the estimation of $S(N)$ reduces to that of $T(N)$.

Further analysis on $F(n)$ is necessary. Recall that $q^+>q^B$ for any $q\in\mathcal{Q}^\sharp$. Also for any $q\in\mathcal{Q}^\sharp$, we have by (4.3) that

$$\frac{m}{2qq^+}<\left|m\alpha-\frac{ml}{q}\right|<\frac{m}{qq^+}.$$

If it happens that $q|m$, we change variables $m=qm'$ so that the above becomes

$$\frac{m'}{2q^+}<\|m'q\alpha\|<\frac{m'}{q^+}. \tag{4.19}$$

For further analysis we write

$$\mathcal{Q}^\sharp=\{m_1,m_2,\ldots\}.$$

Noting that

$$m_1<m_1^B\le m_1^+\le m_2<m_2^B\le m_2^+\le\ldots, \tag{4.20}$$

we deduce that $m_2^+ > m_2^B > (m_1^B)^B = m_1^{B^2}$, and consequently

$$m_j^+ > m_1^{B^j} \tag{4.21}$$

for all $j \geq 1$. The sequence (4.20) should also be truncated at Y. Since $m_j \to \infty$ as $j \to \infty$, there exists a positive integer J such that

$$m_J \leq Y < m_{J+1}. \tag{4.22}$$

From this and (4.21), we can bound J from above as

$$J \leq \frac{\log \frac{\log Y}{\log m_1}}{\log B} + 1 \ll \frac{\log\log\log N}{\log B} \tag{4.23}$$

where we have used the definition of Y in (4.16) and therefore the implied constant depends on τ. If we write $q = m_j$ in (4.19) and change variables as $m = m'm_j$, then

$$F_j(n) := F(n; m_j) = \sum_{1 \leq |m'| \leq M_j} \hat{h}(m_j m') e(m_j m' x_1) \frac{e(nm'm_j\alpha) - 1}{e(m'm_j\alpha) - 1} \tag{4.24}$$

where $M_j := m_j^+/m_j$ for $j = 1, \ldots, J-1$, but

$$M_J := Y/m_J. \tag{4.25}$$

In (4.24) we have

$$\frac{m'}{2m_j^+} < \|m'm_j\alpha\| < \frac{m'}{m_j^+}, \tag{4.26}$$

and if we write $\theta_j = \|m_j\alpha\|$ then the above with $m' = 1$ gives

$$\frac{1}{2m_j^+} < \theta_j < \frac{1}{m_j^+} \tag{4.27}$$

for all $j \geq 1$. Hence (4.24) can be written as

$$F_j(n) = f(n\theta_j), \tag{4.28}$$

with

$$f_j(x) = \sum_{1 \leq |m| < M_j} \hat{h}(m_j m) e(m_j m x_1) \frac{e(xm) - 1}{e(m\theta_j) - 1}, \quad x \in [\theta_j, \theta_j N]. \tag{4.29}$$

We conclude that the function $F(n)$ in (4.18) is of the form

$$F(n) = f_1(n\theta_1) + \cdots f_J(n\theta_J). \tag{4.30}$$

This is the expression from which we start to handle the factor $e(b_2 F(n))$ in (4.18).

With f_j as in (4.29) we set

$$\Phi_j = \sum_{1\le |m|<M_j} |m|^2 |\hat{h}(m_j m)|. \tag{4.31}$$

Let $C > 0$ be a large constant to be specified at the end of §6. We need to consider three possibilities separately:

(A) $m_J^+ \Phi_J \le \log^{4C} N$;
(B) $(m_J^+)^3 \ge \Phi_J N^4 \log^C N$;
(C) $m_J^+ \Phi_J > \log^{4C} N$ and $(m_J^+)^3 < \Phi_J N^4 \log^C N$.

In cases (A) and (B), the factor $e(b_2 F(n))$ will be handled by Fourier analysis and Lemma 2.1, while in case (C) by a finite version of the Vinogradov method (Bourgain-Sarnak-Ziegler [3]), as well as Poisson summation and stationary phase.

4.4. **Theorem 1.2 with α irrational: case (A).** In this subsection we prove the following proposition.

Proposition 4.2. *Let $S(N)$ be as in (3.5), and h an analytic function whose Fourier coefficients satisfy the upper bound condition (1.7). Assume condition (A). Then*

$$S(N) \ll N(\log N)^{8C+5-A}, \tag{4.32}$$

where $A > 0$ is arbitrary, and the implied constant depends on A, τ, and b_2, but uniform in all the other parameters.

We remark that the lower bound condition (1.8) is not needed in Proposition 4.2.

Proof. It suffices to bound $T(N)$ defined as in (4.18) under the condition (A). Our analysis starts from f_1. Recall that

$$f_1(x) = \sum_{1\le |m|\le M_1} \hat{h}(m_1 m) e(m_1 m x_1) \frac{e(xm)-1}{e(m\theta_1)-1}, \qquad x \in [\theta_1, \theta_1 N]. \tag{4.33}$$

It is easy to compute the first and second derivatives of f_1, that is

$$f_1'(x) = 2\pi i \sum_{1\le |m|<M_1} m\hat{h}(mm_1) e(mm_1x_1) \frac{e(xm)}{e(m\theta_1)-1}, \qquad x\in[\theta_1,\theta_1 N],$$

and

$$f_1''(x) = (2\pi i)^2 \sum_{1\le |m|<M_1} m^2\hat{h}(mm_1) e(mm_1x_1) \frac{e(xm)}{e(m\theta_1)-1}, \qquad x\in[\theta_1,\theta_1 N].$$

Trivially we have

$$f_1'(x) \ll \frac{1}{\theta_1} \sum_{1\le |m|<M_1} |\hat{h}(mm_1)| \ll \frac{\Phi_1}{\theta_1}, \quad f_1''(x) \ll \frac{\Phi_1}{\theta_1},$$

where the implied constants are absolute. Note that $e(b_2 f_1(x))$ is a smooth periodic function on $\mathbb{R}$, and hence can be expanded into Fourier series

$$e(b_2 f_1(x)) = \sum_{k\in\mathbb{Z}} a(k) e(kx), \tag{4.34}$$

where

$$a(k) = \int_0^1 e(b_2 f_1(x)) e(-kx) dx. \tag{4.35}$$

We must compute the dependence of $a(k)$ on f_1 and b_2. By partial integration we have

$$\begin{aligned} a(k) &= -\frac{1}{2\pi i k} \int_0^1 e(b_2 f_1(x)) de(-kx) \\ &= -\frac{b_2}{k} \int_0^1 e(b_2 f_1(x)) f_1'(x) e(-kx) dx \\ &= \frac{b_2}{2\pi i k^2} \int_0^1 \frac{d\{e(b_2 f_1(x)) f_1'(x)\}}{dx} e(-kx) dx. \end{aligned}$$

Since

$$\begin{aligned} \left| \frac{d\{e(b_2 f_1(x)) f_1'(x)\}}{dx} \right| &= |e(b_2 f_1(x))\{f_1''(x) + 2\pi b_2 f_1'(x) f_1'(x)\}| \\ &\le \frac{\Phi_1}{\theta_1} + 2\pi b_2 \left(\frac{\Phi_1}{\theta_1}\right)^2, \end{aligned}$$

we can bound $a(k)$ as follows

$$|a(k)| \le \left(\frac{\Phi_1}{\theta_1} + \left(\frac{\Phi_1}{\theta_1}\right)^2\right) \frac{b_2^2}{|k|^2} \tag{4.36}$$

for $k \ne 0$. Obviously for $k = 0$ we have $|a(0)| \le 1$. It follows that

$$\begin{aligned} \sum_{k\in\mathbb{Z}} |a(k)| &\le 1 + \left(\frac{\Phi_1}{\theta_1} + \left(\frac{\Phi_1}{\theta_1}\right)^2\right) \sum_{|k|\ge 1} \frac{b_2^2}{|k|^2} \\ &\le (4b_2)^2 \left(1 + \frac{\Phi_1}{\theta_1}\right)^2 \le (8b_2)^2 (1 + m_1^+ \Phi_1)^2, \end{aligned} \tag{4.37}$$

where in the last step we have applied $\sum_{|k|\ge 1} |k|^{-2} < 4$ as well as (4.27).

Now we can remove the factor $e(b_2 f_1(n\theta_1))$ from any sum of the form

$$\sum_{n\le N} \mu(n) e(b_2 f_1(n\theta_1) + G(n))$$

where $G(n)$ is a function of n. Indeed, on inserting the Fourier expansion of $e(b_2 f_1(x))$, the above sum in absolute value can be written as

$$\begin{aligned} &= \left| \sum_{n \le N} \mu(n) e(G(n)) \sum_{k_1 \in \mathbb{Z}} a(k_1) e(nk_1\theta_1) \right| \\ &\le \sum_{k_1 \in \mathbb{Z}} |a(k_1)| \left| \sum_{n \le N} \mu(n) e(nk_1\theta_1 + G(n)) \right| \\ &\le (8b_2)^2 (1 + m_1^+ \Phi_1)^2 \sup_{k_1, \theta_1} \left| \sum_{n \le N} \mu(n) e(nk_1\theta_1 + G(n)) \right|, \end{aligned}$$

by (4.37). In this way the factor $e(b_2 f_1(n\theta_1))$ has been removed. Of course, the same argument applies to $e(b_2 f_2), \ldots, e(b_2 f_J)$, and hence (4.18) becomes

$$|T(N)| \le \Sigma\Pi, \tag{4.38}$$

where

$$\Sigma = \sup \left| \sum_{n \le N} \mu(n) e\{n(k_1\theta_1 + \cdots + k_J\theta_J) + P(n) + mn\alpha\} \right| \tag{4.39}$$

with the sup taken over $\alpha, m, k_1, \ldots, k_J, \theta_1, \ldots, \theta_J$, and where

$$\Pi = (8b_2)^{2J} \prod_{j=1}^{J} (1 + m_j^+ \Phi_j)^2. \tag{4.40}$$

The sum Σ above can be estimated by Lemma 2.1,

$$\Sigma \ll N \log^{-A} N, \tag{4.41}$$

where the implied constant depends on A, but independent of all the other parameters.

To estimate Π we need to compute $m_1^+ \cdots m_{J-1}^+$. From (4.20) we deduce by induction that $(m_j^+)^{B^{J-j-1}} \le m_{J-1}^+$ for $j = 1, \ldots, J-1$, and therefore

$$m_1^+ \cdots m_{J-1}^+ \le (m_{J-1}^+)^{B^{-J+2} + B^{-J+3} + \cdots + B^0} \le (m_{J-1}^+)^2.$$

By definition there is a constant $K \ge 1$ depending on τ such that the inequality $\Phi_j \le K$ holds for all j. Hence

$$\prod_{j=1}^{J-1} (1 + m_j^+ \Phi_j)^2 \le (2K)^{J-1} (m_1^+ \cdots m_{J-1}^+)^2 \le (2K)^{J-1} (m_{J-1}^+)^4,$$

and this can be used to bound Π as follows:

$$\begin{aligned}\Pi &= (8b_2)^{2J}(1+m_J^+\Phi_J)^2\prod_{j=1}^{J-1}(1+m_j^+\Phi_j)^2\\ &\le (16b_2K)^{2J}(1+m_J^+\Phi_J)^2(m_{J-1}^+)^4.\end{aligned}$$

By (4.21) we have

$$(16b_2K)^{2J} = m_1^{\frac{2J\log(16b_2K)}{\log m_1}} \le m_1^{B^{J-1}} \le m_{J-1}^+$$

if B is sufficiently large in terms of K and b_2, that is in terms of τ and b_2. It turns out that for this purpose the choice $B = 4[\log(16b_2K)]$ is acceptable, where $[x]$ denotes the integral part of x. Note that $m_{J-1}^+ \le m_J \le Y$ with Y as in (4.16). These together with condition (A) give

$$\Pi \le (1+m_J^+\Phi_J)^2 m_J^5 \le (1+m_J^+\Phi_J)^2 Y^5 \ll (\log N)^{8C+5} \tag{4.42}$$

with the implied constant depending on τ only. This is the desired upper bound for Π.

Inserting (4.42) and (4.41) back into (4.38), we get

$$T(N) \ll N(\log N)^{8C+5-A},$$

where the implied constant depends only on A and τ. From this and (4.17) we conclude that

$$S(N) \ll 1 + N(\log N)^{8C+5-A}$$

with the implied constant depends on A, τ, and b_2, but uniform in all the other parameters. This completes the analysis of case (A). $\square$

5. Theorem 1.2 with α irrational: case (B)

In this section we handle case (B). Still, the lower bound condition (1.8) is not needed in Proposition 5.1.

Proposition 5.1. *Let $S(N)$ be as in (3.5), and h an analytic function whose Fourier coefficients satisfy the upper bound condition (1.7). Assume condition (B). Then*

$$S(N) \ll N\log^{-A} N + N(\log N)^{6-C} \tag{5.1}$$

where $A > 0$ is arbitrary and the implied constant depends on A, τ, and b_2 only.

Proof. It is sufficient to estimate $T(N)$ defined as in (4.18) under the condition (B). We can repeat the argument in case (A) but with J there replaced by $J-1$. Thus in (4.18) the factors $e(b_2f_1), e(b_2f_2), \ldots, e(b_2f_{J-1})$ can be removed by repeated application of Fourier analysis, but the factor $e(b_2 f_J)$ remains in the summation in (4.18). Hence, instead of (4.38), we have in the present situation,

$$|T(N)| \le \Sigma^* \Pi^* \tag{5.2}$$

with

$$\Sigma^* = \sup \left| \sum_{n \le N} \mu(n) e\{n(k_1\theta_1 + \cdots + k_{J-1}\theta_{J-1}) + b_2 f_J(n\theta_J) + P(n) + mn\alpha\} \right|, \tag{5.3}$$

where the sup is taken over $\alpha, m, k_1, \ldots, k_{J-1}, \theta_1, \ldots, \theta_{J-1}$. Also similar to (4.40),

$$\Pi^* = (8b_2)^{2(J-1)} \prod_{j=1}^{J-1} (1 + m_j^+ \Phi_j)^2, \tag{5.4}$$

where we note that Π^* does not have any factor involving the subscript J. Similar to (4.42), we have

$$\Pi^* \le m_J^5 \le Y^5 \tag{5.5}$$

provided that B is sufficiently large in terms of τ and b_2.

The estimation of Σ^* requires more detailed analysis. We should take advantage of the fact that now θ_J is very small. We write f_J in (4.29) in the form

$$f_J(n\theta_J) = \sum_{1 \le |m| < M_J} \hat{h}(m_J m) e(m_J m x_1) \sum_{j=0}^{n-1} e(jm\theta_J), \tag{5.6}$$

where recall that $M_J = Y/m_J$ by (4.25). Taylor's expansion gives

$$e(jm\theta_J) = \sum_{k=0}^{2} \frac{(2\pi i jm\theta_J)^k}{k!} + O(j^3 m^3 \theta_J^3),$$

and therefore

$$\sum_{j=0}^{n-1} e(jm\theta_J) = \sum_{k=0}^{2} \frac{(2\pi i m\theta_J)^k}{k!} \sum_{j=0}^{n-1} j^k + O(m^3 \theta_J^3 N^4).$$

Hence (5.6) takes the new form

$$\begin{aligned} f_J(n\theta_J) &= c_0(M_J)n + \frac{1}{2} c_1(M_J)\theta_J n(n-1) \\ &\quad + \frac{1}{6} c_2(M_J)\theta_J^2 (n-1)n(2n-1) + O\{\widetilde{c}_3(M_J)\theta_J^3 N^4\}, \end{aligned} \tag{5.7}$$

where

$$c_k(M) = \frac{(2\pi i)^k}{k!} \sum_{1 \le |m| < M} m^k \hat{h}(m_J m) e(m_J m x_1)$$

for $k=0,1,2$, while

$$\widetilde{c}_3(M)=\sum_{1\le|m|<M}|m|^3|\hat{h}(m_Jm)|.$$

Obviously $\widetilde{c}_3(M_J)\le Y\Phi_J$. Put $c_k=c_k(\infty)$ for $k=0,1,2$. Then by the upper bound condition (1.7),

$$\begin{aligned}|c_k-c_k(M_J)| &\le \sum_{|m|\ge M_J}|m|^k|\hat{h}(m_Jm)|\ll\sum_{m\ge M_J}m^ke^{-\tau m_Jm}\\ &\ll e^{-\frac{1}{2}\tau Y}\ll N^{-4},\end{aligned}$$

and therefore $c_k(M_J)=c_k+O(N^{-4})$ for $k=0,1,2$. Collecting these estimates back to (5.7), we have

$$b_2f_J(n\theta_J)=b_2Q(n)+O(|b_2|N^{-1})+O(|b_2|Y\Phi_J\theta_J^3N^4)$$

with

$$Q(n)=c_0n+\frac{1}{2}c_1\theta_Jn(n-1)+\frac{1}{6}c_2\theta_J^2(n-1)n(2n-1).$$

Inserting these back into (5.3) yields

$$\begin{aligned}\Sigma^* &\ll \sup\left|\sum_{n\le N}\mu(n)e\{n(k_1\theta_1+\cdots+k_{J-1}\theta_{J-1})+b_2Q(n)+P(n)+mn\alpha\}\right|\\ &\quad+O(|b_2|)+O(|b_2|Y\Phi_J\theta_J^3N^5),\end{aligned}$$

where the sup is taken over $\alpha,m,k_1,\ldots,k_{J-1},\theta_1,\ldots,\theta_{J-1}$.

The condition (B) is designed to control the last O-term, which is $\ll N(\log N)^{1-C}$ with the implied constant depending on τ and b_2 only. Applying Lemma 2.1 again to the above sum over n, we get

$$\Sigma^*\ll N\log^{-A}N+N(\log N)^{1-C}$$

where $A>0$ is arbitrary and the implied constant depends on A,τ, and b_2 only. The desired result now follows from this and (5.5). □

6. Theorem 1.2 with α irrational: case (C)

6.1. **The result and the idea of proof.** In this section we treat case (C) by establishing the following result.

Proposition 6.1. *Let $S(N)$ be as in (3.5), and h an analytic function whose Fourier coefficients satisfying both the upper bound condition (1.7) and the lower bound condition (1.8). Assume condition (C). Then*

$$S(N)=o(N). \tag{6.1}$$

In view of (4.17), it is sufficient to establish (6.1) for $T(N)$ with

$$T(N) = \sum_{n \le N} \mu(n) e\{b_2 F(n) + P(n) + mn\alpha\}$$

as in (4.18). Here we recall that $P(n)$ is the polynomial of degree at most 2 as in (3.4), and $F(n) = f_1(n\theta_1) + \cdots f_J(n\theta_J)$ with

$$f_j(x) = \sum_{1 \le |m| < M_j} \hat{h}(m_j m) e(m_j m x_1) \frac{e(xm) - 1}{e(m\theta_j) - 1}, \quad x \in [\theta_j, \theta_j N]$$

as in (4.30) and (4.29) respectively. The tool of our proof is the following result of Bourgain-Sarnak-Ziegler [3].

Lemma 6.2. *Let $f : \mathbb{N} \to \mathbb{C}$ with $|f| \le 1$ and let ν be a multiplicative function with $|\nu| \le 1$. Let $\tau > 0$ be a small parameter and assume that for all primes $p_1, p_2 \le e^{1/\tau}, p_1 \ne p_2$, we have that for M large enough*

$$\left| \sum_{m \le M} f(p_1 m) \overline{f(p_2 m)} \right| \le \tau M. \tag{6.2}$$

Then for N large enough

$$\left| \sum_{n \le N} \nu(n) f(n) \right| \le 2\sqrt{\tau \log \frac{1}{\tau}} N. \tag{6.3}$$

Lemma 6.2 reduces the estimation of $T(N)$ to that of

$$\widetilde{T}(N) = \sum_{n \le N} e\{b_2 F(d_1 n) - b_2 F(d_2 n) + P(d_1 n) - P(d_2 n) + d_1 nm\alpha - d_2 nm\alpha\} \tag{6.4}$$

where $d_1 \ne d_2$ are positive integers. Without loss of generality we assume henceforth that $d_1 > d_2$. Noting that

$$\begin{aligned} b_2 F(d_1 n) - b_2 F(d_2 n) &= \{b_2 f_1(d_1 n\theta_1) - b_2 f_1(d_2 n\theta_1)\} + \cdots \\ &\quad + \{b_2 f_J(d_1 n\theta_J) - b_2 f_J(d_2 n\theta_J)\}, \end{aligned}$$

we can repeat the argument in case (A) but with J there replaced by $J-1$. Thus in (6.4) the factors

$$e(b_2 f_1), e(-b_2 f_1), \ldots, e(b_2 f_{J-1}), e(-b_2 f_{J-1})$$

can be removed by repeated application of Fourier analysis, but the factor

$$e\{b_2 f_J(d_1 n\theta_J) - b_2 f_J(d_2 n\theta_J)\}$$

remains in the summation. Hence instead of (4.38) we have in the present situation

$$|\widetilde{T}(N)| \le \widetilde{\Sigma}\widetilde{\Pi} \tag{6.5}$$

with new definitions of $\widetilde{\Sigma}$ and $\widetilde{\Pi}$. In fact in the above

$$\begin{aligned}\widetilde{\Sigma} &= \sup\bigg|\sum_{n\le N} e\{n(d_1k_1\theta_1+\cdots+d_1k_{J-1}\theta_{J-1}-d_2l_1\theta_1-\cdots-d_2l_{J-1}\theta_{J-1})\\ &\qquad +b_2f_J(d_1n\theta_J)-b_2f_J(d_2n\theta_J)+P(d_1n)-P(d_2n)+(d_1-d_2)mn\alpha\}\bigg|, \qquad (6.6)\end{aligned}$$

where the sup is taken over $\alpha, m, d_1, d_2, k_1, \ldots, k_{J-1}, l_1, \ldots, l_{J-1}, \theta_1, \ldots, \theta_{J-1}$. Also similar to (4.40),

$$\widetilde{\Pi} = (8b_2)^{4(J-1)}\prod_{j=1}^{J-1}(1+m_j^+\Phi_j)^4,$$

where we note that $\widetilde{\Pi}$ does not have any factor involving the subscript J. Similar argument gives

$$\widetilde{\Pi} \le m_J^9 \le Y^9 \tag{6.7}$$

provided that B is sufficiently large in terms of τ and b_2.

To handle $\widetilde{\Sigma}$, we write $\widetilde{f}_J(x)$ for $f_J(d_1x)-f_J(d_2x)$ so that

$$\widetilde{f}_J(x) = \sum_{1\le|m|<M_J}\hat{h}(m_Jm)e(m_Jmx_1)\frac{e(d_1mx)-e(d_2mx)}{e(m\theta_J)-1}, \quad x\in[\theta_J,\theta_JN], \tag{6.8}$$

where recall that $M_J = Y/m_J$ by definition. We want to estimate $\widetilde{\Sigma}$ by Poisson's summation formula and the method of stationary phase. To this end, we need to know the derivatives of $\widetilde{f}_J(x)$. We are going to use the third derivative of $\widetilde{f}_J(x)$, which is

$$\widetilde{f}_J^{(3)}(x) = (2\pi i)^3\sum_{1\le|m|<M_J}m^3\hat{h}(mm_J)e(mm_Jx_1)\frac{d_1^3e(d_1mx)-d_2^3e(d_2mx)}{e(m\theta_J)-1};$$

the reason for using the third derivative will be explained later. Since $\theta_J \asymp \frac{1}{m_J^+}$ we have

$$|m|\theta_J \asymp \frac{|m|}{m_J^+} \le \frac{1}{m_J}$$

for $|m| < M_J$, and hence

$$e(m\theta_J)-1 = 2\pi i m\theta_J\bigg(1+O\bigg(\frac{1}{m_J}\bigg)\bigg).$$

It follows that

$$\widetilde{f}_J^{(3)}(x) = -\frac{(2\pi)^2}{\theta_J}\bigg(1+O\bigg(\frac{1}{m_J}\bigg)\bigg)\phi(x), \tag{6.9}$$

where

$$\phi(x) = \sum_{1\le|m|<M_J} m^2\hat{h}(mm_J)e(mm_Jx_1)\{d_1^3e(d_1mx) - d_2^3e(d_2mx)\}. \tag{6.10}$$

The polynomial $\phi(x)$ is too long for a stationary phase argument, however the upper and lower bound conditions (1.7) and (1.8) enable us to cut $\phi(x)$ at some fixed integer D. We will show in the following subsection that the choice

$$D = [\tau_2/\tau] + 2 \tag{6.11}$$

is acceptable, where $[x]$ denotes the integral part of x.

6.2. **The polynomials ϕ and ϕ_D.** We denote by ϕ_D the part of ϕ with $|m| \le D$, that is

$$\phi_D(x) = \sum_{1\le|m|\le D} m^2\hat{h}(mm_J)e(mm_Jx_1)\{d_1^3e(d_1mx) - d_2^3e(d_2mx)\}, \tag{6.12}$$

and we want to approximate ϕ by this ϕ_D. By the upper bound condition (1.7), the tail $\phi - \phi_D$ can be estimated as

$$\begin{aligned} \phi(x) - \phi_D(x) &\ll d_1^3 \sum_{m\ge D+1} m^2|\hat{h}(mm_J)| \\ &\ll d_1^3 \sum_{m\ge D+1} m^2 e^{-\tau mm_J} \ll d_1^3 e^{-\tau Dm_J}, \end{aligned} \tag{6.13}$$

where the implied constants depend at most on τ. Next we are going to prove that, when x is away from the zeros of $\phi_D(x)$ by a small quantity δ, $|\phi_D(x)|$ is away from 0 by some quantity depending on δ.

Lemma 6.3. *Let $P(z)$ be a complex polynomial of degree n defined by*

$$P(z) = c_0 + c_1z + \cdots + c_nz^n, \tag{6.14}$$

and let $z_1, \ldots, z_n$ be the zeros of $P(z)$. Let δ be a small real number, and around each z_j make a disc $D_j = \{z : |z - z_j| < \delta\}$ where $j = 1, \ldots, n$. Let $\mathbb{T}$ denote the unit circle. Then for any $z \in \mathbb{T}\backslash\{\cup_{j=1}^n D_j\}$ we have

$$|P(z)| \ge \left(\frac{\delta}{3}\right)^n \|P\|_2,$$

where

$$\|P\|_2 = \left(\sum_{m=0}^n |c_m|^2\right)^{\frac{1}{2}}. \tag{6.15}$$

We remark that $\mathbb{T}\backslash\{\cup_j D_j\}$ is the unit circle with some open arcs removed, and some of the removed open arcs may not contain any zero of $P(z)$. The total number of these removed open arcs is at most n.

Proof. Suppose that $|z_j| \le 2$ for $j = 1, \ldots, k$, while $|z_j| > 2$ for $j = k+1, \ldots, n$. Then we can write $P(z) = P_0(z)P_1(z)$ with

$$P_0(z) = c_n \prod_{j=0}^{k} (z - z_j), \quad P_1(z) = \prod_{j=k+1}^{n} (z - z_j).$$

First we note that a lower bound for $|P(z)|$ follows directly from the construction of $\mathbb{T} \backslash \{\cup_{j=1}^{n} D_j\}$, that is

$$|P(z)| \ge c_n \delta^k |P_1(z)|, \quad z \in \mathbb{T} \backslash \{\cup_{j=1}^{n} D_j\}. \tag{6.16}$$

Next we compute the norms of P and P_0, getting

$$\|P_0\|_2^2 = \int_0^1 |P_0(e(x))|^2 dx = \int_0^1 c_n^2 \prod_{j=0}^{k} |e(x) - z_j|^2 dx \le c_n^2 3^{2k},$$

and

$$\begin{aligned} \|P\|_2^2 &= \int_0^1 |P(e(x))|^2 dx \le \max_{z \in \mathbb{T}} |P_1(z)|^2 \int_0^1 |P_0(e(x))|^2 dx \\ &\le c_n^2 3^{2k} \max_{z \in \mathbb{T}} |P_1(z)|^2. \end{aligned}$$

The last inequality combined with (6.16) gives

$$|P(z)| \ge \left(\frac{\delta}{3}\right)^k \|P\|_2 \frac{|P_1(z)|}{\max_{z \in \mathbb{T}} |P_1(z)|}, \quad z \in \mathbb{T} \backslash \{\cup_{j=1}^{n} D_j\}. \tag{6.17}$$

Suppose $\max_{z \in \mathbb{T}} |P_1(z)|$ is achieved at $z = \zeta \in \mathbb{T}$. Then for any $z \in \mathbb{T}$ we have

$$\frac{|P_1(z)|}{\max_{z \in \mathbb{T}} |P_1(z)|} = \prod_{j=k+1}^{n} \frac{|z - z_j|}{|\zeta - z_j|} \ge \prod_{j=k+1}^{n} \frac{|z_j| - 1}{|z_j| + 1} \ge 3^{k-n}.$$

The desired result finally follows from this and (6.17). $\square$

We want to apply the above lemma to ϕ_D. Dividing ϕ_D by $e(-d_1 D x)$, we have

$$e(d_1 D x)\phi_D(x) = \phi_{D,1}(x) - \phi_{D,2}(x), \tag{6.18}$$

where, for $\ell = 1, 2$,

$$\phi_{D,\ell}(x) = d_\ell^3 \sum_{m=-D}^{D} m^2 \hat{h}(m m_J) e(m m_J x_1) e(d_\ell m x + d_1 D x). \tag{6.19}$$

Recall that we have assumed $d_1 > d_2$. The norm of $\phi_{D,\ell}$ can be computed as

$$\|\phi_{D,\ell}\|_2 = d_\ell^3 \Phi$$

with

$$\Phi = \left(\sum_{m=-D}^{D} |m|^4 |\hat{h}(mm_J)|^2 \right)^{\frac{1}{2}}, \tag{6.20}$$

and therefore, by (6.18) and the triangle inequality,

$$\begin{aligned} \|\phi_D\|_2 &= \|\phi_{D,1} - \phi_{D,2}\|_2 \geq \|\phi_{D,1}\|_2 - \|\phi_{D,2}\|_2 \\ &= (d_1^3 - d_2^3)\Phi \geq \Phi. \end{aligned} \tag{6.21}$$

If we write $z = e(x)$, then z lives on $\mathbb{T}$ and $e(d_1 Dx)\phi_D(x)$ can be written as a polynomial, say $P(z)$, in z with degree $2d_1D$. An application of Lemma 6.3 to $P(z)$ asserts that

$$|P(z)| \geq \left(\frac{\delta}{3}\right)^{2d_1D} \|P\|_2, \quad z \in \mathbb{T}\backslash\{\cup_{j=1}^n D_j\}, \tag{6.22}$$

where $\|P\|_2$ is defined as in (6.15). Obviously $\|P\|_2 = \|\phi_D\|_2$.

Under the map $x \mapsto z = e(x)$, the pre-image of $z \in \mathbb{T} \cap \{\cup_{j=1}^n D_j\}$ is a union of small intervals

$$\bigcup_{\ell \leq L} I_\ell \subset (0, 1],$$

where $L \leq \deg(P) = 2d_1D$. Note that each I_ℓ has length at most 2δ. It follows from (6.21) and (6.22) that, for $x \in (0, 1]\backslash\{\cup_{\ell \leq L} I_\ell\}$,

$$|\phi_D(x)| \geq \left(\frac{\delta}{3}\right)^{2d_1D} \Phi.$$

Obviously $\Phi \geq |\hat{h}(m_J)|$, which together with (6.13) gives

$$\frac{1}{2}|\phi_D(x)| - |\phi(x) - \phi_D(x)| \geq \frac{1}{2}\left(\frac{\delta}{3}\right)^{2d_1D} |\hat{h}(m_J)| - d_1^3 e^{-\tau D m_J}. \tag{6.23}$$

The lower bound condition (1.8) implies that $|\hat{h}(m_J)| \gg e^{-\tau_2 m_J}$, and hence the right-hand side of (6.23) is positive provided that

$$d_1^3 < \frac{1}{2}\left(\frac{\delta}{3}\right)^{2d_1D} e^{(\tau D - \tau_2)m_J}. \tag{6.24}$$

In view of (6.11) and (4.22), the exponent $(\tau D - \tau_2)m_J$ approaches infinity when $N \to \infty$. Suppose that (6.24) is satisfied. Then, for $x \in (0, 1]\backslash\{\cup_{\ell \leq L} I_\ell\}$,

$$|\phi(x)| \geq \frac{1}{2}|\phi_D(x)| + \left(\frac{1}{2}|\phi_D(x)| - |\phi(x) - \phi_D(x)|\right) \geq \frac{1}{2}\left(\frac{\delta}{3}\right)^{2d_1D} \Phi. \tag{6.25}$$

Next we formulate (6.25) in terms of the function $\phi(x\theta_J)$ with $x \in (0, \theta_J^{-1}]$. Write $x\theta_J = \xi$ and

$$J_\ell = \theta_J^{-1} I_\ell, \tag{6.26}$$

that is each J_ℓ is an amplification of I_ℓ by θ_J^{-1}. Note that the length of each J_ℓ is $\le 2\theta_J^{-1}\delta$. Hence (6.25) implies that, for $x \in (0, \theta_J^{-1}] \backslash \{\cup_{\ell \le L} J_\ell\}$,

$$|\phi(x\theta_J)| \ge \frac{1}{2}\left(\frac{\delta}{3}\right)^{2d_1 D} \Phi. \tag{6.27}$$

This will be used in the following subsection.

An upper bound of $|\phi(x\theta_J)|$ will also be necessary. Since the right-hand side of (6.23) is positive, we have

$$|\phi(x)| \le |\phi_D(x)| + |\phi(x) - \phi_D(x)| \le \frac{3}{2}|\phi_D(x)|. \tag{6.28}$$

Also, by (6.12) and Cauchy's inequality,

$$|\phi_D(x)| \le 2d_1^3 \sum_{|m| \le D} |m|^2 |\hat{h}(mm_J)| \le 2d_1^3 (2D)^{\frac{1}{2}} \Phi.$$

Combining the above two inequalities, we get

$$|\phi(x\theta_J)| \le 6d_1^3 D^{\frac{1}{2}} \Phi \tag{6.29}$$

for any real x. The above upper bound will also be used in the following subsection.

6.3. **Application of Poisson's summation and stationary phase.** In this subsection we estimate $\widetilde{\Sigma}$ in (6.6) by Poisson's summation formula and stationary phase. The following lemma of van der Corput (see for example Iwaniec and Kowalski [13], Corollary 8.19), in particular, will be applied.

Lemma 6.4. *Let $b - a \ge 1$. Let $F(x)$ be a real function on (a, b) such that*

$$\Lambda \le |F^{(3)}(x)| \le \eta\Lambda \tag{6.30}$$

for some $\Lambda > 0$ and $\eta \ge 1$. Then

$$\sum_{a<n<b} e(F(n)) \ll \eta^{\frac{1}{2}} \Lambda^{\frac{1}{6}} (b-a) + \Lambda^{-\frac{1}{6}} (b-a)^{\frac{1}{2}},$$

where the implied constant is absolute.

Proof of Proposition 6.1. The sum $\widetilde{\Sigma}$ in (6.6) can be written as

$$\widetilde{\Sigma} = \sup \left| \sum_{n \le N} e(E(n)) \right| \tag{6.31}$$

with

$$E(x) = x(d_1k_1\theta_1 + \cdots + d_1k_{J-1}\theta_{J-1} - d_2l_1\theta_1 - \cdots - d_2l_{J-1}\theta_{J-1}) + b_2\widetilde{f}_J(x\theta_J) + P(d_1x) - P(d_2x) + (d_1 - d_2)mx\alpha,$$

where the sup is taken over $\alpha, m, d_1, d_2, k_1, \dots, k_{J-1}, l_1, \dots, l_{J-1}, \theta_1, \dots, \theta_{J-1}$. If we take the third derivative of $E(x)$, then all the quadratic and linear terms in $E(x)$ will be killed, and the argument will be clearer. This is the reason for taking the third derivative of $E(x)$. Thus (6.9) and (3.4) imply that

$$E^{(3)}(x) = b_2\widetilde{f}_J^{(3)}(x\theta_J)\theta_J^3 = -(2\pi\theta_J)^2 b_2\left(1 + O\left(\frac{1}{m_J}\right)\right)\phi(x\theta_J). \tag{6.32}$$

Recall that in case (C) we have $m_J^+\Phi_J > \log^{4C} N$. We need to handle the following two possibilities separately:

(C1) $m_J^+ \le N$;
(C2) $m_J^+ > N$.

Case (C1). In this case we will first conduct our analysis on the subinterval $(0, \theta_J^{-1}] \subset [1, N]$. The set $(0, \theta_J^{-1}]\backslash\{\cup_{\ell \le L} J_\ell\}$ consists of at most $L+1$ intervals, and we suppose (a, b) is any one of them. On this interval (a, b) we apply (6.27) and (6.29) to get

$$\beta\theta_J^2\Phi \ll |E^{(3)}(x)| \ll d_1^3\theta_J^2\Phi \tag{6.33}$$

with

$$\beta = (\delta/3)^{2d_1D}, \tag{6.34}$$

where the implied constants depend on b_2 and D only. This means we can take $\Lambda = \beta\theta_J^2\Phi$ and $\eta = \beta^{-1}d_1^3$ in Lemma 6.4, which implies that

$$\sum_{n\in(a,b)} e(E(n)) \ll \beta^{-\frac{1}{3}}d_1^2(\theta_J^2\Phi)^{\frac{1}{6}}(b-a) + (\beta\theta_J^2\Phi)^{-\frac{1}{6}}(b-a)^{\frac{1}{2}} + 1,$$

where we have added a 1 on the right-hand side to cover the case $b - a < 1$. Summing over all these possible intervals $(a, b) \subset (0, \theta_J^{-1}]\backslash\{\cup_{\ell\le L} J_\ell\}$, which are at most $L+1 \le 2d_1D+1$ in number, we get

$$\sum_{n\in(0,\theta_J^{-1}]\backslash\{\cup J_\ell\}} e(E(n)) \ll \beta^{-\frac{1}{3}}d_1^3(\theta_J^2\Phi)^{\frac{1}{6}}\theta_J^{-1} + d_1(\beta\theta_J^2\Phi)^{-\frac{1}{6}}\theta_J^{-\frac{1}{2}} + d_1, \tag{6.35}$$

where the implied constants depend on b_2 and D only. The length of each interval J_ℓ is $\ll \theta_J^{-1}\delta$ by (6.26), and hence trivially

$$\sum_{n\in J_\ell} e(E(n)) \ll \theta_J^{-1}\delta.$$

The number L of these intervals J_ℓ is at most $2d_1D$, and consequently

$$\sum_{n\in\cup J_\ell} e(E(n)) \ll d_1\theta_J^{-1}\delta,$$

which together with (6.35) yields

$$\sum_{n\in(0,\theta_J^{-1}]} e(E(n)) \ll \beta^{-\frac{1}{3}}d_1^3(\theta_J^2\Phi)^{\frac{1}{6}}\theta_J^{-1} + d_1(\beta\theta_J^2\Phi)^{-\frac{1}{6}}\theta_J^{-\frac{1}{2}} + d_1 + d_1\theta_J^{-1}\delta \tag{6.36}$$

with the implied constant depending on b_2 and D only.

Now we come to the estimation of $\widetilde{\Sigma}$. Since $E(n)$ is a periodic function with period θ_J^{-1}, we cut the interval $[1,N]$ into two smaller ones $[1,\theta_J^{-1}K]\cup(\theta_J^{-1}K,N]$, where $K=[\theta_J N]$ and $[x]$ means the integral part of x. The main interval $[1,\theta_J^{-1}K]$ and the tail interval $(\theta_J^{-1}K,N]$ must be treated differently, and therefore we split (6.31) as

$$\widetilde{\Sigma} \le \widetilde{\Sigma}_0 + \widetilde{\Sigma}_1 \tag{6.37}$$

where

$$\widetilde{\Sigma}_0 = \sup\Bigg|\sum_{n\in[1,\theta_J^{-1}K]} e(E(n))\Bigg|, \quad \widetilde{\Sigma}_1 = \sup\Bigg|\sum_{n\in(\theta_J^{-1}K,N]} e(E(n))\Bigg|$$

with the sup having the same meaning as in (6.31). The main interval $[1,\theta_J^{-1}K]$ consists exactly of K periods of $E(n)$, and (6.36) can be applied to give

$$\begin{aligned}\sum_{n\in[1,\theta_J^{-1}K]} e(E(n)) &\ll K\Bigg|\sum_{n\in(0,\theta_J^{-1}]} e(E(n))\Bigg| \\ &\ll \beta^{-\frac{1}{3}}d_1^3(\theta_J^2\Phi)^{\frac{1}{6}}N + d_1(\beta\theta_J^2\Phi)^{-\frac{1}{6}}\theta_J^{\frac{1}{2}}N + d_1\theta_J N + d_1\delta N. \end{aligned}\tag{6.38}$$

The first term on the right-hand side of (6.38) is bounded from above by $\ll d_1^3\beta^{-\frac{1}{3}}\theta_J^{\frac{1}{3}}N$ with the implied constant depending on τ only, and the second by $d_1\beta^{-\frac{1}{6}}\theta_J^{\frac{1}{6}}\Phi^{-\frac{1}{6}}N$, which also dominates the third term. Therefore the third term can be erased.

We want to replace the Φ by Φ_J in the second term of (6.38). The definitions (6.20) and (4.31) trivially imply

$$\Phi^2 \le \sum_{1\le|m|<M_J} |m|^4|\hat{h}(m_J m)|^2 \le \Phi_J^2.$$

In the other direction, we have

$$\begin{aligned}\Phi_J^2 &\leq 2M_J \sum_{1\leq |m| < M_J} |m|^4 |\hat{h}(m_J m)|^2 \\ &\ll M_J \sum_{1\leq |m|\leq D} |m|^4 |\hat{h}(m_J m)|^2 = M_J \Phi^2\end{aligned}$$

by Cauchy's inequality as well as an argument similar to (6.28) to compare the above longer sum up to M_J with the shorter one up to D. Therefore we can substitute the Φ in the second term of (6.38) by Φ_J, getting

$$\widetilde{\Sigma}_0 \ll d_1^3 \beta^{-\frac{1}{3}} \frac{N}{(m_J^+)^{\frac{1}{3}}} + d_1 \beta^{-\frac{1}{6}} \frac{NY}{(m_J^+ \Phi_J)^{\frac{1}{6}}} + d_1 \delta N \tag{6.39}$$

where the implied constant depends on b_2, τ, τ_2 only. We multiply $\widetilde{\Pi}$ with $\widetilde{\Sigma}_0$, and then apply the bound (6.7) to get

$$\widetilde{\Sigma}_0 \widetilde{\Pi} \ll d_1^3 \beta^{-\frac{1}{3}} \frac{NY^9}{(m_J^+)^{\frac{1}{3}}} + d_1 \beta^{-\frac{1}{6}} \frac{NY^{10}}{(m_J^+ \Phi_J)^{\frac{1}{6}}} + d_1 m_J^9 \delta N,$$

where the implied constant depends on b_2, τ, τ_2 only. It should be remarked that to the last term on the right-hand side above, the bound $\widetilde{\Pi} \leq m_J^9$ has been used instead of the crude bound $\widetilde{\Pi} \leq Y^9$.

Now we specify

$$\delta = 3m_J^{-10} \tag{6.40}$$

so that (6.34) implies that

$$\beta^{-1} = m_J^{20 d_1 D} \leq Y^{20 d_1 D}, \tag{6.41}$$

and hence

$$\widetilde{\Sigma}_0 \widetilde{\Pi} \ll \frac{d_1^3 N Y^{7 d_1 D + 9}}{(m_J^+)^{\frac{1}{3}}} + \frac{d_1 N Y^{4 d_1 D + 10}}{(m_J^+ \Phi_J)^{\frac{1}{6}}} + \frac{d_1 N}{m_J}, \tag{6.42}$$

where the implied constant depends on b_2, τ, τ_2 only.

We must check that our choice of δ in (6.40) makes the inequality (6.24) meaningful, that is (6.24) does not confine d_1 to a finite interval. This can be seen easily from the fact that, under (6.40), the right-hand side of (6.24) equals

$$\frac{e^{(\tau D - \tau_2) m_J}}{2 m_J^{20 d_1 D}}$$

which clearly approaches infinity as $m_J \to \infty$, that is as $N \to \infty$.

For fixed d_1 we take $C = 20d_1D + 20$. Then the assumption $m_J^+ \gg m_J^+\Phi_J \geq \log^{4C} N$ implies that the right-hand side in (6.42) approaches 0 as $N \to \infty$.

$$\widetilde{\Sigma}_0\widetilde{\Pi} = o(N) \tag{6.43}$$

as $N \to \infty$. This completes our analysis concerning the main sum $\widetilde{\Sigma}_0$.

To bound the tail sum $\widetilde{\Sigma}_1$ we note that

$$\sum_{n\in(\theta_J^{-1}K,N]} e(E(n)) = \sum_{n\in(0,N-\theta_J^{-1}K]} e(E(n)) \tag{6.44}$$

by periodicity of $E(n)$. The length of last sum is $N - \theta_J^{-1}K < \theta_J^{-1}$, and therefore it is covered by case (C2). The argument in case (C2) below will give

$$\widetilde{\Sigma}_1\widetilde{\Pi} = o(N) \tag{6.45}$$

as $N \to \infty$.

We conclude from (6.5), (6.37), (6.43), and (6.45) that

$$|\widetilde{T}(N)| \leq \widetilde{\Sigma}_0\widetilde{\Pi} + \widetilde{\Sigma}_1\widetilde{\Pi} = o(N), \tag{6.46}$$

which in turn proves that $T(N) = o(N)$ by Lemma 6.2. The desired result for $S(N)$ follows from (4.17), and this finishes the analysis in case (C1).

Case (C2). This is similar to the proof of (6.43) in case (C1), and only minor modifications are necessary. In the present situation we start the analysis on $[1, N]$ directly, instead of on $(1, \theta_J]$. Thus (6.36) takes the form

$$\sum_{n\leq N} e(E(n)) \ll \beta^{-\frac{1}{3}} d_1^3 (\theta_J^2\Phi)^{\frac{1}{6}} N + d_1(\beta\theta_J^2\Phi)^{-\frac{1}{6}} N^{\frac{1}{2}} + d_1 + d_1 N\delta.$$

As before we multiply by $\widetilde{\Pi}$, apply the bound (6.7), replace Φ by Φ_J, and take δ as in (6.40). Then we have (6.41), and hence instead of (6.42) we have

$$\begin{aligned} |\widetilde{T}(N)| &\leq \widetilde{\Sigma}\widetilde{\Pi} \\ &\ll \frac{d_1^3 N Y^{7d_1D+9}}{(m_J^+)^{\frac{1}{3}}} + \frac{d_1(m_J^+)^{\frac{1}{3}} N^{\frac{1}{2}} Y^{4d_1D+10}}{\Phi_J^{\frac{1}{6}}} + \frac{d_1 N}{m_J}, \end{aligned} \tag{6.47}$$

For fixed d_1 we take $C = 20d_1D+20$ as before. The first and third terms on the right-hand side are the same as in (6.42), and can be handled in the same way. To bound the second term on the right-hand side of (6.47), we write

$$\frac{(m_J^+)^{\frac{1}{3}}}{\Phi_J^{\frac{1}{6}}} = \frac{1}{(m_J^+\Phi_J)^{\frac{1}{24}}} \frac{(m_J^+)^{\frac{3}{8}}}{\Phi_J^{\frac{1}{8}}}. \tag{6.48}$$

To the term $(m_J^+\Phi_J)^{\frac{1}{24}}$ in the denominator we apply the first assumption $m_J^+\Phi_J > \log^{4C} N$ in case (C), while to the term $(m_J^+)^{\frac{3}{8}}$ in the numerator we use the second assumption $(m_J^+)^3 < \Phi_J N^4 \log^C N$ in case (C). Thus (6.48) becomes

$$\frac{(m_J^+)^{\frac{1}{3}}}{\Phi^{\frac{1}{6}}} < \frac{1}{\log^{\frac{C}{6}} N} \frac{\Phi_J^{\frac{1}{8}} N^{\frac{1}{2}} \log^{\frac{C}{8}} N}{\Phi_J^{\frac{1}{8}}} < N^{\frac{1}{2}} \log^{-\frac{C}{24}} N.$$

This ensures that the second term on the right-hand side of (6.47) approaches 0 as $N \to \infty$. This proves that $\widetilde{T}(N) = o(N)$ and hence $T(N) = o(N)$ by Lemma 6.2 again. This completes the analysis in case (C2).

Proposition 6.1 is finally proved. □

Proof of Theorem 1.2. Theorem 1.2 follows from Propositions 3.1, 4.2, 5.1, and 6.1. □

7. Disjointness of μ from Furstenberg's system

7.1. **Furstenberg's example.** Furstenberg gave an example of smooth transformation $T : \mathbb{T}^2 \to \mathbb{T}^2$ such that the ergodic averages do not all exist. Let α be as in §4.1 such that

$$q_{k+1} \asymp e^{\tau q_k} \tag{7.1}$$

with τ as in (1.7). Define $q_{-k} = q_k$ and set

$$h(x) = \sum_{k \neq 0} \frac{e(q_k\alpha) - 1}{|k|} e(q_k x). \tag{7.2}$$

It follows from (4.3) and (7.1) that $h(x)$ is a smooth function. We also have $h(x) = g(x + \alpha) - g(x)$ where

$$g(x) = \sum_{k \neq 0} \frac{1}{|k|} e(q_k x) \tag{7.3}$$

so that $g(x) \in L^2(0, 1)$ and in particular defines and measurable function. But $g(x)$ cannot correspond to a continuous function, as shown in Furstenberg [6].

7.2. **The Möbius function is disjoint from the Furstenberg example.** It is enough to prove that a smooth conjugation of Furstenberg's dynamical system above satisfies the conditions of Theorem 1.2. To this end we introduce another function

$$H(x) = \sum_{m \in \mathbb{Z}} \hat{H}(m) e(mx), \tag{7.4}$$

where

$$\hat{H}(m) = e^{-2\tau|m|}. \tag{7.5}$$

Obviously $H(x) = G(x+\alpha) - G(x)$ where

$$G(x) = \sum_{m\in\mathbb{Z}} \hat{H}(m)\frac{e(mx)}{e(m\alpha)-1}. \tag{7.6}$$

We claim that $G(x)$ is smooth, and this can be proved by the the argument in Lemma 4.1. In fact by (4.6) for any positive t,

$$\sum_{\substack{m\le t\\ q_k\nmid m}} \frac{1}{\|m\alpha\|} \ll \left(\frac{t}{q_k}+1\right) q_k^2 \log q_k,$$

and hence partial integration yields

$$\begin{aligned}
\sum_{\substack{q_k\le m<q_{k+1}\\ q_k\nmid m}} \frac{\hat{H}(m)}{\|m\alpha\|} &\ll \int_{q_k}^{\infty} e^{-2\tau t} d\Big\{ \sum_{\substack{m\le t\\ q_k\nmid m}} \frac{1}{\|m\alpha\|}\Big\} \\
&\ll q_k^2 \log q_k \int_{q_k}^{\infty} t e^{-2\tau t} dt \ll e^{-\tau q_k}.
\end{aligned}$$

On the other hand, by (4.7),

$$\sum_{\substack{q_k\le m<q_{k+1}\\ q_k\mid m}} \frac{1}{\|m\alpha\|} \ll q_{k+1}\log q_{k+1},$$

which together with (7.1) gives

$$\sum_{\substack{q_k\le m<q_{k+1}\\ q_k\mid m}} \frac{\hat{H}(m)}{\|m\alpha\|} \ll e^{-2\tau q_k} q_{k+1}\log q_{k+1} \ll e^{-\tau q_k} q_k.$$

These prove that the series in (7.6) is absolutely convergent, and hence $G(x)$ is continuous. In the same way we can prove that $G(x)$ is even smooth.

Now we add h to H so that $h+H$ is smooth, and also

$$h(x) + H(x) = \{g(x+\alpha) + G(x+\alpha)\} - \{g(x) + G(x)\}. \tag{7.7}$$

However $g(x) + G(x)$ cannot be a continuous function, since $G(x)$ is while $g(x)$ is not.

In the following we want to check that $h(x) + H(x)$ satisfies the upper bound and lower bound conditions (1.7) and (1.8) of our Theorem 1.2. The m-th Fourier coefficient of $h + H$ is

$$\begin{cases} \hat{H}(m), & \text{if } m \ne q_k; \\ \hat{H}(m) + \frac{1-e(q_k\alpha)}{k}, & \text{if } m = q_k. \end{cases}$$

The case $m \neq q_k$ is obvious. To check the case $m = q_k$, we apply (4.3) and (7.1) to get

$$\frac{|1 - e(q_k\alpha)|}{k} \asymp \frac{1}{kq_{k+1}} \asymp \frac{1}{ke^{\tau q_k}},$$

which in combination with (7.5) yields

$$\hat{H}(q_k) + \frac{1 - e(q_k\alpha)}{k} \asymp \frac{1}{ke^{\tau q_k}}.$$

Thus the Fourier coefficients of $h + H$ satisfy (1.7) and (1.8), and therefore Theorem 1.2 states that the Möbius function is disjoint from the flow defined by $h + H$.

8. Theorem 1.3

For a review of preliminaries of nilmanifolds, the reader is referred to the Appendix §9.

8.1. **Structure of affine linear maps.** We begin with the structure of affine linear maps. By §2.4 in particular Theorem 2.12 in Dani [4], any affine linear map T of G/Γ can be written as

$$T = T_g \circ \overline{\sigma} \tag{8.1}$$

where T_g is the action of $g \in G$ on G/Γ, σ is an automorphism of G such that $\sigma(\Gamma) = \Gamma$, and $\overline{\sigma} : G/\Gamma \to G/\Gamma$ satisfies $\overline{\sigma}(x\Gamma) = \sigma(x)\Gamma$. It follows that

$$T(x\Gamma) = T_g\{\overline{\sigma}(x\Gamma)\} = g\sigma(x)\Gamma,$$

and by induction

$$T^n(x\Gamma) = g\sigma(g)\cdots\sigma^{n-1}(g)\sigma^n(x)\Gamma. \tag{8.2}$$

We remark that (8.2) itself is not enough to give a proof of Theorem 1.3, since the number of factors on the right-hand side of (8.2) depends on n.

8.2. **Application of zero entropy.** To prove Theorem 1.3, we need the fact that the flow $\mathscr{X} = (T, X)$ has zero entropy. The main reference concerning the dynamics here is Dani's review article [4], Chapter 10. In this setting the flow has zero entropy if and only if it is quasi-unipotent. So the aim is to prove Theorem 1.3 for such flows.

We need some words to clarify the definition. Let $T = T_g \circ \overline{\sigma}$ be as in (8.1). If all the eigenvalues of the differential $d\sigma : \mathfrak{g} \to \mathfrak{g}$ are of absolute value 1, then we say that T and σ are quasi-unipotent according to §2.4 in Dani [4]; this holds if and only if all the eigenvalues are roots of unity. Further, when G is simply connected, the factor of σ on $G/[G, G]$ is a linear automorphism and the proceeding condition holds if and only if all the eigenvalues of the factor are roots of unity.

Let $\mathcal{X} = \{X_1, \ldots, X_r\}$ be a basis for the Lie algebra $\mathfrak{g}$, and for $x \in G$ let $\psi_{\exp}(x) = (u_1, \ldots, u_r)$ be the coordinates of the first kind. Then $\sigma(x)$ can be computed by applying (9.1) in the Appendix as follows:

$$\begin{aligned} \sigma(x) &= \sigma\{\exp(u_1 X_1 + \cdots + u_r X_r)\} \\ &= \exp\{(d\sigma)(u_1 X_1 + \cdots + u_r X_r)\}. \end{aligned}$$

Since $d\sigma$ is quasi-unipotent, we may assume that the matrix U of $d\sigma$ under $\mathcal{X}$ is quasi-unipotent, and hence

$$(d\sigma)(u_1 X_1 + \cdots + u_r X_r) = (X_1, \ldots, X_r) U u,$$

where u denotes the transpose of the row vector $(u_1, \ldots, u_r)$. It follows that

$$(d\sigma)^n (u_1 X_1 + \cdots + u_r X_r) = (X_1, \ldots, X_r) U^n u,$$

and therefore

$$\begin{aligned} \sigma^n(x) &= \exp\{(d\sigma)^n (u_1 X_1 + \cdots + u_r X_r)\} \\ &= \exp\{(X_1, \ldots, X_r) U^n u\}. \end{aligned} \tag{8.3}$$

Since U is quasi-unipotent, U is a triangular matrix with its diagonal entries being roots of unity. It follows that there is a positive integer ν such that

$$U^\nu = I + N \tag{8.4}$$

where I is the identity matrix and N is nilpotent. From now on we let ν denote the least positive integer such that (8.4) holds. For any n, we can write $n = q\nu + l$ with $0 \le l \le \nu - 1$, and therefore we can compute U^n as

$$U^n = U^{\nu q + l} = U^l (I + N)^q = U^l \sum_{j=0}^{\min(q, r-1)} \binom{q}{j} N^j.$$

It follows that

$$U^n u = y \tag{8.5}$$

where y denotes the transpose of the row vector $(y_{n1}(q), \ldots, y_{nr}(q))$ and each $y_{nk}(q)$ is a polynomial in q with coefficients depending on U, x, ν, and l. Of course $\deg y_{nk} \le r - 1$ for all $k = 1, \ldots, r$. Inserting (8.5) back to (8.3), we have

$$\sigma^n(x) = \exp\{y_{n1}(q) X_1 + \cdots + y_{nr}(q) X_r\}, \tag{8.6}$$

or, in the notation of $\psi_{\exp}$,

$$\psi_{\exp}(\sigma^n(x)) = (y_{n1}(q), \ldots, y_{nr}(q)). \tag{8.7}$$

Similar results holds for $\psi_{\exp}(\sigma^j(g))$ with $g \in G$ and $j = 1, \ldots, n-1$, that is

$$\psi_{\exp}(\sigma^j(g)) = (y_{j1}(q), \ldots, y_{jr}(q)) \tag{8.8}$$

where each $y_{jk}(q)$ is a polynomial in q with degree $\leq r-1$ and with coefficients depending on U, g, ν, and l. In the special case $j=0$ the above just reduces to the coordinates $\psi_{\exp}(g)$ of g.

Now we apply Lemma 9.2 in the Appendix n times, so that the above analysis gives

$$\psi_{\exp}\{g\sigma(g)\cdots\sigma^{n-1}(g)\sigma^n(x)\} = (Y_1(q), \ldots, Y_r(q))$$

where $Y_1(q), \ldots, Y_r(q)$ are real polynomials in q with bounded degrees (which are actually $O_r(1)$ with the O-constant uniform in other parameters) and with their coefficients depending on U, x, g, ν, and l.

By Lemma 9.1 in the Appendix we can transform the coordinates of the first kind to those for the second kind. Apply $\psi \circ \psi_{\exp}^{-1}$ to the above equality,

$$\begin{aligned}\psi\{g\sigma(g)\cdots\sigma^{n-1}(g)\sigma^n(x)\} &= (\psi \circ \psi_{\exp}^{-1})(Y_1(q), \ldots, Y_r(q)) \\ &= (Z_1(q), \ldots, Z_r(q)),\end{aligned}$$

or

$$g\sigma(g)\cdots\sigma^{n-1}(g)\sigma^n(x) = \exp\{Z_1(q)X_1\}\cdots\exp\{Z_r(q)X_r\}, \tag{8.9}$$

where $Z_1(q), \ldots, Z_r(q)$ are real polynomials in q with bounded degrees and with their coefficients depending on U, x, g, ν, and l.

For each $j = 1, \ldots, r$ we may write

$$Z_j(q) = c_{j\ell}q^\ell + \cdots + c_{j1}q + c_{j0},$$

where $\ell = \deg Z_j$ and the coefficients c_{jk}'s are reals. Recalling that $n = q\nu + l$ with $0 \leq l \leq \nu - 1$, we may write $Z_j(q)$ as a polynomial in n as follows

$$Z_j(q) = c'_{j\ell}n^\ell + \cdots + c'_{j1}n + c'_{j0},$$

where c'_{jk}'s are real coefficients depending on U, g, x, ν, and l. It follows that

$$\exp\{Z_j(q)X_j\} = \exp(c'_{j\ell}X_j n^\ell)\cdots\exp(c'_{j1}X_j n)\exp(c'_{j0}) = b_{j\ell}^{n^\ell}\cdots b_{j1}^n b_{j0}$$

with $b_{j\ell} = \exp(c'_{j\ell}X_j)$ etc. Inserting these into (8.9), we see that

$$g\sigma(g)\cdots\sigma^{n-1}(g)\sigma^n(x) = b_1^{h_1(n)}\cdots b_k^{h_k(n)}$$

where $b_1, \ldots, b_k \in G$ and $h_1, \ldots, h_k$ are integral polynomials in n. Here it is important to note that k does not depend on n. Thus (8.2) becomes

$$\begin{aligned}T^n(x\Gamma) &= g\sigma(g)\cdots\sigma^{n-1}(g)\sigma^n(x)\Gamma \\ &= b_1^{h_1(n)}\cdots b_k^{h_k(n)}\Gamma.\end{aligned} \tag{8.10}$$

Compared with (8.2), this has the advantage that the number k of factors on the right-hand side is independent of n. This fact will be important for the following lemma to hold.

Lemma 8.1. *Let ν be a positive integer and $0 \le l < \nu$. Let G/Γ be a nilmanifold and $f : G/\Gamma \to [-1,1]$ a Lipschitz function. Let $b_1, \dots, b_k \in G$ and $h_1, \dots, h_k$ be integral polynomials in n, where k does not depend on n. Then, for any $A > 0$,*

$$\sum_{\substack{n \le N \\ n \equiv l (\bmod \nu)}} \mu(n) f\big(b_1^{h_1(n)} \cdots b_k^{h_k(n)} \Gamma\big) \ll N \log^{-A} N \tag{8.11}$$

where the implied constant depends on G, Γ, T, f, x, ν, and A.

Lemma 8.1 can be established in the same way as Theorem 1.1 in Green-Tao [9], where the case $\nu = l = 1$ is handled. Now a proof of Theorem 1.3 is immediate.

Proof of Theorem 1.3. Recall that ν is the least positive integer satisfying (8.4), that is ν is fixed. Then each $n \in \mathbb{N}$ can be written as $n = \nu q + l$ with $0 \le l \le \nu - 1$, and our original sum takes the form

$$\begin{aligned} \sum_{n \le N} \mu(n) f(T^n(x\Gamma)) &= \sum_{l=0}^{\nu-1} \sum_{\substack{n \le N \\ n \equiv l (\bmod \nu)}} \mu(n) f(T^n(x\Gamma)) \\ &= \sum_{l=0}^{\nu-1} \sum_{\substack{n \le N \\ n \equiv l (\bmod \nu)}} \mu(n) f\big(b_1^{h_1(n)} \cdots b_k^{h_k(n)} \Gamma\big) \end{aligned}$$

by (8.10). Applying Lemma 8.1 to the last sum over n, we get

$$\sum_{n \le N} \mu(n) f(T^n(x\Gamma)) \ll N \log^{-A} N$$

where the implied constant depends on G, Γ, T, f, x, ν, and A. Theorem 1.3 is proved. $\square$

9. Appendix: preliminaries on nilmanifolds

9.1. **Nilmanifolds.** Let G be a connected, simply connected nilpotent Lie group of dimension r. A *filtration* $G_\bullet$ on G is a sequence of closed connected groups

$$G = G_0 = G_1 \supset \cdots \supset G_d \supset G_{d+1} = \{\mathrm{id}_G\}$$

with the property that $[G_j, G_k] \subset G_{j+k}$ for all $j, k \ge 0$. Here $[H, K]$ denotes the commutator group of H and K. The *degree* d of $G_\bullet$ is the least integer such that $G_{d+1} = \{\mathrm{id}_G\}$. We say that G is *nilpotent* if G has a filtration. If Γ is a discrete and cocompact subgroup of G, then $G/\Gamma = \{g\Gamma : g \in G\}$ is called a *nilmanifold*. We write $r = \dim G$ and $r_j = \dim G_j$ for $j = 1, \dots, d$. If a filtration $G_\bullet$ of degree d exists then the lower central series filtration defined by

$$G = G_0 = G_1, \quad G_{j+1} = [G, G_j]$$

terminates with $G_{s+1} = \{\mathrm{id}_G\}$ for some $s \le d$. The least such s is called the *step* of the nilpotent Lie group G.

9.2. **Connections with Lie Algebra.** Let $\mathfrak{g}$ be the Lie algebra of G, and let $\exp : \mathfrak{g} \to G$ and $\log : G \to \mathfrak{g}$ be the exponential and logarithm maps, which are both diffeomorphisms. We can also define the 1-parameter subgroup $(g^t)_{t\in\mathbb{R}}$ associated to an element $g \in G$, and thus

$$\exp(X)^t = \exp(tX)$$

for all $X \in \mathfrak{g}$ and $t \in \mathbb{R}$. For an automorphism σ of G we denote by $d\sigma : \mathfrak{g} \to \mathfrak{g}$ the differential of σ. Then we have

$$\sigma(\exp(X)) = \exp\{(d\sigma)(X)\} \quad \text{for any } X \in \mathfrak{g}. \tag{9.1}$$

These maps are illustrated below:

$$\begin{array}{ccc} G & \xrightarrow{\sigma} & G \\ \exp\uparrow & & \downarrow\log \\ \mathfrak{g} & \xrightarrow[d\sigma]{} & \mathfrak{g} \end{array}$$

9.3. **Coordinates of the first and second kind.** Now we give the notion of coordinates of the first and second kinds. Let $\mathcal{X} = \{X_1, \dots, X_r\}$ be a basis for the Lie algebra $\mathfrak{g}$. If

$$g = \exp(u_1X_1 + \cdots + u_rX_r),$$

then we say that $(u_1, \dots, u_r)$ are the *coordinates of the first kind* or exponential coordinates for g relative to the basis $\mathcal{X}$. We write $(u_1, \dots, u_r) = \psi_{\exp}(g)$. If

$$g = \exp(v_1X_1) \cdots \exp(v_rX_r),$$

then we say that $(v_1, \dots, v_r)$ are the *coordinates of the second kind* for g relative to $\mathcal{X}$, and we write $(v_1, \dots, v_r) = \psi(g)$. The *height* of a reduced rational number $\frac{a}{b}$ is defined to be $\max\{|a|, |b|\}$. The basis $\mathcal{X}$ is said to be *Q-rational* if all the structure constants c_{ijk} in the relations

$$[X_i, X_j] = \sum_k c_{ijk}X_k$$

are rationals of height at most Q.

The following lemmas describes the connection between the two types of coordinate systems; they are Lemmas A.2 and A.3 in Green-Tao [8].

Lemma 9.1. *Let $\mathcal{X}$ be a basis for $\mathfrak{g}$ such that*

$$[\mathfrak{g}, X_j] \subset \mathrm{Span}(X_{j+1}, \dots, X_r) \tag{9.2}$$

for $j = 1, \dots, r-1$. Then the compositions $\psi_{\exp} \circ \psi^{-1}$ and $\psi \circ \psi_{\exp}^{-1}$ are both polynomial maps on $\mathbb{R}^r$ with bounded degree. If $\mathcal{X}$ is Q-rational then all the coefficients of these polynomials are rational of height at most Q^C for some constant $C > 0$.

Lemma 9.2. *Let $\mathcal{X}$ be a basis for $\mathfrak{g}$ satifying (9.2). Let $x, y \in G$, and suppose that $\psi(x) = (u_1, \ldots, u_r)$ and $\psi(y) = (v_1, \ldots, v_r)$. Then*

$$\psi_{\exp}(x) = (u_1, u_2 + R_1(u_1), \ldots, u_r + R_{r-1}(u_1, \ldots, u_{r-1})),$$

where each $R_j : \mathbb{R}^j \to \mathbb{R}$ is a polynomial of bounded degree. Also,

$$\psi_{\exp}(xy) = (u_1 + v_1, u_2 + v_2 + S_1(u_1, v_1), \ldots, u_r + v_r + S_{r-1}(u_1, \ldots, u_{r-1}, v_1, \ldots, v_{r-1})),$$

where each $S_j : \mathbb{R}^j \times \mathbb{R}^j \to \mathbb{R}$ is a polynomial of bounded degree.

Let $Q \geq 2$. If $\mathcal{X}$ is Q-rational then all the coefficients of the polynomials R_j, S_j are rationals of height Q^C for some constant $C > 0$.

Acknowledgements. While working on this project the first author made multiple visits to the Institute for Advanced Study and Princeton University in 2010-2013, and it is a pleasure to record his gratitude to both institutions. The first author is supported by the 973 Program, NSFC grant 11031004, and the Development Program for Changjiang Scholars and Innovative Research Team. The second author is supported by an NSF grant.

References

[1] N. Aoki, Topological entropy of distal affine transformations on compact abelian groups, J. Math. Soc. Japan 23(1971), 11-17.

[2] J. Bourgain, On the correlation of the Möbius function with random rank one systems (2011), arXiv:1112.1031.

[3] J. Bourgain, P. Sarnak, and Ziegler, Disjointness of Möbius from horocycle flows, *From Fourier analysis and number theory to radon transforms and geometry*, 67-83, Dev. Math. 28, Springer, New York, 2013.

[4] S. G. Dani, Dynamical systems on homogeneous spaces, in Chapter 10, *Dynamical Systems, Ergodic Theory and Applications*, Encyclopedia of Mathematical Sciences, Vol. 100, Springer.

[5] H. Davenport, On some infinite series involving arithmetical functions II, Quart. J. Math. 8(1937), 313-350.

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

[7] H. Furstenberg, The structure of distal flows, Amer. J. Math. 85(1963), 477-515.

[8] B. Green and T. Tao, The quatitative behaviour of polynomial obits on nilmanifolds, Ann. Math. (2), 175(2012), 465-540.

[9] B. Green and T. Tao, The Möbius function is strongly orthorgonal to nilsequences, Ann. Math. (2), 175 (2012), 541-566.

[10] F. J. Hahn, On affine transformations of compact abelian groups, Amer. J. Math. 85(1963), 428-446; Errata: Amer. J. Math. 86(1964), 463-464.

[11] H. Hoare and W. Parry, Affine transformations with quasi-discrete spectrum I, J. London Math. Soc. 41(1966), 88-96.

[12] L. K. Hua, *Additive theory of prime numbers*, AMS Translations of Mathematical Monographs, Vol. 13, Providence, R.I. 1965.

[13] H. Iwaniec and E. Kowalski, *Analytic number theory*, AMS Colloquium Publications, Vol. 53, Providence, R.I. 2004.
[14] C. Mauduit and J. Rivat, Sur un problème de Gelfond: la somme des chiffres des nombres premiers, Ann. Math. (2) 171(2010), 1591-1646.
[15] N. M. dos Santos and R. Urzúa-Luz, Minimal homeomorphisms on low-dimensional tori, Ergodic Theory Dynam. Systems 29(2009), 1515-1528.
[16] P. Sarnak, *Three lectures on the Möbius function, randomness and dynamics*, IAS Lecture Notes, 2009; http://publications.ias.edu/sites/default/files/MobiusFunctionsLectures(2).pdf.
[17] P. Sarnak, Möbius randomness and dynamics, Port Elizabeth lecture, 2011.
[18] P. Sarnak and A. Ubis, The horocycle at prime times, arXiv:1110.0777v2.

School of Mathematics, Shandong University, Jinan, Shandong 250100, China
E-mail address: `jyliu@sdu.edu.cn`

Department of Mathematics, Princeton University & Institute for Advanced Study, Princeton, NJ 08544-1000, USA
E-mail address: `sarnak@math.princeton.edu`