JOURNAL OF THE
AMERICAN MATHEMATICAL SOCIETY
Volume 30, Number 1, January 2017, Pages 67–157
http://dx.doi.org/10.1090/jams/857
Article electronically published on March 1, 2016

HIGHER ORDER FOURIER ANALYSIS OF MULTIPLICATIVE FUNCTIONS AND APPLICATIONS

NIKOS FRANTZIKINAKIS AND BERNARD HOST

1. Introduction

A function $f\colon \mathbb{N} \to \mathbb{C}$ is called *multiplicative* if

$$f(mn) = f(m)f(n) \quad \text{whenever} \quad (m,n) = 1.$$

If multiplicativity holds for every $m, n \in \mathbb{N}$, we call f *completely multiplicative.* We denote by $\mathcal{M}$ the set of multiplicative functions of modulus at most 1.

The asymptotic behavior of averages of multiplicative functions is a central topic in analytic number theory that has been studied extensively. In this article we are interested in studying the asymptotic behavior of averages of the following form:

$$\frac{1}{N^2} \sum_{1\leq m,n\leq N} \prod_{i=1}^{s} f(L_i(m,n)), \tag{1.1}$$

where $f \in \mathcal{M}$ is arbitrary and $L_i(m,n)$, $i = 1, \ldots, s$, are linear forms with integer coefficients. We are mainly motivated by applications, perhaps the most surprising one being that there is a link between the aforementioned problem and partition regularity problems of non-linear homogeneous equations in three variables; our methods enable us to address some previously intractable problems.

Two typical questions we want to answer, stated somewhat imprecisely, are as follows:

(i) Can we impose "soft" conditions on $f \in \mathcal{M}$ implying that the averages (1.1) converge to 0 as $N \to +\infty$?
(ii) Is it always possible to replace $f \in \mathcal{M}$ with a "structured" component f_{st}, such that the averages (1.1) remain unchanged, modulo a small error, for all large N?

The answer to both questions is positive in a very strong sense, the necessary condition of question (i) turns out to be extremely simple, we call it "aperiodicity," and the structured component f_{st} that works for (ii) can be taken to be approximately periodic with an approximate period independent of f and N.

Received by the editors April 21, 2014 and, in revised form, December 30, 2015 and January 25, 2016.

2010 *Mathematics Subject Classification.* Primary 11N37; Secondary 05D10, 11N60, 11B30, 37A45.

Key words and phrases. Multiplicative functions, Gowers uniformity, partition regularity, inverse theorems, Chowla conjecture.

The first author was partially supported by Marie Curie IRG 248008.

The second author was partially supported by Centro de Modelamiento Matemático, Universitad de Chile.

For $s \leq 3$ both questions can be answered by combining simple Fourier analysis tools on cyclic groups and a quantitative version of a classical result of Daboussi [11–13] which gives information on the Fourier transform of a multiplicative function. The key point is that for $s \leq 3$ the norm of the averages (1.1) can be controlled by the maximum of the Fourier coefficients of f, and the previous result can be used to give satisfactory necessary and sufficient conditions so that this maximum converges to 0 as $N \to +\infty$.

For $s \geq 4$ it is impossible to control the norm of the averages (1.1) by the maximum of the Fourier coefficients of f, and classical Fourier analytic tools do not seem to facilitate the study of these more complicated averages. To overcome this obstacle, we supplement our toolbox with some deep results from "higher order Fourier analysis"; in particular, the inverse theorem for the Gowers uniformity norms [32] and the quantitative factorization of polynomial sequences on nilmanifolds [30] play a prominent role. In an argument that spans a substantial part of this article, these tools are combined with an orthogonality criterion for multiplicative functions, and a delicate equidistribution result on nilmanifolds, in order to prove a structure theorem for multiplicative functions. This structure theorem is going to do the heavy lifting in answering questions (i), (ii), and in subsequent applications; we mention here a variant sacrificing efficiency for ease of understanding (more efficient variants and the definition of the U^s-norms appear in Sections 2.1 and 8).

Theorem 1.1 (Structure theorem for multiplicative functions I)**.** *Let $s \geq 2$ and $\varepsilon > 0$. There exist positive integers $Q := Q(s,\varepsilon)$ and $R := R(s,\varepsilon)$, such that for every sufficiently large $N \in \mathbb{N}$, depending on s and ε only, every $f \in \mathcal{M}$ admits the decomposition*

$$f(n) = f_{\mathrm{st}}(n) + f_{\mathrm{un}}(n), \qquad \text{for } \ n = 1, \ldots, N,$$

where f_{st} and f_{un} depend on N, $|f_{\mathrm{st}}| \leq 1$, and

(i) $|f_{\mathrm{st}}(n+Q) - f_{\mathrm{st}}(n)| \leq \dfrac{R}{N}$, *for* $n = 1, \ldots, N-Q$*;*

(ii) $\|f_{\mathrm{un}}\|_{U^s(\mathbb{Z}_N)} \leq \varepsilon$*.*

A distinctive feature of Theorem 1.1 is that it applies to arbitrary bounded multiplicative functions. For this reason our argument differs significantly from arguments in [29, 31, 32, 50, 51, 58–62], where pseudorandomness properties of the Möbius and other arithmetical functions are exploited. For instance, the lack of effective estimates that could be used to treat the "minor arc" part of our argument renders the method of [31] inapplicable and necessitates the introduction of new tools. These new ideas span Sections 3, 6, 7, 8 and are properly explained in the course of this article. Other important features of Theorem 1.1 are that the structured component f_{st} is always approximately periodic and that its approximate period is independent of f and N. In fact, we show that f_{st} is a convolution product of f with a kernel on $\mathbb{Z}_{\widetilde{N}}$ ($\widetilde{N} \geq N$ is a prime) that does not depend on f and the cardinality of its spectrum depends only on s and ε. All these properties turn out to be very crucial for subsequent applications. Note that for arbitrary bounded sequences, decomposition results with similar flavor have been proved in [20, 21, 28, 71, 72], but in order to work in this generality one is forced to use a structured component that does not satisfy the strong rigidity condition in (i); the best that can be said is that it is an $(s-1)$-step nilsequence of bounded complexity. This property is much weaker than (i) even when $s = 2$ and insufficient for our applications.

Similar comments apply for analogous decomposition results for infinite sequences [45] that were motivated by structural results in ergodic theory [44].

Despite its clean and succinct form, Theorem 1.1 turns out to be difficult to prove. The main ideas are sketched in Sections 2.1.3, 6.1, 8.1; furthermore, Proposition 3.4 provides a toy model of the much more complicated general case.

We remark that although explicit use of ergodic theory is not made anywhere in the proof of Theorem 1.1 and its variants, ergodic structural results and dynamical properties of sequences on nilmanifolds have guided some of our arguments.

Next, we give some representative examples of the applications that we are going to derive from variants of Theorem 1.1. Again, we sacrifice generality for ease of understanding; the precise statements of the more general results appear in the next section.

Partition regularity of quadratic equations. Since the theorems of Schur [68] and van der Waerden [73], numerous partition regularity results have been proved for linear equations, but progress has been scarce for non-linear ones, the hardest case being equations in three variables. We prove partition regularity for certain equations involving quadratic forms in three variables. For example, we show in Corollary 2.8 that for every partition of $\mathbb{N}$ into finitely many cells, there exist distinct x, y belonging to the same cell and $\lambda \in \mathbb{N}$ such that $16x^2+9y^2 = \lambda^2$. Similar results hold for the equation $x^2 - xy + y^2 = \lambda^2$ and in much greater generality (see Theorems 2.7 and 2.13). We actually prove stronger density statements from which the previous partition regularity results follow.

Uniformity of multiplicative functions. In Theorem 2.5 we show that for $s \geq 2$, for every $f \in \mathcal{M}$ we have

$$\lim_{N\to+\infty} \|f\|_{U^s(\mathbb{Z}_N)} = 0 \text{ if and only if } \lim_{N\to+\infty} \frac{1}{N}\sum_{n=1}^{N} f(an+b) = 0 \quad \text{for every } a, b \in \mathbb{N}.$$

Furthermore, using a result of Halász (see Theorem 2.3), it is easy to recast the second condition as a simple statement that is easy to verify or refute for explicit multiplicative functions (see Property (iv) of Proposition 2.4).

A generalization of a result of Daboussi. A classical result of Daboussi [11–13] states that

$$\lim_{N\to+\infty} \sup_{f\in\mathcal{M}} \Big|\frac{1}{N}\sum_{n=1}^{N} f(n)\, e^{2\pi i n\alpha}\Big| = 0 \quad \text{for every} \quad \alpha \in \mathbb{R}\setminus\mathbb{Q}. \tag{1.2}$$

Kátai [46] showed that the same thing holds if $e^{2\pi i n\alpha}$ is replaced by $e^{2\pi i p(n)}$ where $p(n) = \alpha_1 n + \cdots + \alpha_d n^d$ has at least one coefficient irrational. In Theorem 2.2 we generalize this even further to cover sequences induced by totally equidistributed polynomial sequences on nilmanifolds. One such example is the sequence $e^{2\pi i [n\sqrt{2}] n\sqrt{3}}$.

A variant of Chowla's conjecture. A classical conjecture of Chowla [9] states that if λ is the Liouville function and $P \in \mathbb{Z}[x, y]$ is a homogeneous polynomial such that $P \neq cQ^2$ for every $c \in \mathbb{Z}$, $Q \in \mathbb{Z}[x, y]$, then

$$\lim_{N\to+\infty} \frac{1}{N^2} \sum_{1\leq m,n\leq N} \lambda(P(m, n)) = 0. \tag{1.3}$$

This was established by Landau when $\deg(P) = 2$ [52] (see also [39]), by Helfgott when $\deg(P) = 3$ [40, 41], and by Green, Tao, and Ziegler when P is a product of pairwise independent linear forms [29–32]. The conjecture is also closely related to the problem of representing primes by irreducible polynomials; for relevant work see [16, 36–38]. In Theorem 2.6 we show that if

$$P(m,n) := (m^2+n^2)^r \prod_{i=1}^{s} L_i(m,n),$$

where $r \geq 0$, $s \in \mathbb{N}$, and L_i are pairwise independent linear forms with integer coefficients, and if $f \in \mathcal{M}$ is completely multiplicative and aperiodic, meaning it averages to zero on every infinite arithmetic progression, then

$$\lim_{N\to+\infty} \frac{1}{N^2} \sum_{1\leq m,n\leq N} f(P(m,n)) = 0.$$

As a consequence, for aperiodic multiplicative functions f and pairwise independent linear forms, the averages (1.1) converge to 0 as $N \to +\infty$. Note that even in the case where $r = 0$, our result is new, as it applies to arbitrary aperiodic multiplicative functions of modulus at most 1, not just the Möbius or the Liouville.

In the next section we give a more precise formulation of our main results and also define some of the concepts used throughout the article.

2. Precise statements of the main results

2.1. **Structure theorem for multiplicative functions.** Roughly speaking, our main structure theorem asserts that an arbitrary multiplicative function of modulus at most 1 can be split into two components, one that is approximately periodic, and another that behaves randomly enough to have a negligible contribution for the averages we are interested in handling. For our purposes, randomness is measured by the Gowers uniformity norms. Before proceeding to the precise statement of the structure theorem, we start with some discussion regarding the Gowers uniformity norms and the uniformity properties (or lack thereof) of multiplicative functions.

2.1.1. *Gowers uniformity norms.* For $N \in \mathbb{N}$ we let $\mathbb{Z}_N := \mathbb{Z}/N\mathbb{Z}$ and $[N] := \{1, \ldots, N\}$. These sets are often identified in the obvious way, but arithmetic operations performed on them behave differently. If x, y are considered as elements of $[N]$, expressions like $x+y$, $x-y$, …, are computed in $\mathbb{Z}$. If x, y are considered as elements of $\mathbb{Z}_N$, $x+y$, $x-y$, …, are computed modulo N and are elements of $\mathbb{Z}_N$.

If A is a finite set and $a\colon A \to \mathbb{C}$ is a function, we write

$$\mathbb{E}_{x\in A} a(x) := \frac{1}{|A|} \sum_{x\in A} a(x).$$

The same notation is used for a function of several variables. We recall the definition of the U^s-Gowers uniformity norms from [19].

Definition (Gowers norms on a cyclic group [19])**.** Let $N \in \mathbb{N}$ and $a\colon \mathbb{Z}_N \to \mathbb{C}$. For $s \in \mathbb{N}$ the *Gowers $U^s(\mathbb{Z}_N)$-norm* $\|a\|_{U^s(\mathbb{Z}_N)}$ of a is defined inductively as follows: For every $t \in \mathbb{Z}_N$ we write $a_t(n) := a(n+t)$. We let

$$\|a\|_{U^1(\mathbb{Z}_N)} := |\mathbb{E}_{n\in\mathbb{Z}_N} a(n)| = \Big(\mathbb{E}_{x,t\in\mathbb{Z}_N} a(x)\, \overline{a}(x+t)\Big)^{1/2},$$

and for every $s \geq 1$ we let

$$\|a\|_{U^{s+1}(\mathbb{Z}_N)} := \Big(\mathbb{E}_{t\in\mathbb{Z}_N} \|a \cdot \overline{a}_t\|^{2^s}_{U^s(\mathbb{Z}_N)}\Big)^{1/2^{s+1}}. \tag{2.1}$$

For example,

$$\|a\|^4_{U^2(\mathbb{Z}_N)} = \mathbb{E}_{x,t_1,t_2\in\mathbb{Z}_N} a(x)\, \overline{a}(x+t_1)\, \overline{a}(x+t_2)\, a(x+t_1+t_2),$$

and a similar closed formula can be given for the $U^s(\mathbb{Z}_N)$-norms for $s \geq 3$. It can be shown that $\|\cdot\|_{U^s(\mathbb{Z}_N)}$ is a norm for $s \geq 2$, and for every $s \in \mathbb{N}$ we have

$$\|a\|_{U^{s+1}(\mathbb{Z}_N)} \geq \|a\|_{U^s(\mathbb{Z}_N)}. \tag{2.2}$$

In an informal way, having a small U^s-norm is interpreted as a property of U^s-uniformity, and we say that a function or sequence of functions is U^s-uniform if the corresponding uniformity norms converge to 0 as $N \to +\infty$. By (2.2) we get that U^{s+1}-uniformity implies U^s-uniformity.

Recall that the *Fourier transform* of a function a on $\mathbb{Z}_N$ is given by

$$\widehat{a}(\xi) := \mathbb{E}_{n\in\mathbb{Z}_N} a(n)\, \mathrm{e}\big(-n\frac{\xi}{N}\big) \quad \text{for} \quad \xi \in \mathbb{Z}_N,$$

where, as is standard, $\mathrm{e}(x) := \exp(2\pi i x)$. A direct computation gives the following identity that links the U^2-norm of a function a on $\mathbb{Z}_N$ with its Fourier coefficients:

$$\|a\|_{U^2(\mathbb{Z}_N)} = \|\widehat{a}\|_{\ell^4(\mathbb{Z}_N)} := \Big(\sum_{\xi\in\mathbb{Z}_N} \big|\widehat{a}(\xi)\big|^4\Big)^{1/4}. \tag{2.3}$$

It follows that if $|a| \leq 1$, then

$$\|a\|_{U^2(\mathbb{Z}_N)} \leq \sup_{\xi\in\mathbb{Z}_N} |\widehat{a}(\xi)|^{1/2} \leq \|a\|^{1/2}_{U^2(\mathbb{Z}_N)}. \tag{2.4}$$

We would like to stress though that similar formulas and estimates do not exist for higher order Gowers norms; a function bounded by 1 may have small Fourier coefficients, but large $U^s(\mathbb{Z}_N)$-norm for $s \geq 3$. In fact, eliminating all possible obstructions to $U^s(\mathbb{Z}_N)$-uniformity necessitates the study of correlations with all polynomial phases $\mathrm{e}(P(n))$, where $P \in \mathbb{R}[x]$ has degree $s-1$, and also the larger class of $(s-1)$-step nilsequences of bounded complexity (see Theorem 4.3).

For the purposes of this article it will be convenient to also define Gowers norms on an interval $[N]$ (this was also done in [29]). For $N^* \geq N$ we often identify the interval $[N^*]$ with $\mathbb{Z}_{N^*}$ in which case we consider $[N]$ as a subset of $\mathbb{Z}_{N^*}$.

Definition (Gowers norms on an interval [29])**.** Let $s \geq 2$, $N \in \mathbb{N}$, and $a\colon [N] \to \mathbb{C}$ be a function. By Lemma A.2 in Appendix A, the quantity

$$\|a\|_{U^s[N]} := \frac{1}{\|\mathbf{1}_{[N]}\|_{U^s(\mathbb{Z}_{N^*})}} \, \|\mathbf{1}_{[N]} \cdot a\|_{U^s(\mathbb{Z}_{N^*})}$$

is independent of N^* provided that $N^* > 2N$. It is called *the* $U^s[N]$*-norm of* a.

In complete analogy with the $U^s(\mathbb{Z}_N)$-norms we have that $\|\cdot\|_{U^s[N]}$ is a norm that increases with s, in the sense that for every $s \geq 2$ there exists a constant $c := c(s)$ such that $\|\cdot\|_{U^{s+1}[N]} \geq c\,\|\cdot\|_{U^s[N]}$. In Appendix A we derive various relations between the $U^s(\mathbb{Z}_N)$ and the $U^s[N]$ norms for $s \geq 2$. In particular, in Lemma A.4 we show that if $|a| \leq 1$, then $\|a\|_{U^s(\mathbb{Z}_N)}$ is small if and only if $\|a\|_{U^s[N]}$ is small. Thus, the reader should think of the $U^s[N]$ and $U^s(\mathbb{Z}_N)$ norms as equivalent measures of randomness; which one we use is a matter of convenience and depends on the particular problem at hand.

2.1.2. *Multiplicative functions.* Some examples of multiplicative functions of modulus at most 1 are the Möbius and the Liouville functions, the function $n \mapsto n^{it}$ for $t \in \mathbb{R}$, and Dirichlet characters, that is, periodic completely multiplicative functions that are not identically zero. Throughout, we denote by χ_q a Dirichlet character of least period q. Then $\chi_q(n) = 0$ whenever $(n,q) > 1$ and $\chi_q(n)$ is a $\phi(q)$-root of unity if $(n,q) = 1$, where ϕ is Euler's totient function.

It follows from results in [31,32] that the Möbius and the Liouville functions are U^s-uniform for every $s \in \mathbb{N}$. The next examples illustrate some simple but very important obstructions to uniformity for general bounded multiplicative functions.

Examples (Obstructions to uniformity)**.** (*i*) One easily sees that $\mathbb{E}_{n\in[N]} n^{it} \sim c_N := \frac{N^{it}}{1+it}$; hence for $t \neq 0$ the range of this average is contained densely in the circle with the center at zero and radius $1/\sqrt{1+t^2}$. Therefore, there is no constant c, independent of N, so that the function $n^{it} - c$ averages to 0 on $\mathbb{N}$. On the other hand, the average of $n^{it} - c_N$ on the interval $[N]$ converges to 0 as $N \to +\infty$, and, in fact, it can be seen[1] that $(n^{it} - c_N)_{n\in[N]}$, $N \in \mathbb{N}$, is U^s-uniform for every $s \geq 2$.

(*ii*) A non-principal Dirichlet character χ_q has average 0 on every interval with length a multiple of q; hence $\mathbb{E}_{n\in[N]}\chi_q(n) \to 0$ as $N \to +\infty$. However, χ_q is not U^2-uniform because it is periodic.

(*iii*) Let f be the completely multiplicative function defined by $f(2) := -1$ and $f(p) := 1$ for every prime $p \neq 2$. Equivalently, $f(2^m(2k+1)) = (-1)^m$ for all $k, m \geq 0$. Then $\mathbb{E}_{n\in[N]} f(n) = 1/3 + o(1)$, and this non-zero mean value already gives an obstruction to U^2-uniformity. But this is not the only obstruction. We have $\mathbb{E}_{n\in[N]}(-1)^n(f(n) - 1/3) = -2/3 + o(1)$, which implies that $f - 1/3$ is not U^2-uniform. In fact, it is not possible to subtract from f a periodic component f_{st} that is independent of N so that $f - f_{\text{st}}$ becomes U^2-uniform. But this problem is alleviated if we allow f_{st} to depend on N.

The first and third examples illustrate that the structured component we need to subtract from a multiplicative function so that the difference has a small $U^s(\mathbb{Z}_N)$-norm may vary a lot with N. This is one of the reasons why we cannot obtain an infinite variant of the structural result of Theorem 1.1. The last two examples illustrate that normalized multiplicative functions can have significant correlation with periodic phases; thus this is an obstruction to U^2-uniformity that we should take into account. However, it is a non-trivial fact that plays a central role in this article, that correlation with periodic phases is, in a sense to be made precise later, the only obstruction not only to U^2-uniformity but also to U^s-uniformity of multiplicative functions in $\mathcal{M}$ for all $s \geq 2$.

2.1.3. *The main structure theorem.* The structural result of Theorem 1.1 suffices for some applications, and a more informative variant is given in Theorem 8.1. But both results are not well suited for the combinatorial applications given in Section 2.4. The reason is that in such problems we seek to obtain positive lower bounds for certain averages of multiplicative functions, and the error introduced

[1]One can see this by using Theorem 8.1 and adjusting the argument used to prove Theorem 2.5.

by the uniform component typically subsumes the positive lower bound coming from the structured component as this depends on ε (via Q and R) in a rather inexplicit way. In order to overcome this obstacle, we would like to know that the uniformity norm of the uniform component can be chosen to be smaller than any predetermined positive function of Q and R. This can be achieved if we introduce an additional term that has a small $L^1[N]$-norm. We give the precise form of such a structural result after we set up some notation.

As it is often easier to work on a cyclic group rather than an interval of integers (this makes Fourier analysis tools more readily available), we introduce some notation to help us avoid roundabout issues.

Notation. Throughout, we assume that an integer $\ell \geq 2$ is given. This parameter adds some flexibility needed in the applications of our main structural results; its precise value will depend on the particular application we have in mind. We consider ℓ as fixed, and the dependence on ℓ is always left implicit. For every $N \in \mathbb{N}$, we denote by $\widetilde{N}$ any prime such that $N \leq \widetilde{N} \leq \ell N$. By Bertrand's postulate, such a prime always exists. In some cases we specify the value of $\widetilde{N}$, and its precise dependence on N depends on the application we have in mind.

For every multiplicative function $f \in \mathcal{M}$ and every $N \in \mathbb{N}$, we denote by f_N the function on $\mathbb{Z}_{\widetilde{N}}$, or on $[\widetilde{N}]$, defined by

$$f_N(n) := \begin{cases} f(n) & \text{if } n \in [N]; \\ 0 & \text{otherwise.} \end{cases} \tag{2.5}$$

Each time the domain of f_N will be clear from the context. Working with the truncated function $f \cdot \mathbf{1}_{[N]}$, rather than the function f, is a technical maneuver, and the reader will not lose much by ignoring the cutoff. We should stress that for the purposes of the structure theorem, U^s-norms are going to be defined and Fourier analysis is going to happen on the group $\mathbb{Z}_{\widetilde{N}}$ and not on the group $\mathbb{Z}_N$.

Definition. By a *kernel* on $\mathbb{Z}_{\widetilde{N}}$ we mean a non-negative function with average 1.

In the next statement we assume that the set $\mathcal{M}$ is endowed with the topology of pointwise convergence and thus is a compact metric space.

Theorem 2.1 (Structure theorem for multiplicative functions II)**.** [2] *Let $s \geq 2$, $\varepsilon > 0$, ν be a probability measure on the compact set $\mathcal{M}$, and $F\colon \mathbb{N} \times \mathbb{N} \times \mathbb{R}^+ \to \mathbb{R}^+$ be arbitrary. Then there exist positive integers Q and R that are bounded by a constant which depends only on s, ε, F, such that the following holds: For every sufficiently large $N \in \mathbb{N}$, which depends only on s, ε, F, and for every $f \in \mathcal{M}$, the function f_N admits the decomposition*

$$f_N(n) = f_{N,\mathrm{st}}(n) + f_{N,\mathrm{un}}(n) + f_{N,\mathrm{er}}(n) \quad \text{for every } \ n \in \mathbb{Z}_{\widetilde{N}},$$

where $f_{N,\mathrm{st}}$, $f_{N,\mathrm{un}}$, $f_{N,\mathrm{er}}$ satisfy the following properties:

(i) *$f_{N,\mathrm{st}} = f_N * \psi_{N,1}$ and $f_{N,\mathrm{st}} + f_{N,\mathrm{er}} = f_N * \psi_{N,2}$, where $\psi_{N,1}$, $\psi_{N,2}$ are kernels on $\mathbb{Z}_{\widetilde{N}}$ that do not depend on f, and the convolution product is defined in $\mathbb{Z}_{\widetilde{N}}$;*

[2] When $s = 3$, Sun [70] recently proved a similar result for multiplicative functions defined on the Gaussian integers.

(ii) $|f_{N,\mathrm{st}}(n+Q) - f_{N,\mathrm{st}}(n)| \le \dfrac{R}{\widetilde{N}}$ *for every* $n \in \mathbb{Z}_{\widetilde{N}}$*, where* $n+Q$ *is taken* mod $\widetilde{N}$*;*

(iii) *If* $\xi \in \mathbb{Z}_{\widetilde{N}}$ *satisfies* $\widehat{f}_{N,\mathrm{st}}(\xi) \neq 0$*, then* $\left|\dfrac{\xi}{\widetilde{N}} - \dfrac{p}{Q}\right| \le \dfrac{R}{\widetilde{N}}$ *for some* $p \in \{0, \dots, Q-1\}$*;*

(iv) $\|f_{N,\mathrm{un}}\|_{U^s(\mathbb{Z}_{\widetilde{N}})} \le \dfrac{1}{F(Q,R,\varepsilon)}$*;*

(v) $\mathbb{E}_{n\in\mathbb{Z}_{\widetilde{N}}} \displaystyle\int_{\mathcal{M}} |f_{N,\mathrm{er}}(n)|\, d\nu(f) \le \varepsilon$.

Remarks. (1) The result is of interest even when ν is supported on a single multiplicative function, that is, when $f \in \mathcal{M}$ is fixed and Property (v) is replaced with the estimate $\mathbb{E}_{n\in\mathbb{Z}_{\widetilde{N}}}|f_{N,\mathrm{er}}(n)| \le \varepsilon$. The stronger version stated is needed for the combinatorial applications.

(2) As remarked in the Introduction, various decomposition results with similar flavor have been proved for arbitrary bounded sequences but working in this generality necessitates the use of structured components that are much less rigid. An additional important feature of our result is that the structured component is defined by a convolution product with a kernel that is independent of $f \in \mathcal{M}$. All these properties play an important role in the derivation of some of our applications.

(3) In Section 10 we use Theorem 2.1 for the function $F(x,y,z) := cx^2y^2/z^4$ where c is a constant that depends on ℓ only. Restricting the statement to this function though does not simplify our proof.

(4) It is a consequence of Property (i) that for fixed F, N, ε, ν, the maps $f \mapsto f_{\mathrm{st}}, f \mapsto f_{\mathrm{un}}, f \mapsto f_{\mathrm{er}}$ are continuous, and $|f_{\mathrm{st}}| \le 1$, $|f_{\mathrm{un}}| \le 2$, $|f_{\mathrm{er}}| \le 2$.

(5) We do not know if a uniform version of the result holds, meaning with Property (v) replaced with $\sup_{f\in\mathcal{M}} \mathbb{E}_{n\in\mathbb{Z}_{\widetilde{N}}}|f_{N,\mathrm{er}}(n)| \le \varepsilon$.

The bulk of the work in the proof of Theorem 2.1 appears in the proof of Theorem 8.1 below which is a more informative version of Theorem 1.1 given in the Introduction. Two ideas that play a prominent role in the proof, roughly speaking, are as follows:

(a) A multiplicative function that has U^2-norm bounded away from zero correlates with a linear phase that has frequency close to a rational with small denominator.
(b) A multiplicative function that has U^s-norm bounded away from zero necessarily has U^2-norm bounded away from zero.

The proof of (a) uses classical Fourier analysis tools and is given in Section 3. The key number theoretic input is the orthogonality criterion of Kátai stated in Lemma 3.1.

The proof of (b) is much harder and is done in several steps using higher order Fourier analysis machinery. In Sections 6 and 7 we study the correlation of multiplicative functions with totally equidistributed (minor arc) nilsequences. This is the heart of the matter and the technically more demanding part in the proof of the structure theorem. The argument used by Green and Tao in [31] to prove similar estimates for the Möbius function uses special features of the Möbius and is inadequate for our purposes. To overcome this serious obstacle, we combine the orthogonality criterion of Kátai with a rather delicate asymptotic orthogonality result of polynomial nilsequences in order to establish a key discorrelation estimate (Theorem 6.1). This estimate is then used in Section 8, in conjunction with the U^s-inverse theorem (Theorem 4.3) and a factorization theorem (Theorem 5.6) of Green and Tao, to conclude the proof of Theorem 8.1. We defer the reader to Sections 6.1 and 8.1 for a more detailed sketch of the proof strategy of Property (b).

Upon proving Theorem 8.1, the proof of Theorem 2.1 consists of a Fourier analysis energy increment argument and avoids the use of finitary ergodic theory and the Hahn-Banach theorem, tools that are typically used for other decomposition results (see [20–22, 28, 72]). This hands-on approach enables us to transfer all the information obtained in Theorem 8.1 which is important for our applications.

Problem 1. *Can Theorems* 2.1 *and* 8.1 *be extended to multiplicative functions defined on quadratic number fields? More general number fields?*

2.2. **A generalization of a result of Daboussi.** The classic result of Daboussi [11–13] was recorded in the Introduction (see (1.2)). We prove the following generalization (all notions used below are defined in Sections 4 and 5).

Theorem 2.2 (Daboussi for nilsequences)**.** *Let* $X := G/\Gamma$ *be a nilmanifold and* $(g(n))_{n\in\mathbb{N}}$ *be a polynomial sequence in* G *such that* $(g(n)\cdot e_X)_{n\in\mathbb{N}}$ *is totally equidistributed in* X*. Then for every* $\Phi\in C(X)$ *with* $\int_X \Phi\, dm_X = 0$ *we have*

$$\lim_{N\to+\infty}\sup_{f\in\mathcal{M}}\Big|\frac{1}{N}\sum_{n=1}^{N} f(n)\,\Phi(g(n)\cdot e_X)\Big| = 0. \tag{2.6}$$

Remarks. (1) Theorem 2.2 follows from the stronger finitary statement in Theorem 6.1.

(2) If P is a polynomial with at least one non-constant coefficient irrational, then $(P(n))_{n\in\mathbb{N}}$ is totally equidistributed on the circle. Hence, for such a polynomial, $\mathrm{e}(P(n))$ can take the place of $\Phi(g(n)\cdot e_X)$ in (2.6), recovering a result of Kátai [46].

(3) An easy approximation argument allows one to extend the eligible functions F to all Riemann integrable functions on X. We can use this enhancement to show that any sequence of the form $\mathrm{e}(2\pi i[n\alpha]n\beta)$ with $1,\alpha,\beta$ rationally independent over $\mathbb{Q}$ can take the place of $\Phi(g(n)\cdot e_X)$ in (2.6).

2.3. **Aperiodic multiplicative functions.** It is known that the Möbius and the Liouville functions have zero average on every infinite arithmetic progression. In this subsection we work with the following vastly more general class of multiplicative functions.

Definition. We say that a multiplicative function $f\colon\mathbb{N}\to\mathbb{C}$ is *aperiodic* if it has zero average on every infinite arithmetic progression, that is,

$$\lim_{N\to+\infty}\mathbb{E}_{n\in[N]}f(an+b) = 0, \quad \text{for every } a,b\in\mathbb{N}.$$

In order to give easy to check conditions that imply aperiodicity, we will use a celebrated result of Halász [34]. To facilitate exposition we first define the distance between two multiplicative functions (see for example [24, 25]).

Definition. If $f,g\in\mathcal{M}$, we let $\mathbb{D}\colon \mathcal{M}\times\mathcal{M}\to[0,\infty]$ be given by
$$\mathbb{D}(f,g)^2=\sum_{p\in\mathbb{P}}\frac{1}{p}\big(1-\mathrm{Re}\big(f(p)\,\overline{g}(p)\big)\big),$$
where $\mathbb{P}$ denotes the set of prime numbers.

Remark. Note that if $|f|=|g|=1$, then $\mathbb{D}(f,g)^2=\sum_{p\in\mathbb{P}}\frac{1}{2p}\,|f(p)-g(p)|^2$.

Theorem 2.3 (Halász [34])**.** *A multiplicative function $f\in\mathcal{M}$ has mean value zero if and only if for every $t\in\mathbb{R}$ we either have $\mathbb{D}(f,n^{it})=\infty$ or $f(2^m)=-2^{imt}$ for all $m\in\mathbb{N}$.*

We record several conditions equivalent to aperiodicity; the last one is the easiest to verify for explicit multiplicative functions.

Proposition 2.4. *For a multiplicative function $f\in\mathcal{M}$ the following are equivalent:*

(i) *f is aperiodic.*
(ii) *For every $p,q\in\mathbb{N}$, we have $\lim\limits_{N\to+\infty}\mathbb{E}_{n\in[N]}f(n)\,\mathrm{e}(np/q)=0$.*
(iii) *For every Dirichlet character χ_q, we have $\lim\limits_{N\to+\infty}\mathbb{E}_{n\in[N]}f(n)\,\chi_q(n)=0$.*
(iv) *For every $t\in\mathbb{R}$ and Dirichlet character χ_q, we have either $\mathbb{D}(f,\chi_q(n)n^{it})=\infty$ or $f(2^m)\,\chi_q(2^m)=-2^{-imt}$ for every $m\in\mathbb{N}$.*

Assuming Theorem 2.3, the proof of the equivalences is simple (this was already observed in [11–13]); we give it for the convenience of the reader in Section 9.1. Lending terminology from [24, 25], Condition (iv) states that a multiplicative function is aperiodic unless it "pretends" to be $\chi_q(n)n^{it}$ for some Dirichlet character χ_q and some $t\in\mathbb{R}$. It follows easily from (iv) that if $f\in\mathcal{M}$ has real values and satisfies $\sum_{p\in\mathbb{P}\cap(d\mathbb{Z}+1)}\frac{1-f(p)}{p}=+\infty$ for every $d\in\mathbb{N}$, then f is aperiodic. In particular, this is satisfied by the Möbius and the Liouville functions. Sharper results can be obtained using a theorem of Hall [35] and the argument in [24, Corollary 2]. For instance, it can be shown that if $f(p)$ takes values in a finite subset of the unit disc and $f(p)\neq 1$ for all $p\in\mathbb{P}$, then f is aperiodic.

2.3.1. *Uniformity of aperiodic functions.* We give explicit necessary and sufficient conditions for a multiplicative function $f\in\mathcal{M}$ to be U^s-uniform, that is, have $U^s[N]$-norm converging to zero as $N\to+\infty$.

Aperiodicity is easily shown to be a necessary condition for U^2-uniformity. For general bounded sequences it is far from sufficient though. For instance, the sequences $(\mathrm{e}(n\alpha))_{n\in\mathbb{N}}$ and $(\mathrm{e}(n^2\alpha))_{n\in\mathbb{N}}$, where α is irrational, are aperiodic, but the first is not U^2-uniform and the second is U^2-uniform but not U^3-uniform. It is a rather surprising (and non-trivial) fact that for the general multiplicative function in $\mathcal{M}$ aperiodicity suffices for U^s-uniformity for every $s\geq 2$. This is not hard to show for $s=2$ by combining well known results about multiplicative functions, but for $s\geq 3$ it is much harder to do so, and we need to use essentially the full force of Theorem 1.1.

Theorem 2.5 (U^s-uniformity of aperiodic multiplicative functions)**.** *If a multiplicative function $f\in\mathcal{M}$ is aperiodic, then $\lim\limits_{N\to+\infty}\|f\|_{U^s[N]}=0$ for every $s\geq 2$.*

Remarks. (1) Our proof gives the following finitary inverse theorem: For given $s\in\mathbb{N}$ and $\varepsilon>0$, there exists $\delta:=\delta(s,\varepsilon)>0$ and $Q:=Q(s,\varepsilon)\in\mathbb{N}$, such that if $f \in \mathcal{M}$ satisfies $\limsup_{N\to+\infty} \|f\|_{U^s[N]} \geq \varepsilon$, then there exists $a, b \in \mathbb{N}$ with $1 \leq a, b \leq Q$ such that $\limsup_{N\to+\infty} \big|\mathbb{E}_{n\in[N]} f(an+b)\big| \geq \delta$. Note that δ and Q do not depend on f.

(2) Theorem 2.5 implies that if f is an aperiodic multiplicative function, then f does not correlate with any polynomial phase function $\mathrm{e}(p(n))$, $p \in \mathbb{R}[t]$. More generally, it implies that $\lim_{N\to+\infty} \mathbb{E}_{n\in[N]} f(n)\,\phi(n) = 0$ for every nilsequence $(\phi(n))_{n\in\mathbb{N}}$. For the Möbius and the Liouville functions this result was obtained by Green and Tao in [31].

2.3.2. *Chowla's conjecture for aperiodic multiplicative functions.* We provide a class of homogeneous polynomials $P \in \mathbb{Z}[m, n]$ such that $\mathbb{E}_{m,n\in[N]} f(P(m, n)) \to 0$ for every aperiodic completely multiplicative $f \in \mathcal{M}$.

Here and below, d is a positive integer, and by $\sqrt{-d}$ we mean $i\sqrt{d}$. It is well known that the ring of integers of $\mathbb{Q}(\sqrt{-d})$ is equal to $\mathbb{Z}[\tau_d]$ where

$$\tau_d := \frac{1}{2}(1 + \sqrt{-d}) \text{ if } d = 3 \bmod 4 \text{ and } \tau_d := \sqrt{-d} \text{ otherwise.}$$

The norm of $z \in \mathbb{Z}[\tau_d]$ is $\mathcal{N}(z) = |z|^2$. We write $Q_d(m, n)$ for the corresponding fundamental quadratic form, that is,

$$Q_d(m, n) := \mathcal{N}(m + n\tau_d) = |m + n\tau_d|^2.$$

Explicitly, we have

$$Q_d(m, n) = \begin{cases} m^2 + mn + \dfrac{d+1}{4} n^2 & \text{if } d = 3 \bmod 4; \\ m^2 + dn^2 & \text{otherwise.} \end{cases}$$

We say that a quadratic form $Q(m, n)$ with integer coefficients is *equivalent* to the form Q_d if it is obtained from Q_d by a change of variables given by a 2×2 matrix with integer entries and determinant equal to ± 1.

Convention. Every multiplicative function $f \in \mathcal{M}$ is extended to an even multiplicative function on $\mathbb{Z}$, by putting $f(0) = 0$ and $f(-n) = f(n)$ for every $n \in \mathbb{N}$. We denote this extension by f as well.

We prove the following theorem.

Theorem 2.6 (Some cases of Chowla's conjecture for aperiodic functions)**.** *Let $f \in \mathcal{M}$ be an aperiodic multiplicative function, $d \in \mathbb{N}$, Q be a quadratic form, equivalent to a quadratic form Q_d defined above, and let $r \geq 0$ and $s \geq 1$ be integers. Let $L_j(m, n)$, $j = 1, \dots, s$, be linear forms with integer coefficients, and suppose that either $s = 1$ or $s > 1$ and the linear forms L_1, L_j are linearly independent for $j = 2, \dots, s$. Then*

$$\lim_{N\to+\infty} \mathbb{E}_{m,n\in[N]} f\big(Q(m, n)^r\big) \prod_{j=1}^{s} f\big(L_j(m, n)\big) = 0. \tag{2.7}$$

Remarks. (1) The same statement holds with $f\big(Q(m, n)\big)^r$ in place of $f\big(Q(m, n)^r\big)$.

(2) A more general result is given in Theorem 9.7.

(3) The result fails when there are no linear factors; for instance, it fails for averages of the form $\mathbb{E}_{m,n\in[N]} f(m^2 + n^2)$, as it is easy to construct aperiodic mul-

tiplicative functions such that $f(m^2+n^2) = 1$ for all $m, n \in \mathbb{N}$; let f be 1 for integers that are sums of two squares and 0 if they are not. One can also construct examples of aperiodic completely multiplicative functions f with values ± 1 such that $f(m^2+n^2) = 1$ for all $m, n \in \mathbb{N}$.

(4) We restrict to positive definite quadratic forms because we can handle them using results for imaginary quadratic fields; real quadratic fields have infinitely many units, and this causes problems in our proof.

When $r = 0$, Theorem 2.6 follows by combining Theorem 2.5 with the estimates of Lemma 9.6. For $r \geq 1$ the main observation is that since $Q_d(m,n) = \mathcal{N}(m+n\tau_d)$ is completely multiplicative, the map $m + n\tau_d \mapsto f\big(Q_d(m,n)^r\big)$ is multiplicative in $\mathbb{Z}[\tau_d]$, in the sense defined in Section 9.5. Then using a variant of the orthogonality criterion of Katái for the ring $\mathbb{Z}[\tau_d]$ (see Proposition 9.5) we can show that the average on the left hand side of (2.7) converges to 0 if some other average that involves products of $2s$ linear forms converges to 0. With a bit of effort we show that the linear independence assumption of the linear forms is preserved, thus reducing the problem to the case $r = 0$ that we already know how to deal with using Theorem 1.1.

Perhaps the previous argument can be adjusted to treat the case of any irreducible quadratic polynomial Q. On the other hand, when Q has two or more irreducible quadratic factors or has irreducible factors of degree greater than two, we lose the basic multiplicativity property mentioned before and we do not see how to proceed. It could be the case though that Theorem 2.6 continues to hold, at least in the case that the function f is completely multiplicative, for averages of the form

$$\mathbb{E}_{m,n\in[N]} f\big(P(m,n)\big)$$

under the much weaker assumption that P is any homogeneous polynomial such that some linear form appears in the factorization of P with degree exactly one.

Problem 2. *Can Theorem* 2.6 *be extended to the case where Q is an arbitrary irreducible quadratic or a product of such? What about the case where Q is an arbitrary homogeneous polynomial without linear factors?*

2.4. **Partition regularity results.** An important question in Ramsey theory is to determine which algebraic equations, or systems of equations, are partition regular over the natural numbers. Here, we restrict our attention to polynomials in three variables, in which case partition regularity of the equation $p(x,y,z) = 0$ amounts to saying that, for any partition of $\mathbb{N}$ into finitely many cells, some cell contains *distinct* x, y, z that satisfy the equation.

The case where the polynomial p is linear was completely solved by Rado [66]; for $a, b, c \in \mathbb{N}$ the equation $ax + by = cz$ is partition regular if and only if a, b, or $a + b$ is equal to c. The situation is much less clear for second or higher degree equations, and only scattered results are known. Partition regularity is known when the equation satisfies a shift invariance property, as is the case for the equations $z - x = (y - x)^2$ (see [6] or [74]) and $x - y = z^2$ [1], and in other instances it can be deduced from related linear statements as is the case for the equation $xy = z^2$ (consider the induced partition for the powers of 2). But such fortunate occurrences are rather rare.

2.4.1. *Quadratic equations.* A notorious old question of Erdös and Graham [14] is whether the equation $x^2 + y^2 = z^2$ is partition regular. As Graham remarks in [23], "There is actually very little data (in either direction) to know which way to guess." More generally, one may ask for which $a, b, c \in \mathbb{N}$ is the equation

$$ax^2 + by^2 = cz^2 \tag{2.8}$$

partition regular. A necessary condition is that at least one of a, b, $a + b$ is equal to c, but currently there are no $a, b, c \in \mathbb{N}$ for which partition regularity of (2.8) is known.

We study here the partition regularity of Equation (2.8), and other quadratic equations, under the relaxed condition that the variable z is allowed to vary freely in $\mathbb{N}$; henceforth we use the letter λ to designate the special role of this variable.

Definition. The equation $p(x, y, \lambda) = 0$ is *partition regular* if for every partition of $\mathbb{N}$ into finitely many cells, one of the cells contains *distinct* x, y that satisfy the equation for some $\lambda \in \mathbb{Z}$.

A classical result of Furstenberg-Sárközy [17, 67] is that the equation $x - y = \lambda^2$ is partition regular. Other examples of translation invariant equations can be given using the polynomial van der Waerden theorem of Bergelson and Leibman [6], but not much is known in the non-translation invariant case. A result of Khalfalah and Szemerédi [49] is that the equation $x + y = \lambda^2$ is partition regular.[3] Again, the situation is much less clear when one considers non-linear polynomials in x and y, as is the case for the equation $ax^2 + by^2 = \lambda^2$ where $a, b \in \mathbb{N}$. We give the first positive results in this direction. For example, we show that the equations

$$16x^2 + 9y^2 = \lambda^2 \quad \text{and} \quad x^2 + y^2 - xy = \lambda^2$$

are partition regular (note that $16x^2 + 9y^2 = z^2$ is not partition regular). In fact, we prove a more general result for homogeneous quadratic forms in three variables.

Theorem 2.7 (Partition regularity of quadratic equations)**.** [4] *Let p be the quadratic form*

$$p(x, y, z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz, \tag{2.9}$$

where a, b, c are non-zero and d, e, f are arbitrary integers. Suppose that all three forms $p(x, 0, z)$, $p(0, y, z)$, $p(x, x, z)$ have non-zero square discriminants. Then the equation $p(x, y, \lambda) = 0$ is partition regular.

The last hypothesis means that the three integers

$$\Delta_1 := e^2 - 4ac, \quad \Delta_2 := f^2 - 4bc, \quad \Delta_3 := (e + f)^2 - 4c(a + b + d)$$

are non-zero squares. As a special case, we get the following result.

Corollary 2.8. *Let a, b, c, and $a + b$ be non-zero squares. Then the equation $ax^2 + by^2 = c\lambda^2$ is partition regular. Moreover, if a, b, and $a + b + c$ are non-zero squares, then the equation $ax^2 + by^2 + cxy = \lambda^2$ is partition regular.*

---

[3] Bergelson and Moreira [4, 5] recently proved partition regularity in $\mathbb{Q}$ for patterns of the form $\{x + y, xy\}$, or equivalently for the equation $\lambda x - y = \lambda^2$. Partition regularity in $\mathbb{Z}$ remains open.

[4] Sun [70] recently proved a similar partition regularity result on the Gaussian integers which covers the equation $x^2 - y^2 = \lambda^2$, where $x, y, \lambda \in \mathbb{Z}[i]$.

A partition $\mathcal{P}_1, \ldots, \mathcal{P}_r$ of the squares induces a partition $\widetilde{\mathcal{P}}_1, \ldots, \widetilde{\mathcal{P}}_r$ of $\mathbb{N}$ by the following rule: $x \in \widetilde{\mathcal{P}}_i$ if and only if $x^2 \in \mathcal{P}_i$. Applying the first part of Corollary 2.8 for the induced partition we deduce partition regularity results for the set of squares.

Corollary 2.9. *Let a, b, and $a+b$ be non-zero squares. Then for every partition of the squares into finitely many cells there exist distinct x and y belonging to the same cell such that $ax+by$ is a square.*

Furthermore, we get the following result.

Corollary 2.10. *Let a, b, and $a+b$ be non-zero squares, with a, b coprime. Then for every partition of the squares into finitely many cells there exist $m, n \in \mathbb{N}$ such that the integers $bm^2, n^2 - am^2$ belong to the same cell.*

Indeed, as in the previous corollary, we deduce from Theorem 2.7 that there exist squares x, y in the same cell and $\lambda \in \mathbb{N}$ such that $ax + by = b\lambda^2$. Since $(a, b) = 1$ we have $x = bm^2$ for some $m \in \mathbb{N}$ and $y = \lambda^2 - am^2$. The asserted conclusion holds for $n := \lambda$.

Although combinatorial tools, Fourier analysis tools, and the circle method have been used successfully to prove partition regularity of equations that enjoy some linearity features (also for non-linear equations with at least five variables [8, 42, 43, 47, 48, 69]), we have not found such tools adequate for the fully non-linear setup we are interested in. Instead, our main tool is going to be the structural result of Theorem 2.1. We give a summary of our proof strategy in Sections 2.4.4 and 10.6.

2.4.2. *Parametric reformulation.* In order to prove Theorem 2.7 we exploit some special features of the solution sets of the equations involved given in parametric form.

Definition. We say that the integers $\ell_0, \ldots, \ell_4$ are *admissible* if ℓ_0 is positive, $\ell_1 \neq \ell_2$, $\ell_3 \neq \ell_4$, and $\{\ell_1, \ell_2\} \neq \{\ell_3, \ell_4\}$.

The following result is proved in Appendix C.

Proposition 2.11 (Parametric form of solutions)**.** *Let the quadratic form p satisfy the hypothesis of Theorem 2.7. Then there exist admissible integers $\ell_0, \ldots, \ell_4$, such that for every $k, m, n \in \mathbb{Z}$, the integers $x := k\ell_0(m + \ell_1 n)(m + \ell_2 n)$ and $y := k\ell_0(m + \ell_3 n)(m + \ell_4 n)$ satisfy the equation $p(x, y, \lambda) = 0$ for some $\lambda \in \mathbb{Z}$.*

For example, the equation $16x^2 + 9y^2 = \lambda^2$ is satisfied by the integers $x := km(m + 3n)$, $y := k(m + n)(m - 3n)$, $\lambda := k(5m^2 + 9n^2 + 6mn)$, and the equation $x^2 + y^2 - xy = \lambda^2$ is satisfied by the integers $x := km(m+2n)$, $y := k(m-n)(m+n)$, $\lambda := k(m^2 + n^2 + mn)$.

The key properties of the patterns involved in Proposition 2.11 are as follows: (a) they are dilation invariant, which follows from homogeneity, (b) they "factor linearly," which follows from our assumption that the discriminants Δ_1, Δ_2 are squares, and (c) the coefficient of m in all forms can be taken to be 1, which follows from our assumption that the discriminant Δ_3 is a square.

Using Proposition 2.11, we see that Theorem 2.7 is a consequence of the following result.

Theorem 2.12 (Parametric reformulation of Theorem 2.7)**.** *Let $\ell_0, \ldots, \ell_4 \in \mathbb{Z}$ be admissible. Then for every partition of $\mathbb{N}$ into finitely many cells, there exist $k, m, n \in \mathbb{Z}$ such that the integers $k\ell_0(m+\ell_1 n)(m+\ell_2 n)$ and $k\ell_0(m+\ell_3 n)(m+\ell_4 n)$ are positive, distinct, and belong to the same cell.*

In fact, in Theorem 10.1 we prove something stronger, that any set of integers with positive multiplicative density (a notion defined in Section 10.1) contains the aforementioned configurations.

2.4.3. *More general patterns and higher degree equations.* Theorem 2.12 is proved using the structural result of Theorem 2.1 for $s = 3$; using this structural result for general $s \in \mathbb{N}$ we can prove, without essential changes in our argument, the following strengthening.

Theorem 2.13. *Let $s \geq 2$. For $i = 1, 2$ and $j = 1, \ldots, s$, let $L_{i,j}(m, n)$ be linear forms with integer coefficients. Suppose that for $i = 1, 2$, the linear forms $L_{i,j}$, $j = 1, \ldots, s$, are pairwise independent and that the product of the coefficients of m in the forms $L_{1,j}$ and in the forms $L_{2,j}$ are equal and non-zero. Then for every partition of $\mathbb{N}$ into finitely many cells, there exist $k, m, n \in \mathbb{Z}$ such that the integers $k \prod_{j=1}^{s} L_{1,j}(m, n)$ and $k \prod_{j=1}^{s} L_{2,j}(m, n)$ are positive, distinct, and belong to the same cell.*

Theorem 2.13 can be used to show that several homogeneous equations in three variables of degree greater than two are partition regular. Unfortunately, we have no general criterion like the one given in Theorem 2.7 and Corollary 2.8.[5] We record here one example of degree three that we found with the help of Tzanakis and some computer software[6] (examples of higher degree equations in three variables can also be found). Let

$$p(x, y, z) := 2x^3 - 2x^2y + 17x^2z - 4xyz + 44xz^2 - y^2z + 36z^3.$$

The equation $p(x, y, \lambda) = 0$ is satisfied for

$$x := km(2m + n)(m - n), \quad y := k(m + n)(2m - n)(m + 2n), \quad \lambda := km^2n$$

for every $k, m, n \in \mathbb{Z}$. It follows from Theorem 2.13 that the equation $p(x, y, \lambda) = 0$ is partition regular.

When one considers four or more variables, there is even more flexibility. For example Tzanakis brought to our attention the following identity of Gérardin:

$$(m^2 - n^2)^4 + (2mn + m^2)^4 + (2mn + n^2)^4 = 2(m^2 + mn + n^2)^4.$$

Using Theorem 2.12, we deduce that for every partition of $\mathbb{N}$ into finitely many cells, there exist distinct x, y belonging to the same cell and $\lambda, \mu \in \mathbb{N}$ such that $x^4 + y^4 = 2\lambda^4 - \mu^4$. As in Corollary 2.10, we deduce that for every partition of the fourth powers into finitely many cells, there exist $m, n, r \in \mathbb{N}$ such that the integers m^4 and $2n^4 - m^4 - r^4$ belong to the same cell.

2.4.4. *From partition regularity to multiplicative functions.* Much like the translation invariant case, where partition regularity results can be deduced from corresponding density statements with respect to a translation invariant density, we deduce Theorem 2.12 from the density regularity result of Theorem 10.1 that involves a dilation invariant density (a notion defined in Section 10.1).

In Section 10.2 we use a well known integral representation result of Bochner that characterizes positive definite sequences on the group $\mathbb{Q}^+$ in order to recast

[5]Note that a celebrated result of Faltings [15] implies that for $d \geq 4$ the equation $ax^d + by^d = cz^d$ has finitely many coprime solutions. This implies that such equations cannot be partition regular.

[6]`http://www.wolframalpha.com`.

the density regularity statement as a positivity property for an integral of averages of products of multiplicative functions (see Theorem 10.5). It is this positivity property that we seek to prove, and the heavy-lifting is done by the structural result of Theorem 2.1 for $s = 3$. The proof of the analytic statement of Theorem 10.5 is completed in Section 10.6, and the reader will find there a detailed sketch of the proof strategy for this crucial step. We remark that although we do not make explicit use of ergodic theory anywhere in this argument, ideas from the ergodic theoretic proof, given by Furstenberg [17], of Sárközy's theorem [67] have guided the last part of our argument.

2.4.5. *Further directions.* Theorem 2.7 implies that the equation

$$ax^2 + by^2 = c\lambda^2 \tag{2.10}$$

is partition regular provided that all three integers, ac, bc, $(a + b)c$, are non-zero squares. Two interesting cases, not covered by the previous result, are the following.

Problem 3. *Are the equations $x^2 + y^2 = \lambda^2$ and $x^2 + y^2 = 2\lambda^2$ partition regular?*[7]

Let us explain why we cannot yet handle these equations using the methods of this article. The equation $x^2 + y^2 = 2z^2$ has the following solutions: $x := k(m^2 - n^2 + 2mn)$, $y := k(m^2 - n^2 - 2mn)$, $z := k(m^2 + n^2)$, where $k, m, n \in \mathbb{Z}$. The values of x and y do not factor in linear terms, and uniformity estimates analogous to the ones stated in Lemma 10.7 fail. The equation $x^2 + y^2 = z^2$ has the following solutions: $x := k(m^2 - n^2)$, $y := 2kmn$, $z := k(m^2 + n^2)$. In this case, it is possible to establish the needed uniformity estimates, but we are not able to carry out the argument of Section 10.6 in order to prove the relevant positivity property (see footnote 13 in Section 10.6 below for more details).

A set $E \subset \mathbb{N}$ has *positive* (additive) *upper density* if $\limsup_{N\to+\infty} |E\cap[N]|/N > 0$. It turns out that the equations of Corollary 2.8 have non-trivial solutions on every infinite arithmetic progression, making the following statement plausible.

Problem 4. *Does every set $E \subset \mathbb{N}$ with positive upper density contain distinct $x, y \in \mathbb{N}$ that satisfy the equation $16x^2 + 9y^2 = \lambda^2$ for some $\lambda \in \mathbb{N}$?*

We say that the equation $p(x, y, z) = 0$, $p \in \mathbb{Z}[x, y, z]$, has *no local obstructions* if for every infinite arithmetic progression P, there exist distinct $x, y, z \in P$ that satisfy the equation. For example, the equations $x^2 + y^2 = 2z^2$ and $16x^2 + 9y^2 = 25z^2$ have no local obstructions.

Problem 5. *Let $p \in \mathbb{Z}[x, y, z]$ be a homogeneous quadratic form and suppose that the equation $p(x, y, z) = 0$ has no local obstructions. Is it true that every subset of $\mathbb{N}$ of positive upper density contains distinct x, y, z that satisfy the equation?*

See [33] for information regarding the density regularity of the equation $x^2+y^2 = 2z^2$.

The theorem of Sárközy [67] implies that for every finite partition of the integers some cell contains integers of the form $m, m+n^2$. What can we say about m? Can it be a square?

---

[7] Note that the equation $x^2+y^2 = 3\lambda^2$ does not have solutions in $\mathbb{N}$. Furthermore, the equation $x^2+y^2 = 5\lambda^2$ has solutions in $\mathbb{N}$, but it is not partition regular. Indeed, if we partition the integers in 6 cells according to whether their first non-zero digit in the 7-adic expansion is $1, 2, \ldots, 6$, it turns out that for every $\lambda \in \mathbb{N}$ the equation has no solution on any single partition cell.

Problem 6. *Is it true that for every partition of the integers into finitely many cells one cell contains integers of the form m^2 and m^2+n^2?*

We remark that the answer will be positive if one shows that the equation $x^2 - y^2 = \lambda^2$ is partition regular.

2.5. **Structure of the article.** In Section 3 we study the Fourier coefficients of multiplicative functions. Our basic tool is the orthogonality criterion of Kátai, and we establish the structural result of Theorem 3.3 which is a more informative version of Theorem 1.1 for $s = 2$.

In Sections 4 and 5 we review some facts about nilmanifolds and state some results that are instrumental for our subsequent work; the inverse theorem for the U^s-norms (Theorem 4.3), the quantitative Leibman theorem (Theorem 5.2), and the factorization theorem for polynomial sequences on nilmanifolds (Theorem 5.6). We also derive some consequences that will be used later on.

Sections 6 and 7 are in some sense the heart of the proof of our structural results. Our main result is Theorem 6.1 where we prove that an arbitrary multiplicative function has small correlation with all minor arc nilsequences. After proving some preparatory results in Section 6, we complete the proof of Theorem 6.1 in Section 7.

In Section 8 we prove our main structural results. Theorem 8.1 is a more informative version of Theorem 1.1, and we deduce Theorem 2.1 using an iterative argument of energy increment.

In Section 9 we deal with two applications of the structural result of Theorem 8.1 to aperiodic multiplicative functions. The first is Theorem 2.5 which states that a multiplicative function is aperiodic if and only if it is U^s-uniform for every $s \geq 2$. The second is Theorem 2.6 which provides a class of homogeneous polynomials in two variables for which Chowla's zero mean conjecture holds for every aperiodic completely multiplicative function.

Finally, in Section 10 we use the structural result of Theorem 2.1 to prove our main partition regularity result for homogeneous quadratic equations stated in Theorem 2.7.

2.6. **Notation and conventions.** For the reader's convenience, we gather here some notation that we use throughout the article.

We denote by $\mathbb{N}$ the set of positive integers.

For $N \in \mathbb{N}$ we denote by $[N]$ the set $\{1, \ldots, N\}$.

For a function a defined on a finite set A we write $\mathbb{E}_{x\in A}a(x) = \frac{1}{|A|}\sum_{x\in A} a(x)$.

With $\mathcal{M}$ we denote the set of multiplicative functions $f\colon \mathbb{N} \to \mathbb{C}$ with modulus at most 1, and with $\mathcal{M}_1^c$ the set of completely multiplicative functions $f\colon \mathbb{N} \to \mathbb{C}$ with modulus exactly 1.

A kernel on $\mathbb{Z}_N$ is a non-negative function on $\mathbb{Z}_N$ with average 1.

Throughout, we assume that we are given an integer $\ell \geq 2$, where its value depends on the problem at hand, and we leave the dependence on ℓ of the various parameters implicit.

For $N \in \mathbb{N}$ we let $\widetilde{N}$ be any prime with $N \leq \widetilde{N} \leq \ell N$. In some cases we specify the value of $\widetilde{N}$, and its precise dependence on N depends on the application we have in mind.

Given $f \in \mathcal{M}$ and $N \in \mathbb{N}$, we let $f_N\colon [\widetilde{N}] \to \mathbb{C}$ be defined by $f_N = f \cdot \mathbf{1}_{[N]}$. The domain of f_N is sometimes thought to be $\mathbb{Z}_{\widetilde{N}}$.

For technical reasons, throughout the article all Fourier analysis happens on $\mathbb{Z}_{\widetilde{N}}$, and all uniformity norms are defined on $\mathbb{Z}_{\widetilde{N}}$.

If x is real, $\mathrm{e}(x)$ denotes the number $e^{2\pi ix}$, $\|x\|$ denotes the distance between x and the nearest integer, $\lfloor x\rfloor$ denotes the largest integer smaller than or equal to x, and $\lceil x\rceil$ denotes the smallest integer greater than or equal to x.

Given $k\in\mathbb{N}$, we write $\mathbf{h}=(h_1,\dots,h_k)$ for a point of $\mathbb{Z}^k$ and $\|\mathbf{h}\|=|h_1|+\cdots+|h_k|$.

For $\mathbf{u}=(u_1,\dots,u_k)\in\mathbb{T}^k$, we write $\|\mathbf{u}\|=\|u_1\|+\cdots+\|u_k\|$, and for $\mathbf{h}\in\mathbb{Z}^k$, we write $\mathbf{h}\cdot\mathbf{u}=h_1u_1+\cdots+h_ku_k$.

If Φ is a function on a metric space X with distance d, we let

$$\|\Phi\|_{\mathrm{Lip}(X)}=\sup_{x\in X}|\Phi(x)|+\sup_{\substack{x,y\in X\\ x\neq y}}\frac{|\Phi(x)-\Phi(y)|}{d(x,y)}.$$

There is a proliferation of constants in this article, and our general principles are as follows: The constants $\ell,\ell_1,\ell_2,\dots,$ are considered as fixed throughout the article, and quantities depending only on these numbers are considered as universal constants. Quantities that depend on one or more variables are denoted by Roman capital letters $C,D,K,L,\dots$ if they represent large quantities and by lowercase Greek letters $\delta,\varepsilon,\eta,\tau,\dots$ and c if they represent small quantities. It will be very clear from the context when we deviate from these rules.

3. Fourier analysis of multiplicative functions

In this section we study the Fourier coefficients of multiplicative functions. Our goal is to establish Theorem 3.3, which proves the case $s=2$ of the decomposition result of Theorem 1.1 and gives more precise information on the structured and uniform components. We will use this result in Section 8 as our starting point in the proof of the more general structure theorem for the U^s-norm for $s\geq 2$.

We recall some notation and conventions. The integer $\ell\geq 2$ is considered as fixed throughout, and we never indicate the dependence on this number. For every $N\in\mathbb{N}$, we denote by $\widetilde{N}$ a prime with $N\leq\widetilde{N}\leq\ell N$. For every $f\in\mathcal{M}$, we write $f_N=f\cdot\mathbf{1}_{[N]}$, and we consider this as a function defined on $\mathbb{Z}_{\widetilde{N}}$. Henceforth, all convolution products are defined on $\mathbb{Z}_{\widetilde{N}}$ and the Fourier coefficients of f_N are given by

$$\widehat{f_N}(\xi):=\mathbb{E}_{n\in\mathbb{Z}_{\widetilde{N}}}f_N(n)\,\mathrm{e}(-n\xi/\widetilde{N})\quad\text{ for }\ \xi\in\mathbb{Z}_{\widetilde{N}}.$$

3.1. The Kátai orthogonality criterion.

We start with a key number theoretic input that we need in this section and which will also be used later in Sections 7 and 9.

Lemma 3.1 (Kátai orthogonality criterion [46]; see also [11,63])**.** *For every $\varepsilon>0$ there exists $\delta:=\delta(\varepsilon)>0$ and $K:=K(\varepsilon)$ such that the following holds: If $N\geq K$, $a\colon[N]\to\mathbb{C}$ is a function with $|a|\leq 1$, and*

$$\max_{\substack{p,q\text{ primes}\\ 1<p<q<K}}\left|\mathbb{E}_{n\in[\lfloor N/q\rfloor]}a(pn)\,\overline{a}(qn)\right|<\delta,$$

then

$$\sup_{f\in\mathcal{M}}\left|\mathbb{E}_{n\in[N]}f(n)\,a(n)\right|<\varepsilon.$$

Remark. The result is stated and proved in [46] for functions $a\colon [N] \to \mathbb{C}$ of modulus 1, but the same argument works for sequences with modulus at most 1.

The dependence of δ and K on ε can be made explicit (for good bounds see [7]) but we do not need such extra information here. We give a complete proof of Lemma 3.1 in a more general context in Section 9 (see Lemma 9.4).

3.2. **Fourier coefficients of multiplicative functions.** Next, we use the orthogonality criterion of Kátai in order to prove that the Fourier coefficients of the restriction of a multiplicative function on an interval $[N]$ are small unless the frequency is close to a rational with small denominator. Furthermore, the implicit constants do not depend on the multiplicative function or the integer N.

Corollary 3.2 (U^2 non-uniformity)**.** *For every $\theta > 0$ there exist positive integers $N_0 := N_0(\theta)$, $Q := Q(\theta)$, and $V := V(\theta)$, such that for every $N \geq N_0$, for every $f \in \mathcal{M}$, and for every $\xi \in \mathbb{Z}_{\widetilde{N}}$, we have the following implication:*

$$\text{if } \ |\widehat{f_N}(\xi)| \geq \theta, \quad \text{then } \quad \Big\| \frac{Q\xi}{\widetilde{N}} \Big\| \leq \frac{QV}{\widetilde{N}}. \tag{3.1}$$

Proof. Let $\delta := \delta(\theta)$ and $K := K(\theta)$ be defined by Lemma 3.1, and let $Q := K!$. Suppose that $N > K$. Let p and p' be primes with $p < p' \leq K$, and let $\xi \in \mathbb{Z}_{\widetilde{N}}$. If $\xi = 0$, the conclusion is obvious; otherwise, since $\widetilde{N}$ is a prime greater than K, we have $\|Q\xi/\widetilde{N}\| \neq 0$. Since Q is a multiple of $p' - p$, we have

$$0 < \|Q\xi/\widetilde{N}\| \leq \frac{Q}{p'-p} \|(p'-p)\xi/\widetilde{N}\| \leq Q\, \|(p'-p)\xi/\widetilde{N}\|.$$

Since $\widetilde{N} \leq \ell N$, we deduce that

$$|\mathbb{E}_{n\in[\lfloor N/p'\rfloor]}\, \mathrm{e}(p'n\xi/\widetilde{N})\, \mathrm{e}(-pn\xi/\widetilde{N})| \leq \frac{2p'}{N\|(p'-p)\xi/\widetilde{N}\|} \leq \frac{2KQ}{N\|Q\xi/\widetilde{N}\|} \leq \frac{2\,\ell KQ}{\widetilde{N}\|Q\xi/\widetilde{N}\|}.$$

Let $V := 2\,\ell K/\delta$. If $\|Q\xi/\widetilde{N}\| > QV/\widetilde{N}$, then the rightmost term of the last inequality is smaller than δ, and thus, by Lemma 3.1 we have

$$|\widehat{f_N}(\xi)| = \big|\mathbb{E}_{n\in[\widetilde{N}]}\, f_N(n)\mathrm{e}(-n\xi/\widetilde{N})\big| = \frac{N}{\widetilde{N}}\big|\mathbb{E}_{n\in[N]} f(n)\mathrm{e}(-n\xi/\widetilde{N})\big| < \theta$$

contradicting (3.1). Hence, $\|Q\xi/\widetilde{N}\| \leq QV/\widetilde{N}$, completing the proof. □

3.3. **Definition of kernels.** We recall that a *kernel* on $\mathbb{Z}_{\widetilde{N}}$ is a non-negative function ϕ on $\mathbb{Z}_{\widetilde{N}}$ with $\mathbb{E}_{n\in\mathbb{Z}_{\widetilde{N}}}\phi(n) = 1$. The *spectrum* of a function $\phi\colon \mathbb{Z}_{\widetilde{N}} \to \mathbb{C}$ is the set

$$\mathrm{Spec}(\phi) := \big\{\xi \in \mathbb{Z}_{\widetilde{N}} \colon \widehat{\phi}(\xi) \neq 0\big\}.$$

Next, we make some explicit choices for the constants Q and V of Corollary 3.2. This will enable us to compare the Fourier transforms of the kernels $\phi_{N,\theta}$ defined below for different values of θ and to establish the monotonicity Property (3.11) in Theorem 3.3 below. For every $\theta > 0$ let $N_0(\theta)$ be as in Corollary 3.2. For

$N \geq N_0(\theta)$ we define

$$\mathcal{A}(N,\theta) := \Big\{\xi \in \mathbb{Z}_{\widetilde{N}} \colon \exists f \in \mathcal{M} \ \text{ such that } \ |\widehat{f_N}(\xi)| \geq \theta^2\Big\};$$

$$W(N,q,\theta) := \max_{\xi \in \mathcal{A}(N,\theta)} \Big\{\widetilde{N}\Big\|q\,\frac{\xi}{\widetilde{N}}\Big\|\Big\};$$

$$Q(\theta) := \min_{k\in\mathbb{N}}\Big\{k!\colon \ \limsup_{N\to+\infty} W(N,k!,\theta) < +\infty\Big\}; \tag{3.2}$$

$$V(\theta) := 1 + \Big\lceil \frac{1}{Q(\theta)} \limsup_{N\to+\infty} W(N,Q(\theta),\theta)\Big\rceil. \tag{3.3}$$

It follows from Corollary 3.2 that the set of integers used in the definition of $Q(\theta)$ is non-empty; hence $Q(\theta)$ is well defined. It follows from the preceding definitions that there exists $N_1 := N_1(\theta)$ such that

(3.4) Implication (3.1) holds for $N \geq N_1$ with θ^2 substituted for θ, $V(\theta)$ for V, and $Q(\theta)$ for Q.

Furthermore, for $0 < \theta' \leq \theta$, we have $Q(\theta') \geq Q(\theta)$, and thus

$$\text{for } \ 0 < \theta' \leq \theta, \ \text{ the integer } \ Q(\theta') \ \text{ is a multiple of } \ Q(\theta). \tag{3.5}$$

Moreover, it can be checked that

$$V(\theta) \ \text{ increases as } \ \theta \ \text{ decreases.} \tag{3.6}$$

Next, we use the constants just defined to build the kernels $\phi_{N,\theta}$ of Theorem 3.3 below.

For every $m \in \mathbb{N}$ and $\widetilde{N} > 2m$ the "Fejer kernel" $\phi_{N,m}$ on $\mathbb{Z}_{\widetilde{N}}$ is defined by

$$\phi_{N,m}(x) := \sum_{-m\leq\xi\leq m} \big(1 - \frac{|\xi|}{m}\big)\, \mathrm{e}\big(x\,\frac{\xi}{\widetilde{N}}\big),$$

where the interval $\{-m,\dots,m\}$ is imbedded in $\mathbb{Z}_{\widetilde{N}}$ in the obvious way. The spectrum of $f_{N,m}$ is the subset $\{-m+1,\dots,m-1\}$ of $\mathbb{Z}_{\widetilde{N}}$. Let $Q_N(\theta)^*$ be the inverse of $Q(\theta)$ in $\mathbb{Z}_{\widetilde{N}}$, that is, the unique integer in $\{1,\dots,\widetilde{N}-1\}$ such that $Q(\theta)Q_N(\theta)^* = 1 \bmod \widetilde{N}$. Let

$$N_0 := N_0(\theta) = \max\{N_1, 2Q(\theta)V(\theta)\lceil\theta^{-2}\rceil\}. \tag{3.7}$$

For $N \geq N_0$ we define

$$\phi_{N,\theta}(x) := f_{N,Q(\theta)V(\theta)\lceil\theta^{-4}\rceil}(Q_N(\theta)^* x). \tag{3.8}$$

An equivalent formulation is that $f_{N,Q(\theta)V(\theta)\lceil\theta^{-4}\rceil}(x) = \phi_{N,\theta}(Q(\theta)x)$. The spectrum of the kernel $\phi_{N,\theta}$ is the set

$$\Xi_{N,\theta} := \Big\{\xi \in \mathbb{Z}_{\widetilde{N}} \colon \Big\|\frac{Q(\theta)\xi}{\widetilde{N}}\Big\| < \frac{Q(\theta)V(\theta)\lceil\theta^{-4}\rceil}{\widetilde{N}}\Big\}, \tag{3.9}$$

and we have

$$\widehat{\phi_{N,\theta}}(\xi) = \begin{cases} 1 - \Big\|\dfrac{Q(\theta)\xi}{\widetilde{N}}\Big\| \dfrac{\widetilde{N}}{Q(\theta)V(\theta)\lceil\theta^{-4}\rceil} & \text{if } \ \xi \in \Xi_{N,\theta}\ ; \\ 0 & \text{otherwise.} \end{cases} \tag{3.10}$$

Note that the cardinality of $\Xi_{N,\theta}$ is bounded by a constant that depends only on θ.

3.4. **U^2-structure theorem for multiplicative functions.** We can now state and prove the main result of this section.

Theorem 3.3 (U^2-structure theorem for multiplicative functions)**.** *Let $\theta > 0$. There exist positive integers $N_0 := N_0(\theta)$, $Q := Q(\theta)$, $R := R(\theta)$, such that for $N \geq N_0$ the following holds: Let the kernel $\phi_{N,\theta}$ be defined in Section 3.3, and for every $f \in \mathcal{M}$ let*

$$f_{N,\mathrm{st}} := f_N * \phi_{N,\theta} \quad \text{and} \quad f_{N,\mathrm{un}} := f_N - f_{N,\mathrm{st}}.$$

Then we have

(i) *If $\xi \in \mathbb{Z}_{\widetilde{N}}$ satisfies $\widehat{f}_{N,\mathrm{st}}(\xi) \neq 0$, then $\left|\frac{\xi}{\widetilde{N}} - \frac{p}{Q}\right| \leq \frac{R}{\widetilde{N}}$ for some $p \in \{0, \dots, Q-1\}$;*

(ii) *$|f_{N,\mathrm{st}}(n+Q) - f_{N,\mathrm{st}}(n)| \leq \frac{R}{\widetilde{N}}$ for every $n \in \mathbb{Z}_{\widetilde{N}}$, where $n+Q$ is taken mod $\widetilde{N}$;*

(iii) *$\|f_{N,\mathrm{un}}\|_{U^2(\mathbb{Z}_{\widetilde{N}})} \leq \theta$.*

Moreover, if $0 < \theta' \leq \theta$ and $N \geq \max\{N_0(\theta), N_0(\theta')\}$, then

$$\text{for every } \xi \in \mathbb{Z}_{\widetilde{N}},\ \widehat{\phi_{N,\theta'}}(\xi) \geq \widehat{\phi_{N,\theta}}(\xi) \geq 0. \tag{3.11}$$

Remarks. (1) The monotonicity Property (3.11) plays a central role in the derivation of Theorem 2.1 from Theorem 8.1 in Section 8.10. This is one of the reasons why we construct the kernels $\phi_{N,\theta}$ explicitly in Section 3.3.

(2) The values of Q and R given by Theorem 3.3 will be used later in Section 8, and they do not coincide with the values of Q and R in Theorems 2.1 and 8.1.

Proof. We first show that (3.11) holds. Indeed, suppose that $\theta \geq \theta' > 0$ and that $N \geq \max\{N_0(\theta), N_0(\theta')\}$. We have to show that $\widehat{\phi_{N,\theta'}}(\xi) \geq \widehat{\phi_{N,\theta}}(\xi)$ for every $\xi \in \mathbb{Z}_{\widetilde{N}}$. Using (3.5) and (3.6), we get that $\Xi_{N,\theta'}$ contains the set $\Xi_{N,\theta}$. Thus, we can assume that ξ belongs to the latter set as the estimate is obvious otherwise. In this case, the claim follows from (3.5), (3.6) and the formula (3.10) giving the Fourier coefficients of $\phi_{N,\theta}$.

Next, we show the remaining assertions (i), (ii), (iii) of the statement. Let $\theta > 0$. Let $Q := Q(\theta)$, $V := V(\theta)$, $N_0(\theta)$ be defined by (3.2), (3.3), (3.7), respectively. Suppose that $N \geq N_0(\theta)$, and let $\phi_{N,\theta}$ and $\Xi_{N,\theta}$ be defined by (3.8) and (3.9), respectively.

If for some $f \in \mathcal{M}$ and $\xi \in \mathbb{Z}_{\widetilde{N}}$ we have $\widehat{f_{N,\mathrm{st}}}(\xi) \neq 0$, then $\widehat{\phi_{N,\theta}}(\xi) \neq 0$ and ξ belongs to the set $\Xi_{N,\theta}$ defined by (3.9). Hence, Property (i) holds, for some constant R depending only on θ.

Moreover, for $f \in \mathcal{M}$ and $n \in \mathbb{Z}_{\widetilde{N}}$, using the Fourier inversion formula and the estimate $|\mathrm{e}(x) - 1| \leq 2\pi\|x\|$, we get

$$|(\phi_{N,\theta}*f_N)(n+Q)-(\phi_{N,\theta}*f_N)(n)| \leq 2\pi \sum_{\xi \in \mathbb{Z}_{\widetilde{N}}} |\widehat{\phi_{N,\theta}}(\xi)| \cdot \Big\| Q\frac{\xi}{\widetilde{N}} \Big\| \leq 2\pi \frac{|\Xi_{N,\theta}| Q V \lceil \theta^{-4} \rceil}{\widetilde{N}},$$

where the last estimate follows from (3.9). The last term in this inequality is bounded by $R/\widetilde{N}$ for some constant R that depends only on θ. This establishes Property (ii).

Last, since $N \geq N_0(\theta) \geq N_1(\theta)$, by (3.4) we have that for every $f \in \mathcal{M}$, if $|\widehat{f_N}(\xi)| \geq \theta^2$, then $\|Q\xi/\widetilde{N}\| \leq QV/\widetilde{N}$ and thus $\widehat{\phi_{N,\theta}}(\xi) \geq 1 - \theta^4$ by (3.10). It follows that $|\widehat{f_N}(\xi) - \widehat{\phi_{N,\theta} * f_N}(\xi)| \leq \theta^4 \leq \theta^2$. This last bound is clearly also true when $|\widehat{f_N}(\xi)| < \theta^2$, and thus, using identity (2.3), we get

$$\begin{aligned}\|f_N - \phi_{N,\theta} * f_N\|^4_{U^2(\mathbb{Z}_{\widetilde{N}})} &= \sum_{\xi\in\mathbb{Z}_{\widetilde{N}}} |\widehat{f_N}(\xi) - \widehat{\phi_{N,\theta} * f_N}(\xi)|^4 \\ &\leq \theta^4 \sum_{\xi\in\mathbb{Z}_{\widetilde{N}}} |\widehat{f_N}(\xi) - \widehat{\phi_{N,\theta} * f_N}(\xi)|^2 \leq \theta^4 \sum_{\xi\in\mathbb{Z}_{\widetilde{N}}} |\widehat{f_N}(\xi)|^2 \leq \theta^4,\end{aligned}$$

where the last estimate follows from Parseval's identity. Hence, $\|f_N - \phi_{N,\theta} * f_N\|_{U^2(\mathbb{Z}_{\widetilde{N}})} \leq \theta$, proving Property (iii) and completing the proof of the theorem. □

3.5. **A model structure theorem.** Before we enter the proof of the U^s-structure theorem for $s \geq 3$ we sketch the proof of a toy model that can serve as a guide for the much more complicated argument that comes later on.

Proposition 3.4 (Model structure theorem for multiplicative functions)**.** *Let $\varepsilon > 0$. There exists $\theta := \theta(\varepsilon)$ such that for every sufficiently large $N \in \mathbb{N}$, depending only on ε, the decomposition $f_N = f_{N,\mathrm{st}} + f_{N,\mathrm{un}}$ associated to θ by Theorem* 3.3 *satisfies Properties* (i), (ii) *of this theorem, and also*

$$\sup_{f\in\mathcal{M},\ \alpha\in\mathbb{R}} \big|\mathbb{E}_{n\in[N]} f_{N,\mathrm{un}}(n)\, \mathrm{e}(n^2\alpha)\big| \leq \varepsilon. \tag{3.12}$$

Proof (Sketch). Let $\varepsilon > 0$ and $N \in \mathbb{N}$ be sufficiently large. Let $\alpha \in \mathbb{R}$ and $f \in \mathcal{M}$, and suppose that

$$\big|\mathbb{E}_{n\in[N]} f_{N,\mathrm{un}}(n)\, \mathrm{e}(n^2\alpha)\big| \geq \varepsilon. \tag{3.13}$$

Minor arcs. Recall that $f_{N,\mathrm{st}} = f_N * \phi_{N,\theta}$ where $\phi_{N,\theta}$ is a kernel on $\mathbb{Z}_{\widetilde{N}}$ and the convolution is taken on $\mathbb{Z}_{\widetilde{N}}$. Therefore, we have $f_{N,\mathrm{un}} = f_N * \psi_{N,\theta}$ where the function $\psi_{N,\theta}$ satisfies $\mathbb{E}_{n\in\mathbb{Z}_{\widetilde{N}}}|\psi_{N,\theta}| \leq 2$. Taking into account the roundabout effects, and using that $\widetilde{N} \leq \ell N$, we deduce that there exists $k \in \mathbb{Z}$ with

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_{[N]}(n+k)\, f_N(n)\, \mathrm{e}((n+k)^2\alpha)\big| \geq \varepsilon/(4\ell).$$

Let K and δ be given by Lemma 3.1 with $\varepsilon/(4\ell)$ substituted for ε. From this lemma and the last estimate it follows that there exist $k \in \mathbb{Z}$ and primes p, p' with $p < p' < K$ such that

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_I(n)\, \mathrm{e}((p^2 - p'^2)n^2\alpha + 2(p-p')kn\alpha)\big| \geq \delta,$$

where I is the interval $I := \{n \in [N]\colon pn, p'n, pn+k, p'n+k \in [N]\}$ (its length is necessarily greater than δN). We interpret this formula by saying that

$$\big((p^2 - p'^2)n^2\alpha + 2(p-p')kn\alpha\big)_{n\in[\widetilde{N}]}$$

is not "well" equidistributed on the torus.

Using Weyl-type results (see for example [30, Proposition 4.3]) we get that there exist positive integers $Q := Q(\varepsilon), R := R(\varepsilon)$ such that

$$\Big|\alpha - \frac{p}{q}\Big| \leq \frac{R}{\widetilde{N}^2} \quad \text{for some } \ p \in \mathbb{Z} \ \text{ and some } q \text{ with } \ 1 \leq q \leq Q. \tag{3.14}$$

In other words, α belongs to a "major arc"; that is, it is close to a rational with a small denominator.

Major arcs. We factorize the sequence $(n^2\alpha)_{n\in[\widetilde{N}]}$ as follows:

$$n^2\alpha = \epsilon(n) + \gamma(n), \qquad \text{where} \qquad \epsilon(n) := n^2\big(\alpha - \frac{p}{q}\big), \ \ \gamma(n) := n^2\frac{p}{q}.$$

The sequence ϵ varies slowly (this follows from (3.14)), and the sequence γ has period q. After partitioning the interval $[\widetilde{N}]$ into sub-progressions where $\epsilon(n)$ is almost constant and $\gamma(n)$ is constant, it is not hard to deduce from (3.13) that

$$|\mathbb{E}_{n\in[\widetilde{N}]}\mathbf{1}_P(n)\cdot f_{N,\mathrm{un}}(n)| > \frac{1}{10}\frac{\varepsilon^2}{QR}$$

for some arithmetic progression $P \subset [\widetilde{N}]$. From this and Lemma A.6 we deduce that

$$\|f_{N,\mathrm{un}}\|_{U^2(\mathbb{Z}_{\widetilde{N}})} > \frac{1}{C}\frac{\varepsilon^2}{QR} =: \theta(\varepsilon),$$

where C is a positive universal constant. This contradicts Property (iii) of Theorem 3.3 and completes the proof. $\square$

Our next goal is for every $s \geq 2$ to replace the estimate in (3.12) with the estimate $\|f_{N,\mathrm{un}}\|_{U^s(\mathbb{Z}_{\widetilde{N}})} \leq \varepsilon$. To this end, we shall see (using the inverse theorem in [32]) that it suffices to get a strengthening of (3.12) where the place of $(\mathrm{e}(n^2\alpha))$ takes any s-step nilsequence $(\Phi(a^n \cdot e_X))$ where Φ is a function on an s-step nilmanifold G/Γ with Lipschitz norm at most 1 and $a \in G$. This is an immensely more difficult task, and it is carried out in the next five sections.

4. Nilmanifolds and the inverse theorem for the U^s-norms

In this section, we review some basic concepts on nilmanifolds and also record the inverse theorem for the U^s-norms. As most notions will be used subsequently to state theorems from [30], we follow the notation used in [30]. Mal'cev bases were introduced in [57], and proofs of foundational properties of Mal'cev basis and rational subgroups used in this article can be found in [10].

4.1. Basic definitions.

Let G be a connected, simply connected, s-step nilpotent Lie group and Γ be a discrete co-compact subgroup. The *commutator subgroups* G_i of G are defined by $G_0 = G_1 := G$ and $G_{i+1} := [G, G_i]$ for $i \in \mathbb{N}$. We have $G_{s+1} = \{1_G\}$.

The compact manifold $X := G/\Gamma$ is called an *s-step nilmanifold.* In some cases the degree of nilpotency does not play a particular role, and we refer to X as a *nilmanifold.*

We view elements of G/Γ as "points" on the nilmanifold X rather than equivalence classes and denote them by x, y, etc. The projection in X of the unit element 1_G of G is called the *base point* of X and is denoted by e_X. The action of G on X is denoted by $(g, x) \mapsto g \cdot x$. The *Haar measure* m_X of X is the unique probability measure on X that is invariant under this action.

Convention. We never consider "nude" nilmanifolds, but assume (often implicitly) that some supplementary structure is given. First, every nilmanifold X can be represented as a quotient G/Γ in several ways, but we assume that one of them is fixed. Moreover, we assume that G is endowed with a *rational filtration*, a *Mal'cev basis* for X adapted to the filtration, and the corresponding Riemannian metric. We define these objects next.

Definition ([30])**.** Let G be a connected, simply connected s-step nilpotent Lie group, and let Γ be a discrete co-compact subgroup of G. A *rational subgroup* of G is a connected, simply connected, closed subgroup G' of G such that $G' \cap \Gamma$ is co-compact in G'.

It is known that the commutator subgroups G_i are rational (see for example [10, Theorem 5.1.1 and Corollary 5.2.2]). More properties of rational subgroups are given in Appendix B.

Definition ([30])**.** A *filtration* $G_\bullet$ on G is a sequence of rational subgroups

$$G_\bullet := \{G = G^{(0)} = G^{(1)} \supset G^{(2)} \supset \cdots \supset G^{(t)} \supset G^{(t+1)} = \{1_G\} = G^{(t+2)} = \cdots\},$$

which has the property that $[G^{(i)}, G^{(j)}] \subset G^{(i+j)}$ for all integers $i, j \geq 0$. The *degree* of the filtration $G_\bullet$ is the smallest integer t such that $G^{(t+1)} = \{1_G\}$. The filtration is *rational* if the groups $G^{(i)}$ are rational.

The *natural filtration* is the *lower central series* that consists of the commutator subgroups G_i, $i \geq 0$, of G. It is a rational filtration and has degree s when G is s-step nilpotent.

Let $G^{(i)}$, $i \geq 0$ be a filtration. We remark that as $[G, G^{(i)}] \subset G^{(i)}$, we have that $G^{(i)}$ is a normal subgroup of G for $i \in \mathbb{N}$. Since $G^{(2)} \supset G_2$, the quotient group $G/G^{(2)}$ is Abelian and isomorphic to $\mathbb{R}^q$ for some $q \geq 0$.

Definition. Let $X := G/\Gamma$ be an s-step nilmanifold and $G_\bullet$ be a filtration. We let $m := \dim(G)$ and $m_i := \dim(G^{(i)})$ for $i \geq 0$. A basis $\mathcal{X} := \{\xi_1, \ldots, \xi_m\}$ for the Lie algebra $\mathfrak{g}$ of G over $\mathbb{R}$ is called a *Mal'cev basis for* X adapted to $G_\bullet$ if the following conditions hold:

(i) For each $j = 0, \ldots, m-1$, $\mathfrak{h}_j := \mathrm{Span}(\xi_{j+1}, \ldots, \xi_m)$ is a Lie algebra ideal in $\mathfrak{g}$, and hence $H_j := \exp(\mathfrak{h}_j)$ is a normal Lie subgroup of G;
(ii) For every $0 \leq i \leq s$, we have $G^{(i)} = H_{m-m_i}$;
(iii) Each $g \in G$ can be written uniquely as $\exp(t_1\xi_1)\exp(t_2\xi_2)\cdots\exp(t_m\xi_m)$ for $t_1, \ldots, t_m \in \mathbb{R}$;
(iv) Γ consists precisely of those elements which, when written in the above form, have all $t_i \in \mathbb{Z}$.

It follows from (iii) that the map

$$(t_1, \ldots, t_m) \mapsto \exp(t_1\xi_1)\cdots\exp(t_m\xi_m)$$

is a diffeomorphism from $\mathbb{R}^m$ onto G; the numbers $t_1, \ldots, t_m$ associated to an element $g \in G$ in this way are called the *coordinates of g in the basis $\mathcal{X}$*.

It can be shown that there exists a Mal'cev basis adapted to any rational filtration $G_\bullet$; see the remark following Proposition 2.1 in [30] which is based on [10, Proposition 5.3.2] and ultimately on [57].

4.2. **The metric on G and on X.** Let $\mathfrak{g}$ be endowed with the Euclidean structure making the Mal'cev basis $\mathcal{X}$ an orthonormal basis. This induces a Riemannian structure on G that is invariant under right translations. The group G is endowed with the corresponding geodesic distance, which we denote by d_G. This distance is invariant under right translations.[8]

[8]We remark that in [30] the authors use a different metric, but it is equivalent with d_G, and the implied constant depends only on X and the choice of the Mal'cev basis, so this does not make any difference for us.

Let the space $X := G/\Gamma$ be endowed with the quotient metric d_X. Writing $p\colon G \to X$ for the quotient map, the metric d_X is defined by

$$d_X(x,y) = \inf_{g,h\in G}\{d_G(g,h)\colon p(g)=x,\ p(h)=y\}.$$

Since Γ is discrete, it follows that the infimum is attained.

For $k\in\mathbb{N}$ and $\Phi\in\mathcal{C}^k(X)$, $\|\Phi\|_{\mathcal{C}^k(X)}$ denotes the usual $\mathcal{C}^k$-norm. We frequently use the fact that if Φ belongs to $\mathcal{C}^1(X)$, then $\|\Phi\|_{\mathrm{Lip}(X)} \le \|\Phi\|_{\mathcal{C}^1(X)}$. We also use some simple facts that follow immediately from the smoothness of the multiplication $G\times G\to G$.

Lemma 4.1. *Let F be a bounded subset of G. There exists a constant $C>0$ such that*

(i) *For every $g,h,h'\in F$ we have $d_G(gh,gh')\le Cd_G(h,h')$;*
(ii) *For every $x,x'\in X$ and $g\in F$ we have $d_X(g\cdot x, g\cdot x')\le Cd_X(x,x')$;*

Moreover, for every $k\in\mathbb{N}$ there exists a constant C_k such that

(iii) *For every $\Phi\in\mathcal{C}^k(X)$ and $g\in F$, writing $\Phi_g(x):=\Phi(g\cdot x)$, we have $\|\Phi_g\|_{\mathcal{C}^k(X)}\le C_k\|\Phi\|_{\mathcal{C}^k(X)}$.*

Lemma 4.2. *There exists $\delta>0$ such that, for $j=1,\dots,s$, if $\gamma\in\Gamma$ and $u\in G_j$ satisfy $d_G(\gamma,u)<\delta$, then $\gamma\in G_j$.*

Proof. This follows immediately from the classical fact that $G_j\cap\Gamma$ is co-compact in G_j. □

4.3. **Sub-nilmanifolds.** We proceed with some basic facts regarding sub-nilmanifolds.

Definition. A *sub-nilmanifold* of X is a nilmanifold $X' := G'/\Gamma'$ where G' is a rational subgroup of G and $\Gamma' := G'\cap\Gamma$. We constantly identify X' with the closed sub-nilmanifold $G'\cdot e_X$ of X. In particular, the base point $e_{X'}$ of X' is identified with the base point e_X of X.

Convention. If $X := G/\Gamma$ is a nilmanifold and G is endowed with a rational filtration $G_\bullet$ and if $X' := G'/(G'\cap\Gamma)$ is a sub-nilmanifold of X, then we implicitly assume that G' is endowed with the *induced rational filtration* defined by $G'^{(j)} := G'\cap G^{(j)}$, $j\in\mathbb{N}$.

In general, there is no natural method to define a Mal'cev basis for X' from a Mal'cev basis for X, and we cannot assume that the inclusion map $X'\to X$ is an isometry. However, this inclusion is a smooth embedding, and it follows that there exists a positive constant $C := C(X',X)$ such that

$$C^{-1}d_X(x,y)\le d_{X'}(x,y)\le Cd_X(x,y)\ \text{ for every } x,y\in X'. \tag{4.1}$$

4.4. **Vertical and horizontal torus and corresponding characters.** Let $X := G/\Gamma$ be an s-step nilmanifold, $m:=\dim(G)$ and $r:=\dim(G_s)$. The *vertical torus* is the connected compact Abelian Lie group $G_s/(G_s\cap\Gamma)$. Since the restriction to $G_s\cap\Gamma$ of the action of G on X is trivial, the vertical torus acts on X, and this action is clearly free. It follows from the definition of the distance on X that the vertical torus acts by isometries. Let $\widetilde{X}$ be the quotient of X under this action. Then $\widetilde{X}$ is an $(s-1)$-step nilmanifold and can be written as $\widetilde{X} := \widetilde{G}/\widetilde{\Gamma}$ where $\widetilde{G} := G/G_s$ and $\widetilde{\Gamma} := \Gamma/(\Gamma\cap G_s)$.

We endow $\widetilde{G}$ with a Mal'cev basis such that, in Mal'cev coordinates, the projection $G \to \widetilde{G}$ is given by $(t_1, \dots, t_m) \mapsto (t_1, \dots, t_{m-r})$. The distance $d_{\widetilde{G}}$ on $\widetilde{G}$ corresponding to this basis is the quotient distance induced by d_G, and the distance $d_{\widetilde{X}}$ on $\widetilde{X}$ is the quotient distance induced by d_X.

Furthermore, the Mal'cev basis of X induces an isometric identification between G_s and $\mathbb{R}^r$, and thus of the vertical torus endowed with the quotient metric, with $\mathbb{T}^r$ endowed with its usual metric. In order to avoid confusion, elements of G_s are written as $u, v, \dots$ when we use the multiplicative notation, and as $\mathbf{u} = (u_1, \dots, u_r)$, $\mathbf{v} = (v_1, \dots v_r), \dots$ when we identify G_s with $\mathbb{R}^r$ and use the additive notation; the same convention is used for the vertical torus.

Definition (Vertical characters and nilcharacters)**.** Let X be an s-step nilmanifold and $r := \dim(G_s)$. A *vertical character* is a continuous group homomorphism $\xi \colon G_s \to \mathbb{T}$ with a trivial restriction on $G_s \cap \Gamma$; it can also be thought of as a character of the vertical torus. The group of vertical characters is then identified with $\mathbb{Z}^r$, where $\mathbf{h} = (h_1, \dots, h_r) \in \mathbb{Z}^r$ corresponds to the group homomorphism ξ given by

$$\xi(\mathbf{t}) := \mathbf{h} \cdot \mathbf{t} \bmod 1 = h_1 t_1 + \cdots + h_r t_r \bmod 1 \quad \text{for} \quad \mathbf{t} = (t_1, \dots, t_r) \in \mathbb{R}^r = G_s.$$

We define the *norm of* ξ to be

$$\|\xi\| := \|\mathbf{h}\| = |h_1| + \cdots + |h_r|.$$

A function $\Phi \colon X \to \mathbb{C}$ is a *nilcharacter with frequency* $\mathbf{h}$ if $\Phi(\mathbf{t} \cdot x) = \mathrm{e}(\mathbf{h} \cdot \mathbf{t})\, \Phi(x)$ for every $\mathbf{t} \in G_s$ and every $x \in X$.[9]

Definition (Maximal torus and horizontal characters)**.** Let $X := G/\Gamma$ be an s-step nilmanifold, and let $m := \dim(G)$ and $m_2 := \dim(G_2)$. The Mal'cev basis induces an isometric identification between the *horizontal torus* $G/(G_2\Gamma)$, endowed with the quotient metric, and $\mathbb{T}^{m-m_2}$, endowed with its usual metric. A *horizontal character* is a continuous group homomorphism $\eta \colon G \to \mathbb{T}$ with a trivial restriction on Γ. In Mal'cev coordinates, it is given by $\eta(x_1, \dots, x_m) = \ell_1 x_1 + \cdots + \ell_{m-m_2} x_{m-m_2} \bmod 1$ for $(x_1, \dots, x_m) \in \mathbb{R}^m$, where $\ell_1, \dots, \ell_{m-m_2}$ are integers called the coefficients of η. We let

$$\|\eta\| := |\ell_1| + \cdots + |\ell_{m-m_2}|.$$

The horizontal character η factors through the horizontal torus and induces a character given by $\boldsymbol{\alpha} \mapsto \ell \cdot \boldsymbol{\alpha} := \ell_1 \alpha_1 + \cdots + \ell_{m-m_2} \alpha_{m-m_2}$ for $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{m-m_2}) \in \mathbb{T}^{m-m_2}$.

4.5. **The U^s-inverse theorem.** We are going to use the following inverse theorem of Green, Tao, and Ziegler that gives a criterion for checking that a function $a \colon \mathbb{Z}_N \to \mathbb{C}$ has U^s-norm bounded away from zero.

Theorem 4.3 (Inverse theorem for the U^s-norms [31, Theorem 1.3])**.** *Let $s \geq 2$ be an integer and ε be a positive real that is smaller than 1. There exist an $(s-1)$-step nilmanifold $X := G/\Gamma$ and $\delta > 0$, both depending on s and ε only, such that the following holds: For every $N \in \mathbb{N}$, if $a \colon \mathbb{Z}_N \to \mathbb{C}$ has modulus at most 1 and satisfies*

$$\|a\|_{U^s(\mathbb{Z}_N)} \geq \varepsilon,$$

[9] In [30] a function with this property is said to have vertical oscillation $\mathbf{h}$.

then there exist $g \in G$ and a function $\Phi\colon X \to \mathbb{C}$ with $\|\Phi\|_{\mathrm{Lip}(X)} \le 1$, such that

$$\big|\mathbb{E}_{n\in[N]} a(n)\,\Phi(g^n \cdot e_X)\big| \ge \delta.$$

There are two differences between this theorem and the form it is stated in [31]. First, the result is stated in [31] with a finite family of nilmanifolds instead of a single one; but as the authors of [31] also remark, one can use a single nilmanifold. More importantly, the result is stated for the norm $U^s[N]$ instead of the norm $U^s(\mathbb{Z}_N)$. The present statement follows immediately from the result in [31] and Lemma A.4 in the Appendix.

A sequence of the form $\Phi(g^n \cdot e_X)$ where Φ is only assumed to be continuous is called a *basic nilsequence* in [3]; if in addition we assume that Φ is Lipschitz, then we call it a *nilsequence of bounded complexity*, a notion first used in [27].

Let us remark that for the partition regularity results of Sections 2.4.1 and 2.4.2 we only need to use the U^3-inverse theorem; an independent and much simpler proof of this inverse theorem can be found in [27].

5. Quantitative equidistribution and factorization on nilmanifolds

In this section we state a quantitative equidistribution result and a factorization theorem for polynomial sequences on nilmanifolds, both proved by Green and Tao in [30], and also derive some consequences that will be used later on.

5.1. Polynomial sequences in a group.

We start with the definition of a polynomial sequence on an arbitrary group.

Definition. Let G be a group endowed with a filtration $G_\bullet$ and $(g(n))_{n\in\mathbb{N}}$ be a sequence in G. For $h \in \mathbb{N}$, we define the sequence $\partial_h g$ by $\partial_h g(n) := g(n+h)g(n)^{-1}$, $n \in \mathbb{N}$. We say that the sequence g is a polynomial sequence with coefficients in the filtration $G_\bullet$ if $\partial_{h_i} \cdots \partial_{h_1} g$ takes values in $G^{(i)}$ for every $i \in \mathbb{N}$ and $h_1, \dots, h_i \in \mathbb{N}$. We write $\mathrm{poly}(G_\bullet)$ for the family of polynomial sequences with coefficients in $G_\bullet$. If the filtration $G_\bullet$ has degree d, we say that the polynomial sequence has *degree at most d*.

The following equivalent definition is given in [30, Lemma 6.7] (see also [53, 54]).

Equivalent Definition. A polynomial sequence with coefficients in the filtration $G_\bullet$ of degree d is a sequence $(g(n))_{n\in\mathbb{N}}$ of the form

$$g(n) = a_0 a_1^n a_2^{\binom{n}{2}} \cdots a_d^{\binom{n}{d}} \quad \text{where} \quad a_j \in G^{(j)} \quad \text{for} \quad j = 0, \dots, d. \tag{5.1}$$

Remarks. (1) The extra flexibility coming from the fact that we consider polynomial sequences with respect to arbitrary filtrations, not just the natural one, will be used in an essential way.

(2) The set $\mathrm{poly}(G_\bullet)$ is a group with operation the pointwise multiplication of sequences [30, Proposition 6.2], a result initially due to Leibman [53, 54] when $G_\bullet$ is the natural filtration.

(3) It can be seen (see [30, Remarks below Corollary 6.8]) that if G is s-step nilpotent, then every sequence $g\colon \mathbb{N} \to G$ of the form $g(n) := a_1^{p_1(n)} \cdots a_k^{p_k(n)}$ with $a_1, \dots, a_k \in G$ and $p_1, \dots, p_k \in \mathbb{Z}[t]$ of degree at most d, is a polynomial sequence with coefficients in some filtration $G_\bullet$ of G of degree at most ds.

When $G = \mathbb{T}$, unless stated explicitly, we assume that $\mathbb{T}$ is endowed with the filtration of degree $d \in \mathbb{N}$ given by $\mathbb{T}^{(j)} = \mathbb{T}$ for $j \leq d$ and $\mathbb{T}^{(j)} = \{0\}$ for $j > d$. In this case, a polynomial sequence of degree at most d in $\mathbb{T}$ can be expressed alternatively in two different ways:

$$\phi(n) = \alpha_0 + \alpha_1\binom{n}{1} + \alpha_2\binom{n}{2} + \cdots + \alpha_d\binom{n}{d} \tag{5.2}$$

$$= \alpha'_0 + \alpha'_1 n + \alpha'_2 n^2 + \cdots + \alpha'_d n^d \tag{5.3}$$

for some $\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_d, \alpha'_0, \alpha'_1, \alpha'_2, \ldots, \alpha'_d \in \mathbb{T}$. The choice between these two representations depends on the problem at hand. Similar comments apply for polynomial sequences in $\mathbb{T}^m$.

Definition (Smoothness norms)**.** Let $(\phi(n))_{n\in\mathbb{N}}$ be a polynomial sequence of degree at most d in $\mathbb{T}$ of the form (5.2). For every $N \in \mathbb{N}$ we define the *smoothness norm*

$$\|\phi\|_{C^\infty[N]} := \max_{1\leq j\leq d} N^j \|\alpha_j\|,$$

where, as usual, $\|\alpha\|$ denotes the distance of α to the nearest integer.

If a polynomial sequence is given by (5.3), then we define

$$\|\phi\|'_{C^\infty[N]} := \max_{1\leq j\leq d} N^j \|\alpha'_j\|.$$

It is easy to check that there exist positive constants $c := c(d), C := C(d)$ such that

$$c\|\phi\|_{C^\infty[N]} \leq \|\phi\|'_{C^\infty[N]} \leq C\|\phi\|_{C^\infty[N]},$$

so the two norms can be used interchangeably without affecting our arguments.

The smoothness norm is designed to capture the concept of a slowly varying polynomial sequence. Indeed, for every $d \in \mathbb{N}$ there exists $C := C(d) > 0$ such that, for every polynomial sequence ϕ of degree d on $\mathbb{T}$ (or $\mathbb{T}^m$) and every $n \in [N]$, we have

$$\|\phi(n) - \phi(n-1)\| \leq \frac{C}{N}\|\phi\|_{C^\infty[N]}.$$

It is immediate to check that for $1 \leq N' \leq N$ and for every polynomial sequence ϕ of degree at most d, we have

$$\|\phi\|_{C^\infty[N']} \leq \|\phi\|_{C^\infty[N]} \leq \big(\frac{N}{N'}\big)^d \|\phi\|_{C^\infty[N']}.$$

We can also show that for $b \in \mathbb{Z}$ and $\phi_b(n) := \phi(n+b)$, we have

$$\|\phi_b\|_{C^\infty[N]} \leq \big(\frac{N+1}{N}\big)^{|b|} \|\phi\|_{C^\infty[N]}. \tag{5.4}$$

Indeed, let us write $\phi_b(n) = \sum_{j=0}^{d} \beta_j \binom{n}{j}$. By a direct computation, we get

$$\text{for } b \geq 0,\ \beta_i = \sum_{j=0}^{d-i} \binom{b}{j} \alpha_{i+j}\ ;\quad \text{for } b < 0,\ \beta_i = \sum_{j=0}^{d-i} (-1)^j \binom{-b-1}{j} \alpha_{i+j},$$

where, as usual, $\binom{n}{p} = 0$ for $p > n$. Hence, for $b \geq 0$ and $i = 1, \ldots, d$, we have

$$N^i\|\beta_i\| \leq N^i \sum_{j=0}^{d-i} \binom{b}{j} \|\alpha_{i+j}\| \leq \|\phi\|_{C^\infty[N]} \sum_{j=0}^{d-i} \binom{b}{j} \frac{1}{N^j} \leq \|\phi\|_{C^\infty[N]} \big(\frac{N+1}{N}\big)^b.$$

For $b < 0$ we get a similar estimate with $-b-1$ in place of b. In both cases the asserted estimate (5.4) follows immediately.

The next lemma is a modification of a particular case of [30, Lemma 8.4].

Lemma 5.1. *Let $d, q, r, N \in \mathbb{N}$ and a, b be integers with $a \neq 0$, $|a| \leq q$, and $|b| \leq rN$. There exist $C := C(d, q, r) > 0$ and $\ell := \ell(a, d) \in \mathbb{N}$ such that if $\phi\colon \mathbb{N} \to \mathbb{T}$ is a polynomial sequence of degree at most d and ψ is given by $\psi(n) := \phi(an+b)$, then*

$$\|\ell\phi\|_{C^\infty[N]} \leq C\|\psi\|_{C^\infty[N]}.$$

Proof. Writing $\phi_b(n) := \phi(n+b)$ and using (5.4) and that $|b| \leq rN$, we get

$$\|\phi\|_{C^\infty[N]} \leq \Big(\frac{N+1}{N}\Big)^{|b|}\|\phi_b\|_{C^\infty[N]} \leq C_1\|\phi_b\|_{C^\infty[N]}$$

for some $C_1 := C_1(r)$. Furthermore, since $\psi(n) = \phi_b(an)$, one easily checks that

$$\||a|^d\phi_b\|_{C^\infty[N]} \leq |a|^{d-1}\|\psi\|_{C^\infty[N]}.$$

Combining the above we get the asserted estimate for $\ell := |a|^d$ and $C := C_1q^d$. □

5.2. **The quantitative Leibman theorem.** We are going to work with the following notion of equidistribution on a nilmanifold.

Definition. Let $X := G/\Gamma$ be a nilmanifold, $N \in \mathbb{N}$, $(g(n))_{n\in[N]}$ be a finite sequence in G, and $\delta > 0$. The sequence $(g(n)\cdot e_X)_{n\in[N]}$ is *totally δ-equidistributed in X*, if for every arithmetic progression $P \subset [N]$ and for every Lipschitz function Φ on X with $\|\Phi\|_{\mathrm{Lip}(X)} \leq 1$ and $\int_X \Phi\, dm_X = 0$, we have

$$\big|\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)\,\Phi(g(n)\cdot e_X)\big| \leq \delta. \tag{5.5}$$

Remark. The distance on X, and as a consequence the notion of equidistribution of a sequence in X, depends on the choice of a Mal'cev basis on G, which in turn depends on the chosen rational filtration $G_\bullet$. As remarked in Section 4.3, if X' is a sub-nilmanifold of X, then there is no natural choice for the Mal'cev basis of X', and thus a sequence in X that is δ-equidistributed in X is only $(C\delta)$-equidistributed in X', where the constant C depends on the choice of the two Mal'cev bases.

To avoid confusion we remind the reader of the following convention that we make throughout the article.

Convention. If $X := G/\Gamma$ is a nilmanifold, G is implicitly endowed with some rational filtration $G_\bullet$. A polynomial sequence in G is always assumed to have coefficients in this filtration; that is, it belongs to $\mathrm{poly}(G_\bullet)$. As the degree of a polynomial sequence in G is bounded by d where d is the degree of $G_\bullet$, all statements below implicitly impose a restriction on the degree of the polynomial sequence under consideration.

The next result gives a convenient criterion for establishing equidistribution properties of polynomial sequences of nilmanifolds.

Theorem 5.2 (Quantitative Leibman Theorem [30, Theorem 2.9])**.** *Let $X := G/\Gamma$ be a nilmanifold and $\varepsilon > 0$. There exists $D := D(X, \varepsilon) > 0$ such that the following holds: For every $N \in \mathbb{N}$, if $g \in \mathrm{poly}(G_\bullet)$ and $(g(n)\cdot e_X)_{n\in[N]}$ is not totally ε-equidistributed in X, then there exists a non-trivial horizontal character η such that*

$$0 < \|\eta\| \leq D \quad \text{and} \quad \|\eta\circ g\|_{C^\infty[N]} \leq D.$$

Remarks. (1) For every $g \in \text{poly}(G_\bullet)$ and every horizontal character η, the sequence $\eta \circ g$ is a polynomial sequence in $\mathbb{T}$ of degree at most d, where d is the degree of $G_\bullet$.

(2) Theorem 5.2 will be used in this form, but it is proved in [30] under the stronger hypothesis that the sequence is not "*ε-equidistributed in X*," meaning (5.5) fails for $P := [N]$. We deduce Theorem 5.2 from this result next.

Proof. Since the sequence $(g(n) \cdot e_X)_{n \in [N]}$ is not totally ε-equidistributed in X, there exist an arithmetic progression $P \subset [N]$ and a Lipschitz function Φ on X such that

$$\|\Phi\|_{\text{Lip}(X)} \leq 1, \quad \int_X \Phi \, dm_X = 0, \quad \text{and} \quad \big|\mathbb{E}_{n \in [N]} \mathbf{1}_P(n)\, \Phi(g(n) \cdot e_X)\big| \geq \varepsilon.$$

We write $P = \{an + b \colon n \in [N']\}$ where N' is the length of P, a is its step, and $b \in [N]$. Note that one necessarily has $N' \geq \varepsilon N$, and thus $a \leq 1/\varepsilon$. Then

$$\big|\mathbb{E}_{n \in [N']} \Phi(h(n) \cdot e_X)\big| \geq \varepsilon,$$

where $h(n) := g(an + b)$. Hence, the sequence $(h(n) \cdot e_X)_{n \in [N']}$ is not ε-equidistributed in X. Note also that $h \in \text{poly}(G_\bullet)$; this follows from the first definition in Section 5.1. Using the variant of Theorem 5.2 that is proved in [30], we deduce that there exists $D := D(X, \varepsilon) > 0$ and a non-trivial horizontal character θ such that $\|\theta\| \leq D$ and $\|\theta \circ h\|_{C^\infty[N']} \leq D$. Writing $\phi(n) := \theta(g(n))$ and $\psi(n) := \theta(h(n))$, we have $\psi(n) = \phi(an + b)$. Using Lemma 5.1 with $q := 1/\varepsilon$ and $r := 1$, we get that there exist $C := C(d, \varepsilon) > 0$ and $\ell := \ell(a, d) \in \mathbb{N}$ such that

$$\|\ell \cdot \theta \circ g\|_{C^\infty[N]} = \|\ell \phi\|_{C^\infty[N]} \leq C \|\psi\|_{C^\infty[N]} \leq C \Big(\frac{N}{N'}\Big)^d \|\psi\|_{C^\infty[N']} \leq C \varepsilon^{-d} D,$$

where d is the degree of the filtration $G_\bullet$. Letting $\eta := \ell\, \theta$ we have $\|\eta\| \leq D\ell$ and the result follows. $\square$

We are also going to use frequently the following converse of Theorem 5.2.

Lemma 5.3 (A converse to Theorem 5.2). *Let $X := G/\Gamma$ be a nilmanifold. There exists $c := c(X) > 0$ such that for every $D \in \mathbb{N}$ and every sufficiently large $N \in \mathbb{N}$, depending only on D and X, the following holds: If $g \in \text{poly}(G_\bullet)$ and there exists a non-trivial horizontal character η of X with $\|\eta\| \leq D$ and $\|\eta \circ g\|_{C^\infty[N]} \leq D$, then the sequence $(g(n) \cdot e_X)_{n \in [N]}$ is not totally (cD^{-2})-equidistributed in X.*

Proof. Let d be the degree of the filtration $G_\bullet$. Since $\|\eta \circ g\|_{C^\infty[N]} \leq D$, we have

$$\eta(g(n)) = \sum_{0 \leq j \leq d} \alpha_i \binom{n}{j} \quad \text{for some } \alpha_0, \ldots, \alpha_d \in \mathbb{T} \quad \text{with } \|\alpha_j\| \leq \frac{D}{N^j}, \qquad \text{for } j = 1, \ldots, d,$$

$$\big|\mathrm{e}\big(\eta(g(n))\big) - \mathrm{e}(\alpha_0)\big| \leq \frac{1}{2}, \quad \text{for} \quad 1 \leq n \leq c_1 \frac{N}{D},$$

for some positive constant $c_1 := c_1(d)$. Suppose that $N \geq 4D/c_1$. Then

$$\big|\mathbb{E}_{n \leq \lfloor c_1 N/D \rfloor} \mathrm{e}\big(\eta(g(n))\big)\big| \geq \frac{1}{2},$$

which gives

$$\left|\mathbb{E}_{n\in[N]}\mathbf{1}_{[\lfloor c_1N/D\rfloor]}(n)\mathrm{e}\big(\eta(g(n))\big)\right| \geq \frac{c_1}{2D} - \frac{1}{N} \geq \frac{c_1}{4D}.$$

Furthermore, since $\|\eta\| \leq D$, the function $x \mapsto \mathrm{e}(\eta(x))$, defined on X, is Lipschitz with constant at most C_1D for some $C_1 := C_1(X)$ and has integral 0 since η is a non-trivial horizontal character. Therefore, the sequence $(g(n)\cdot e_X)_{n\in[N]}$ is not totally (cD^{-2})-equidistributed in X where $c := c_1/(4C_1)$, completing the proof. □

5.3. **Some consequences of the quantitative Leibman theorem.** We give two corollaries that are going to be used in subsequent sections. We caution the reader that in both statements the polynomiality of the sequence and the quantitative Leibman theorem are used in a crucial way.

Corollary 5.4. *Let $X := G/\Gamma$ be a nilmanifold and Γ' be a discrete subgroup of G containing Γ. Let $X' := G/\Gamma'$ and suppose that G is endowed with the same rational filtration $G_\bullet$ for both nilmanifolds X and X'. For every $\varepsilon > 0$ there exists $\delta := \delta(X, X', \varepsilon) > 0$ such that the following holds: For every sufficiently large $N \in \mathbb{N}$, depending only on X, X', ε, if $g \in \mathrm{poly}(G_\bullet)$ and $(g(n)\cdot e_X)_{n\in[N]}$ is totally δ-equidistributed in X', then $(g(n)\cdot e_X)_{n\in[N]}$ is totally ε-equidistributed in X.*

Proof. Since Γ is co-compact and Γ' is closed in G, Γ is co-compact in Γ'; since Γ' is discrete, Γ has a finite index in Γ'. It follows that the natural projection $X \to X'$ is finite to one and there exists an $\ell \in \mathbb{N}$, depending on X and X', such that $\gamma^\ell \in \Gamma$ for every $\gamma \in \Gamma'$. Therefore, for every horizontal character η of X (meaning a group homomorphism $G \to \mathbb{T}$ with a trivial restriction to Γ), η^ℓ has a trivial restriction to Γ' and thus is a horizontal character of X'.

Suppose now that the sequence $(g(n)\cdot e_X)_{n\in[N]}$ is not totally ε-equidistributed in X. By Theorem 5.2 there exist $D := D(X,\varepsilon)$ and a horizontal character η of X with $\|\eta\| \leq D$ and $\|\eta\circ g\|_{C^\infty[N]} \leq D$. Then η^ℓ is a horizontal character of X' such that $\|\eta^\ell\| \leq C\ell\|\eta\| \leq C\ell D$ for some $C := C(X, X')$[10] and $\|\eta^\ell \circ g\|_{C^\infty[N]} \leq \ell D$. Lemma 5.3 then provides a $\delta := \delta(X, X', \varepsilon) > 0$ such that the sequence $(g(n)\cdot e_X)_{n\in[N]}$ is not totally δ-equidistributed in X'. This completes the proof. □

Properties of rational elements are given in Appendix B, and we only recall here that an element g of G is rational if $g^n \in \Gamma$ for some $n \in \mathbb{N}$.

Corollary 5.5. *Let $X := G/\Gamma$ be a nilmanifold and G' be a rational subgroup of G. Let $X' := G'/(G'\cap\Gamma)$, α be a rational element of G, $G'_\alpha := \alpha^{-1}G'\alpha$, and $X'_\alpha := G'_\alpha/(G'_\alpha\cap\Gamma)$. Then there exists a function $\rho_{X,X',\alpha}\colon \mathbb{R}_+ \to \mathbb{R}_+$ with $\rho_{X,X',\alpha}(t) \to 0$ as $t \to 0^+$ such that the following holds: For every sufficiently large $N \in \mathbb{N}$, depending only on X, X', α, if $h \in \mathrm{poly}(G'_\bullet)$ and $(h(n)\cdot e_X)_{n\in[N]}$ is totally t-equidistributed in X', then $(\alpha^{-1}h(n)\alpha\cdot e_X)_{n\in[N]}$ is totally $\rho_{X,X',\alpha}(t)$-equidistributed in X'_α.*

Remark. Recall that since G' is a rational subgroup of G, $G'\cap\Gamma$ is co-compact in G' and X' is identified with the sub-nilmanifold $G'\cdot e_X$ of X. By Lemma B.4, G'_α is also a rational subgroup of G and thus $G'_\alpha\cap\Gamma$ is co-compact in G'_α and $X'_\alpha = G'_\alpha\cdot e_X$. Furthermore, we have $h_\alpha \in \mathrm{poly}(G'_{\alpha\bullet})$ where $h_\alpha(n) := \alpha^{-1}h(n)\alpha$ and $G'_{\alpha\bullet} := \alpha^{-1}G'_\bullet\alpha$.

[10]The constant C arises from the fact that the identifications $G/(G_2\Gamma) = \mathbb{T}^{m-m_2}$ and $G/(G_2\Gamma') = \mathbb{T}^{m-m_2}$ are different.

Proof. To ease notation, in this proof we leave the dependence on X, X', α implicit.

We start by using Lemma B.6 in the Appendix, it gives that $G'\cap\Gamma\cap(\alpha^{-1}\Gamma\alpha)$ has a finite index in the two groups $G'\cap\Gamma$ and $G'\cap(\alpha^{-1}\Gamma\alpha)$. In particular, $G'\cap\Gamma\cap(\alpha^{-1}\Gamma\alpha)$ is discrete and co-compact in G'. We let $\widetilde{X}_\alpha := G'/(G'\cap\Gamma\cap(\alpha^{-1}\Gamma\alpha))$.

By Corollary 5.4, there exists a function $\psi\colon \mathbb{R}_+\to\mathbb{R}_+$ with $\psi(t)\to 0$ as $t\to 0^+$, such that the following holds: If N is sufficiently large and the polynomial sequence $(h(n))_{n\in[N]}$ in G' is such that the sequence $(h(n)\cdot e_X)_{n\in[N]}$ is totally t-equidistributed in X', then the sequence $(h(n)\cdot e_{\widetilde{X}_\alpha})_{n\in[N]}$ is totally $\psi(t)$-equidistributed in $\widetilde{X}_\alpha$.

Furthermore, $G'\cap\alpha^{-1}\Gamma\alpha$ is discrete and co-compact in G', and we let $X''_\alpha := G'/(G'\cap(\alpha^{-1}\Gamma\alpha))$. The natural projection $\widetilde{X}_\alpha\to X''_\alpha$ is Lipschitz with Lipschitz constant C. If the sequence h is as above, then the image of the sequence $(h(n)\cdot e_{\widetilde{X}_\alpha})_{n\in[N]}$ under this projection is $(h(n)\cdot e_{X''_\alpha})_{n\in[N]}$ and this sequence is totally $(C\psi(t))$-equidistributed in X''_α.

The conjugacy map $g\mapsto\alpha^{-1}g\alpha$ is an isomorphism from G' onto G'_α and maps $\alpha^{-1}\Gamma\alpha$ onto Γ. Thus, it induces a diffeomorphism from the nilmanifold X''_α onto X'_α. If the Mal'cev basis of X''_α is chosen so that its image under the conjugacy is the Mal'cev basis of X'_α, this diffeomorphism is an isometry. Then the image of the sequence $(h(n)\cdot e_{X''_\alpha})_{n\in[N]}$ under this isometry is $(\alpha^{-1}h(n)\alpha\cdot e_X)_{n\in[N]}$, and thus this last sequence is totally $(C\psi(t))$-equidistributed in X'_α. This completes the proof. □

5.4. **The factorization theorem** [30]**.** Recall that a nilmanifold $X := G/\Gamma$ is implicitly endowed with a filtration $G_\bullet$. This defines a Mal'cev basis for X which in turn is used to define a metric on X. Following our conventions, every sub-nilmanifold $X' := G'/(G'\cap\Gamma)$ is endowed with the filtration $G'_\bullet$ given by $G'^{(j)} := G'\cap G^{(j)}$ for every $j\in\mathbb{N}$.

Definition. Let $M,N\in\mathbb{N}$.

(i) An element $\gamma\in G$ is *M-rational* if $\gamma^m\in\Gamma$ for some integer m with $1\le m\le M$; for more properties of rational elements see Appendix B.
(ii) A finite sequence $(g(n))_{n\in[N]}$ in G is *M-rational* if all its terms are M-rational.
(iii) A finite sequence $(\epsilon(n))_{n\in[N]}$ in G is *(M,N)-smooth* if $d_G(1_G,\epsilon(n))\le M$ for every $n\in[N]$ and $d_G(\epsilon(n),\epsilon(n+1))\le M/N$ for every $n\in[N-1]$.

The next result of Green and Tao [30] is used multiple times subsequently.

Theorem 5.6 (Factorization of polynomial sequences [30, Theorem 1.19])**.** *Let $X := G/\Gamma$ be a nilmanifold. For every $M\in\mathbb{N}$ there exists a finite family $\mathcal{F}(M)$ of sub-nilmanifolds of X that increase with M, each of the form $X' := G'/\Gamma'$, where G' is a rational subgroup of G and $\Gamma' := G'\cap\Gamma$, such that the following holds: For every function $\omega\colon\mathbb{N}\to\mathbb{R}^+$ there exists a positive integer $M_1 := M_1(X,\omega)$, and for every $N\in\mathbb{N}$ and every polynomial sequence $g\in\mathrm{poly}(G_\bullet)$, there exist $M\in\mathbb{N}$ with $M\le M_1$, a sub-nilmanifold $X'\in\mathcal{F}(M)$, and a factorization*

$$g(n)=\epsilon(n)g'(n)\gamma(n),\qquad n\in[N],$$

where $\epsilon, g', \gamma \in \operatorname{poly}(G_\bullet)$ *and*

(i) $\epsilon\colon [N] \to G$ *is* (M,N)*-smooth;*

(ii) $g' \in \operatorname{poly}(G'_\bullet)$ *and the sequence* $(g'(n)\cdot e_X)_{n\in[N]}$ *is totally* $\omega(M)$*-equidistributed in* X' *with the metric* $d_{X'}$ *induced by the filtration* $G'_\bullet$*;*

(iii) $\gamma\colon [N] \to G$ *is* M*-rational and* $(\gamma(n)\cdot e_X)_{n\in[N]}$ *has period at most* M*.*

Remarks. (1) In [30] this result is stated only for a function ω of the form $\omega(M) = M^A$ for some $A > 0$, but the same argument shows that it holds for all functions $\omega\colon \mathbb{N} \to \mathbb{R}^+$ (in fact, this more general statement is proven in [28, Lemma 2.10]).

(2) In [30], the family $\mathcal{F}(M)$ is defined as the collection of all sub-nilmanifolds $X' := G'/(G'\cap\Gamma)$ of X admitting a Mal'cev basis that consists of M-rational combinations of the elements of the Mal'cev basis $\mathcal{X}$ of X. We do not need this precise description of $\mathcal{F}(M)$; what is important for us is that each $\mathcal{F}(M)$ is finite and the family $(\mathcal{F}(M))_{M\in\mathbb{N}}$ depends only on X (and the filtration $G_\bullet$). The number M_1 in Theorem 5.6 corresponds to the quantity written as $M_0^{O_{A,m,d}(1)}$ in [30, Theorem 1.19]. In our case, what is important is that M_1 depends only on X and on ω.

(3) Simple examples (see [30, Section 1]) show that if $G_\bullet$ is the natural filtration in G, then the filtration $G'_\bullet$ may not be the natural one in G'. Furthermore, even if we start with a "linear" sequence $(g^n)_{n\in[N]}$ in G, we may end up with a "quadratic" sequence $(h^{n^2})_{n\in[N]}$ in G' where $h \neq \mathrm{id}_G$.

5.5. **Some constructions related to the factorization theorem.** We record here some terminology and constructions related to Theorem 5.6 that are used multiple times in the sequel. Throughout, we assume that a nilmanifold $X := G/\Gamma$ with a rational filtration $G_\bullet$ is given; then Theorem 5.6 provides for any given $M \in \mathbb{N}$ a finite family of sub-nilmanifolds $\mathcal{F}(M)$ of X that increases with M.

By Corollary B.3, for every $M \in \mathbb{N}$ there exists a finite subset $\Sigma(M)$ of G, consisting of M-rational elements, such that every M-rational element $\beta \in G$ can be written as $\beta = \alpha\alpha'$ with $\alpha \in \Sigma(M)$ and $\alpha' \in \Gamma$. We can assume that $\mathbf{1}_G \in \Sigma(M)$. Let $X' := G'/(G'\cap\Gamma)$ be a nilmanifold belonging to the family $\mathcal{F}(M)$, and let α belong to the finite set $\Sigma(M)$. By Lemma B.4, $\alpha^{-1}G'\alpha$ is a rational subgroup of G; we let

$$G'_\alpha := \alpha^{-1}G'\alpha \quad\text{and}\quad X'_\alpha := G'_\alpha/(G'_\alpha\cap\Gamma) = G'_\alpha\cdot e_X.$$

Recall that G' and each group G'_α is endowed with the induced filtration given by $G'^{(i)}_\alpha := G^{(i)}\cap G'_\alpha$. Since $G^{(i)}$ is a normal subgroup of G, we have $G'^{(i)}_\alpha = \alpha^{-1}G'^{(i)}\alpha$. Finally, let

$$\mathcal{F}'(M) := \{X'_\alpha \colon X' \in \mathcal{F}(M),\ \alpha \in \Sigma(M)\}.$$

5.5.1. *Defining the constant* $H(X,M)$*.* By Lemma 4.1, there exists a positive integer $H := H(X,M)$ with the following properties:

(i) For every $\alpha \in \Sigma(M)$ and for every $g \in G$ with $d_G(g, 1_G) \le M$, we have $d_G(\alpha^{-1}g\alpha, 1_G) \le H d_G(g, 1_G)$;

(ii) For every $\alpha \in \Sigma(M)$, every $g \in G$ with $d_G(g, \mathbf{1}_G) \le M$, and every $x, y \in X$, we have $d_X(g\alpha\cdot x, g\alpha\cdot y) \le H d_X(x,y)$;

(iii) Therefore, for every $\Phi \in \operatorname{Lip}(X)$, every $\alpha \in \Sigma(M)$, and every $g \in G$ with $d_G(g, \mathbf{1}_G) \le M$, we have $\|\Phi_{g\alpha}\|_{\operatorname{Lip}(X)} \le H\|\Phi\|_{\operatorname{Lip}(X)}$ where $\Phi_{g\alpha}(x) := \Phi(g\alpha\cdot x)$.

Note that the distance on a nilmanifold $X'_\alpha \in \mathcal{F}'(M)$ is not the distance induced by its inclusion in X. However, the inclusion $X'_\alpha \subset X$ is a smooth embedding, and thus we can assume that

(iv) For every nilmanifold $X'_\alpha \in \mathcal{F}'(M)$ and for every $x, y \in X'_\alpha$, we have

$$H^{-1} d_{X'_\alpha}(x,y) \le d_X(x,y) \le H d_{X'_\alpha}(x,y).$$

By Corollary 5.5, there exists a function $\rho_X \colon \mathbb{N} \times \mathbb{R}_+ \to \mathbb{R}_+$, such that $\rho_X(M,t)$ decreases to 0 as $t \to 0^+$ and M is fixed, and satisfies the following:

(v) For every nilmanifold $X' := G'/(G' \cap \Gamma) \in \mathcal{F}(M)$, for every $\alpha \in \Sigma(M)$, and for every $t > 0$, if $N \in \mathbb{N}$ is sufficiently large depending on X, M, and t, and if $(h(n))_{n\in[N]}$ is a polynomial sequence in G' such that the sequence $(h(n) \cdot e_X)_{n\in[N]}$ is totally t-equidistributed in X', then the sequence $(\alpha^{-1} h(n) \alpha)_{n\in\mathbb{N}}$ belongs to $\text{poly}(G'_{\alpha\bullet})$ and $(\alpha^{-1} h(n)\alpha \cdot e_X)_{n\in[N]}$ is totally $\rho_X(M,t)$-equidistributed in X'_α.

5.5.2. *Correlation estimates and factorization.* Let $N \in \mathbb{N}$, and suppose that the sequence $g \colon \mathbb{N} \to G$ factorizes as

$$g(n) = \epsilon(n) g'(n) \gamma(n), \quad n \in [N],$$

where ϵ is (M,N)-smooth, $g' \colon \mathbb{N} \to G$ is an arbitrary sequence, γ is M-rational, and $(\gamma(n) \cdot e_X)_{n\in[N]}$ has period at most M. Note that we do not assume that the sequence g is polynomial and we do not impose an equidistribution assumption on g'.

We are often given a correlation estimate of the form

$$|\mathbb{E}_{n\in[N]} a(n)\, \Phi(g(n) \cdot e_X)| \ge \delta, \tag{5.6}$$

for some $\delta \in (0,1)$, $a \colon [N] \to \mathbb{C}$ bounded by 1, and $\Phi \in \text{Lip}(X)$ with $\|\Phi\|_{\text{Lip}(X)} \le 1$. We want to deduce a similar estimate with $g'(n)$, or some other closely related sequence, in the place of $g(n)$. We do this as follows.

Let

$$L := \lfloor N \frac{\delta}{16\, H^2 M^2} \rfloor, \qquad N_1 := \lceil 16\, H^2 M^2 / \delta \rceil, \tag{5.7}$$

where $H := H(M,X)$ satisfies Properties (i)–(iv) in Section 5.5.1. Henceforth, we assume that $N \ge N_1$, then $L \ge 1$ and

$$N \frac{\delta}{32\, H^2 M^2} \le L \le N \frac{\delta}{16\, H^2 M^2}. \tag{5.8}$$

Since $\delta \le 1$ and $H \ge 1$, we have that

$$2ML \le N.$$

Let K be the (least) period of the sequence $(\gamma(n) \cdot e_X)_{n\in[N]}$, and let $P \subset [N]$ be an arithmetic progression of step K and the length between L and $2L$. Recall that $K \le M$. Let n_0 be the smallest element of the progression P. We write

$$\gamma(n_0) = \alpha\alpha' \;\text{ where }\; \alpha \in \Sigma(M) \;\text{ and }\; \alpha' \in \Gamma; \tag{5.9}$$

$$g'_\alpha(n) := \alpha^{-1} g'(n) \alpha \;\text{ for }\; n \in [N]; \tag{5.10}$$

$$\Phi_\alpha(x) := \Phi(\epsilon(n_0)\alpha \cdot x) \;\text{ for }\; x \in X. \tag{5.11}$$

For every $n \in P$ we have

$$\gamma(n) \cdot e_X = \gamma(n_0) \cdot e_X = \alpha\alpha' \cdot e_X;$$

hence

$$\Phi(g(n)\cdot e_X)=\Phi_\alpha\big(\alpha^{-1}\epsilon(n_0)^{-1}\epsilon(n)\,\alpha\,g'_\alpha(n)\,\alpha'\cdot e_X\big).$$

Furthermore, for $n\in P$ we have $n-n_0=jK$ for some j with $0\leq j<2L$ and thus $0\leq n-n_0\leq 2LK\leq 2LM$. Since the sequence ϵ is (M,N)-smooth, we have $d_G(\epsilon(n_0)^{-1}\epsilon(n),1_G)\leq 2M^2L/N$, and by Property (i) above we get

$$d_G(\alpha^{-1}\epsilon(n_0)^{-1}\epsilon(n)\alpha,1_G)\leq 2HM^2L/N.$$

Property (iii) gives that

$$\|\Phi_\alpha\|_{\mathrm{Lip}(X)}\leq H. \tag{5.12}$$

Combining the above, we get for every $n\in P$ that

$$\big|a(n)\,\Phi(g(n)\cdot e_X)-a(n)\,\Phi_\alpha(g'_\alpha(n)\cdot e_X)\big|\leq 2H^2M^2\frac{L}{N}.$$

Averaging on $[N]$, we get (recall that P has at most $2L$ elements)

$$\begin{aligned}\big|\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)\,a(n)\,\Phi(g(n)\cdot e_X)-\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)\,a(n)\,\Phi_\alpha(g'_\alpha(n)\cdot e_X)\big|\\ \leq 4H^2M^2\big(\frac{L}{N}\big)^2\leq\frac{\delta^2}{64\,H^2M^2}.\end{aligned} \tag{5.13}$$

Since $N\geq N_1$, we have $L\geq 1$ and (5.8) holds. Since $2MK\leq 2ML\leq N$, we can partition the interval $[N]$ into arithmetic progressions of step K and length between L and $2L$. The number of these progressions is bounded by N/L, and it follows from (5.6) that for one of them, say for P_1, we have

$$\big|\mathbb{E}_{n\in[N]}\mathbf{1}_{P_1}(n)\,a(n)\,\Phi(g(n)\cdot e_X)\big|\geq\delta\frac{L}{N}.$$

Estimate (5.13) gives

$$\big|\mathbb{E}_{n\in[N]}\mathbf{1}_{P_1}(n)\,a(n)\,\Phi_\alpha(g'_\alpha(n)\cdot e_X)\big|\geq\delta\frac{L}{N}-\frac{\delta^2}{64\,H^2M^2}\geq\frac{\delta^2}{64\,H^2M^2}, \tag{5.14}$$

where the last inequality follows from (5.8).

6. Minor arc nilsequences-preparatory work

In this section and the next one our goal is to show that bounded multiplicative functions have small correlation with all minor-arc nilsequences, that is, nilsequences that arise from totally equidistributed polynomial sequences on nilmanifolds. This result is a central point in the proof of Theorems 1.1 and 2.1, and it can be viewed as the higher order analogue of Corollary 3.2; linear sequences arising from numbers that are not well approximated by rationals with small denominators correspond to totally equidistributed polynomial sequences on nilmanifolds.

We remind the reader that for every nilmanifold $X:=G/\Gamma$, G is endowed with a rational filtration $G_\bullet$ and that polynomial sequences in G are assumed to have coefficients in this filtration and in particular have degree bounded by the degree of the filtration.

Theorem 6.1 (Key discorrelation estimate)**.** *Let $X:=G/\Gamma$ be a nilmanifold and $\tau>0$. There exist $\sigma:=\sigma(X,\tau)>0$ and $N_0:=N_0(X,\tau)$ such that for every $N\geq N_0$ the following holds: Suppose that $g\in\mathrm{poly}(G_\bullet)$ and*

$$|\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)\,f(n)\,\Phi(g(n+k)\cdot e_X)|\geq\tau, \tag{6.1}$$

for some integer $k \in [-N, N]$, $f \in \mathcal{M}$, $\Phi\colon X \to \mathbb{C}$ with $\|\Phi\|_{\mathrm{Lip}(X)} \leq 1$ and $\int \Phi \, dm_X = 0$, and arithmetic progression P in $[N]$. Then

$$\text{the sequence } (g(n) \cdot e_X)_{n \in [N]} \text{ is not totally } \sigma\text{-equidistributed in } X. \tag{6.2}$$

Recall that any sequence $g(n) = a_1^{p_1(n)} \cdots a_d^{p_d(n)}$, where $a_1, \ldots, a_d \in G$ and $p_1, \ldots, p_d \in \mathbb{Z}[t]$, belongs to $\mathrm{poly}(G_\bullet)$ for some appropriately chosen rational filtration $G_\bullet$. So Theorem 6.1 applies to all such sequences, and we get Theorem 2.2 as a direct consequence.

6.1. **Main ideas of the proof.** Before moving to the rather delicate details used in establishing Theorem 6.1, we explain the skeleton of the proof for a variant of this result that contains some key ideas used in the proof of Theorem 6.1 and suppresses several technicalities that obscure understanding. The main source of simplification in the sketch given below comes from the infinite nature of the problem and the restriction to linear polynomial sequences.

Suppose that we seek to prove the following infinitary result: If $X := G/\Gamma$ is an s-step nilmanifold, with G_s non-trivial, and for some $a \in G$ the sequence $(a^n \cdot e_X)_{n \in \mathbb{N}}$ is totally equidistributed in X, then for every $\Phi \in C(X)$ with $\int \Phi \, dm_X = 0$ we have

$$\lim_{N \to +\infty} \sup_{f \in \mathcal{M}} |\mathbb{E}_{n \in [N]} f(n) \, \Phi(a^n \cdot e_X)| = 0. \tag{6.3}$$

Using a vertical Fourier decomposition we reduce matters to the case where Φ is a nilcharacter of X with non-zero frequency. The orthogonality criterion of Kátai shows that in order to prove (6.3) it suffices to show that for all distinct $p, q \in \mathbb{N}$ we have

$$\lim_{N \to +\infty} \mathbb{E}_{n \in [N]} \Phi(a^{pn} \cdot e_X) \cdot \overline{\Phi}(a^{qn} \cdot e_X) = 0. \tag{6.4}$$

This motivates the study of equidistribution properties of the sequence

$$\big((a^{pn} \cdot e_X, a^{qn} \cdot e_X)\big)_{n \in \mathbb{N}}. \tag{6.5}$$

Let us take for granted that this sequence is equidistributed on a sub-nilmanifold $Y := H/\Delta$ of $X \times X$. We will be done if we manage to show that

$$\int_Y (\Phi \otimes \overline{\Phi}) \, dm_Y = 0. \tag{6.6}$$

This seemingly simple task turns out to be quite challenging, as the explicit structure of the possible nilmanifolds Y seems very difficult to determine (we have only managed to do this in the case $s = 2$).[11] Nevertheless, it is possible to extract some partial information about the group H defining the nilmanifold Y that suffices for our purposes. To do this, as a first step, we studyequidistribution properties of the

[11]It is tempting to believe that $Y := H/\Delta$ where $H := \{(g^p u_1, g^q u_2) \colon g \in G, u_1, u_2 \in G_2\}$, $\Delta := H \cap (\Gamma \times \Gamma)$. But this fails even for the simplest non-Abelian nilmanifolds. For example, let G be the Heisenberg group, meaning $G := \mathbb{R}^3$ with multiplication given by the formula $(x, y, z) \cdot (x', y', z') := (x + x', y + y'; z + z' + xy')$. Let $\Gamma := \mathbb{Z}^3$, $X := G/\Gamma$, and $a := (\alpha, \beta, 0)$ with α and β rationally independent. It is known that the sequence $(a^n \cdot e_X)_{n \in \mathbb{N}}$ is totally equidistributed in X. But if α, β, and $\alpha\beta$ are rationally dependent, then for distinct integers p, q the sequence $(a^{pn} \cdot e_X, a^{qn} \cdot e_X)_{n \in \mathbb{N}}$ turns out to be equidistributed in a sub-nilmanifold of $X \times X$ that is strictly "smaller" than Y.

projection of the sequence (6.5) to the horizontal torus of the nilmanifold $X \times X$. Using the total equidistribution assumption for the sequence $(a^n \cdot e_X)_{n\in\mathbb{N}}$, we get the following inclusion:

$$\{(g^p, g^q) : g \in G\} \subset H \cdot (G_2 \times G_2). \tag{6.7}$$

Taking iterated commutators of the elements on the left we deduce the following key property:

$$(u^{p^s}, u^{q^s}) \in H \quad \text{for every} \quad u \in G_s. \tag{6.8}$$

This easily implies that the function $\Phi \otimes \overline{\Phi}$ is a nilcharacter of Y with non-zero frequency, and hence it satisfies the sought-after zero integral property stated in (6.6).

When the place of a^n in (6.3) takes the general polynomial sequence $g(n)$, the first few steps of our argument remain the same. One important difference is that the inclusion (6.7) fails in general and it has to be replaced with a more complicated one. For instance, suppose that $g(n) := a^n b^{n^2}$ for some $a, b \in G$. Then we show that there exist normal subgroups G^1, G^2 of G such that $G^1 \cdot G^2 = G$ and

$$\{(g_1^p g_2^{p^2}, g_1^q g_2^{q^2}) : g_1 \in G^1, g_2 \in G^2\} \subset H \cdot (G_2 \times G_2).$$

By taking iterated commutators of elements on the left we establish a property analogous to (6.8), namely,

$$\text{the set} \;\; U := \{u \in G_s : (u^{p^j}, u^{q^j}) \in H \text{ for some } j \in \mathbb{N}\} \;\; \text{generates} \;\; G_s.$$

From this we can again extract the sought-after property (6.6).

This gives a rather accurate summary of the skeleton of the proof of Theorem 6.1 in the idealized setting of infinitary mathematics. Unfortunately, the natural habitat of Theorem 6.1, is the world of finitary mathematics, and this adds a serious amount of technical complexity in the implementation of the previous plan. In Section 7 we state and prove two key ingredients needed in the proof of Theorem 6.1; these are Propositions 6.5 and 6.6. In the next section we combine these ingredients in order to implement the previously sketched plan and finish the proof of Theorem 6.1.

6.2. **Notation and conventions.** In this section and the next one, d and m are positive integers representing the dimension of a torus $\mathbb{T}^m$ and the degree of a polynomial sequence on $\mathbb{T}^m$, respectively. Furthermore, p, q are distinct positive integers. We consider all these integers as fixed throughout, and we stress that all other parameters introduced in this section and the next one depend implicitly on d, m, p, q. This is not going to create problems for us, as in the course of proving Theorem 6.1 the integers d, m, p, q can be taken to be bounded by a constant that depends on X and τ only.

We continue to represent elements of $\mathbb{T}^m$ and $\mathbb{Z}^m$ by bold letters. For $\mathbf{x} \in \mathbb{T}^m$ and $\mathbf{h} \in \mathbb{Z}^m$ we write $\|\mathbf{x}\|$ for the distance of $\mathbf{x}$ from $\mathbf{0}$, $\|\mathbf{h}\| = |h_1| + \cdots + |h_m|$, and $\mathbf{h} \cdot \mathbf{x} = h_1 x_1 + \cdots + h_m x_m$. Vectors consisting of d elements of $\mathbb{T}^m$ are written $(\mathbf{g}_1, \ldots, \mathbf{g}_d)$. Beware of the possible confusion between $\mathbf{g}_j \in \mathbb{T}^m$ and g_j representing the jth-coordinate of $\mathbf{g} \in \mathbb{T}^m$. Sequences in $\mathbb{T}^m$ are written as $\mathbf{g}(n)$.

For finite sequences in $\mathbb{T}^m$ we use a slightly modified definition of smoothness: A finite sequence $(\mathbf{g}(n))_{n\in[N]}$ on $\mathbb{T}^m$ is (M, N)-smooth if $\|\mathbf{g}(n+1) - \mathbf{g}(n)\| \leq M/N$

for every $n \in [N-1]$. An element $\boldsymbol{\alpha}$ of $\mathbb{T}^m$ is M-rational if $n\boldsymbol{\alpha} = 0$ for some positive integer $n \leq M$. Throughout, by a sub-torus of $\mathbb{T}^m$ we mean a closed connected subgroup of $\mathbb{T}^m$ (perhaps the trivial one). A sub-torus of $\mathbb{T}^m$ is M-rational if its lift in $\mathbb{R}^m$ has a basis of vectors with integer coordinates of absolute value at most M.

6.3. **A property of factorizations on the torus.** The next result will be used multiple times in order to derive inclusions between various sub-tori of $\mathbb{T}^m$.

Lemma 6.2. *Let $L_1, L_2 \in \mathbb{N}$ and S_1, S_2 be two sub-tori of $\mathbb{T}^m$. There exist $\delta_1 := \delta_1(S_1, S_2, L_1, L_2) > 0$ and $N_1 := N_1(S_1, S_2, L_1, L_2)$ such that the following holds: Let $N \geq N_1$ be an integer and $\mathbf{g}\colon [N] \to \mathbb{T}^m$ be an arbitrary sequence that admits the following factorizations:*

$$\mathbf{g}(n) = \boldsymbol{\epsilon}_i(n) + \mathbf{g}'_i(n) + \boldsymbol{\gamma}_i(n), \quad n \in [N], \qquad i = 1, 2, \tag{6.9}$$

where

(i) *$\boldsymbol{\epsilon}_i\colon [N] \to \mathbb{T}^m$ are (L_i, N)-smooth for $i = 1, 2$;*
(ii) *$(\mathbf{g}'_1(n))_{n\in[N]}$ takes values on S_1 and is totally δ_1-equidistributed on S_1;*
(iii) *$(\mathbf{g}'_2(n))_{n\in[N]}$ takes values on S_2;*
(iv) *$\boldsymbol{\gamma}_i\colon [N] \to \mathbb{T}^m$ have period at most L_i for $i = 1, 2$.*

Then $S_1 \subset S_2$.

Remark. It is important for applications that we do not impose an equidistribution assumption on $\mathbf{g}'_2$.

Proof. Replacing L_1 and L_2 by $L := L_1L_2$ we reduce to the case where $L_1 = L_2 = L$ and we can assume that the sequences $\boldsymbol{\gamma}_1$ and $\boldsymbol{\gamma}_2$ have the same period $Q \leq L$.

We argue by contradiction. Suppose that S_1 is not a subset of S_2. Then there exists a (multiplicative) character θ of $\mathbb{T}^m$, with values on the unit circle, such that

$$\theta(\mathbf{x}) = 1 \ \text{ for every } \ \mathbf{x} \in S_2 \ \text{ and } \ \int \theta \, dm_{S_1} = 0. \tag{6.10}$$

Let $A := A(S_1, S_2)$ be the Lipschitz constant of θ.

Since $\theta(\mathbf{x}) = 1$ for $\mathbf{x} \in S_2$ and $\mathbf{g}'_2$ takes values in S_2, identity (6.9) gives that

$$\theta(\boldsymbol{\epsilon}_3(n) + \mathbf{g}'_1(n) + \boldsymbol{\gamma}_3(n)) = 1, \quad \text{ for } \ n \in [N], \tag{6.11}$$

where $\boldsymbol{\epsilon}_3 := \boldsymbol{\epsilon}_1 - \boldsymbol{\epsilon}_2$ is $(2L, N)$-smooth and $\boldsymbol{\gamma}_3 := \boldsymbol{\gamma}_1 - \boldsymbol{\gamma}_2$ has period Q. Let P be the arithmetic progression $\{Q, 2Q, \ldots, Q\lfloor cN \rfloor\}$ where $c := 1/(4AL^2)$. The progression is well defined as long as $N \geq N_1 := 4AL^2$. Then

$$\boldsymbol{\gamma}_3(n) \ \text{ is constant on } \ P, \tag{6.12}$$

and for $n \in P$ we have

$$|1 - \theta\big(\boldsymbol{\epsilon}_3(n) - \boldsymbol{\epsilon}_3(Q)\big)| = |\theta\big(\boldsymbol{\epsilon}_3(n)\big) - \theta\big(\boldsymbol{\epsilon}_3(Q)\big)| \leq A\|\boldsymbol{\epsilon}_3(n) - \boldsymbol{\epsilon}_3(Q)\| \leq 2AL\frac{n-Q}{N} \leq \frac{1}{2}. \tag{6.13}$$

We get

$$\begin{aligned}|\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)\cdot\theta(\mathbf{g}_1'(n))| &= |\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)\cdot\theta(\mathbf{g}_1'(n)-\mathbf{g}_1'(Q))| \\ &= |\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)\cdot\theta(\mathbf{g}_1'(n)+\boldsymbol{\gamma}_3(n) \\ &\quad -\mathbf{g}_1'(Q)-\boldsymbol{\gamma}_3(Q)) \text{ by (6.12)} \\ &= |\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)\cdot\theta(\boldsymbol{\epsilon}_3(Q)-\boldsymbol{\epsilon}_3(n))| \text{ by (6.11)} \\ &\geq \frac{\lfloor cN\rfloor}{N}\big(1-\frac{1}{2}\big) \\ &> \frac{1}{16\,AL^2} \text{ by (6.13) and since } |P|=\lfloor cN\rfloor.\end{aligned}$$

On the other hand, since by assumption $(\mathbf{g}_1'(n))_{n\in[N]}$ is totally δ_1-equidistributed on S_1, θ has Lipschitz constant A, and $\int\theta\,dm_{S_1}=0$ (by (6.10)), we get

$$|\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)\cdot\theta(\mathbf{g}_1'(n))|\leq A\delta_1.$$

Hence, for $\delta_1:=1/(16\,A^2L^2)$ we get a contradiction, completing the proof. ☐

6.4. **Simultaneous factorization of monomials on the torus.** In the course of proving Proposition 6.5 we need to factorize simultaneously several polynomial sequences on the torus and also make sure that the output of the factorization preserves some of the properties of the original sequences. As it does not seem possible to extract such information from Theorem 5.6, using it as a black box, we modify its proof on the torus in order to get the following result that suits our needs.

Theorem 6.3 (Simultaneous factorization of monomials on the torus)**.** *For every $M\in\mathbb{N}$ there exists a finite family $\mathcal{F}_2(M)$ of sub-tori of $\mathbb{T}^m$ that increases with M, such that the following holds: For every function $\omega_2\colon\mathbb{N}\to\mathbb{R}^+$ there exists a positive integer $M_2:=M_2(\omega_2)$ such that for every $N\in\mathbb{N}$ and every $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_d\in\mathbb{T}^m$, there exist $M\in\mathbb{N}$ with $M\leq M_2$, sub-tori $T_1,\dots,T_d$ of $\mathbb{T}^m$, belonging to the family $\mathcal{F}_2(M)$, and for $j=1,\dots,d$, a factorization*

$$n^j\boldsymbol{\alpha}_j=\boldsymbol{\eta}_j(n)+n^j\boldsymbol{\alpha}_j'+\boldsymbol{\theta}_j(n),\quad n\in[N], \tag{6.14}$$

where

(i) $\boldsymbol{\eta}_j\colon[N]\to\mathbb{T}^m$ *is (M,N)-smooth;*
(ii) $\boldsymbol{\alpha}_j'\in T_j$ *and $(n^j\boldsymbol{\alpha}_j')_{n\in[N]}$ is totally $\omega_2(M)$-equidistributed on T_j;*
(iii) $\boldsymbol{\theta}_j\colon[N]\to\mathbb{T}^m$ *is M-rational and has period at most M.*

Proof. For every $M\in\mathbb{N}$, we let

$$\mathcal{F}_2(M):=\{T\subset\mathbb{T}^m\colon T \text{ is an } M\text{-rational torus}\}.$$

The proof is going to be carried out by an iterative procedure that terminates after finitely many steps.

The data. At each step $i=1,2,\dots,$ we have a constant $C_i:=C_i(\omega_2)$ and for $j=1,\dots,d$ a C_i-rational torus $T_{j,i}\subset\mathbb{T}^m$ and factorizations

$$n^j\boldsymbol{\alpha}_j=\boldsymbol{\eta}_{j,i}(n)+n^j\boldsymbol{\alpha}_{j,i}+\boldsymbol{\theta}_{j,i}(n),\quad n\in[N], \tag{6.15}$$

where

(a) $\boldsymbol{\eta}_{j,i}\colon [N] \to \mathbb{T}^m$ is (C_i, N)-smooth;
(b) $\boldsymbol{\alpha}_{j,i} \in T_{j,i}$;
(c) $\boldsymbol{\theta}_{j,i}\colon [N] \to \mathbb{T}^m$ is C_i-rational and has period at most C_i.

Initialization. We initialize our data. Let $C_1 := 1$, and for $j = 1, \dots, d$ let $T_{j,1} := \mathbb{T}^m \in \mathcal{F}_2(C_1)$ and $\boldsymbol{\alpha}_{j,1} := \boldsymbol{\alpha}_j$. We have the trivial factorization $n^j\boldsymbol{\alpha}_j = \boldsymbol{\eta}_{j,1}(n) + n^j\boldsymbol{\alpha}_{j,1} + \boldsymbol{\theta}_{j,1}(n)$, where the sequence $\boldsymbol{\eta}_{j,1}$ is identically zero and thus (C_1, N)-smooth, and the sequence $\boldsymbol{\theta}_{j,1}$ is identically zero and thus C_1-rational and has period at most C_1.

Test of termination. If $(n^j\boldsymbol{\alpha}_{j,i})_{n\in[N]}$ is totally $\omega_2(C_i)$-equidistributed on $T_{j,i}$ for $j = 1, \dots, d$, then we set $M := C_i$, and terminate the process.[12] If not, we proceed to the next step.

Iteration. Our assumption is that there exists $j_0 \in \{1, \dots, d\}$ such that the sequence $(n^{j_0}\boldsymbol{\alpha}_{j_0,i})_{n\in[N]}$ is not totally $\omega_2(C_i)$-equidistributed on the torus $T_{j_0,i}$. Since $T_{j_0,i}$ is C_i-rational, by Theorem 5.2 applied for the torus $T_{j_0,i}$ there exist $A_1 := A_1(C_i, \omega_2) > 0$ and a non-trivial character η of $T_{j_0,i}$, of the form $\mathbf{x} \mapsto \mathbf{k}\cdot\mathbf{x}$ for some $\mathbf{k} \in \mathbb{Z}^m$, with $\|\mathbf{k}\| \le A_1$ and $\|\mathbf{k}\cdot\boldsymbol{\alpha}_{j_0,i}\| \le A_1/N^{j_0}$. We write $T^*_{j_0,i+1}$ for the kernel of η in $T_{j_0,i}$, and $T_{j_0,i+1}$ for the connected component of 0 in $T^*_{j_0,i+1}$. Then the torus $T_{j_0,i+1}$ is A_2-rational, for some $A_2 := A_2(C_i, \omega_2)$. We can write $\boldsymbol{\alpha}_{j_0,i} = \boldsymbol{\beta} + \boldsymbol{\alpha}^*$, where $\boldsymbol{\alpha}^* \in T^*_{j_0,i+1}$ and $\|\boldsymbol{\beta}\| \le A_3/N^{j_0}$ for some $A_3 := A_3(C_i, \omega_2)$. Furthermore, we can write $\boldsymbol{\alpha}* = \boldsymbol{\alpha}_{j_0,i+1} + \boldsymbol{\delta}$, where $\boldsymbol{\alpha}_{j_0,i+1} \in T_{j_0,i+1}$ and $\boldsymbol{\delta}$ is A_4-rational for some $A_4 := A_4(C_i, \omega_2)$. We let

$$C_{i+1} := \max\{C_i + dA_3, A_4C_i\}.$$

Since $\boldsymbol{\alpha}_{j_0,i} = \boldsymbol{\beta} + \boldsymbol{\alpha}_{j_0,i+1} + \boldsymbol{\delta}$, using (6.15) we get

$$n^{j_0}\boldsymbol{\alpha}_{j_0} = \boldsymbol{\eta}_{j_0,i+1}(n) + n^{j_0}\boldsymbol{\alpha}_{j_0,i+1} + \boldsymbol{\theta}_{j_0,i+1}(n), \quad n \in [N],$$

where

$$\boldsymbol{\eta}_{j_0,i+1}(n) := \boldsymbol{\eta}_{j_0,i}(n) + n^{j_0}\boldsymbol{\beta}, \qquad \boldsymbol{\theta}_{j_0,i+1}(n) := \boldsymbol{\theta}_{j_0,i}(n) + n^{j_0}\boldsymbol{\delta}.$$

Note that $(\boldsymbol{\eta}_{j_0,i+1}(n))_{n\in[N]}$ is C_{i+1}-smooth and $(\boldsymbol{\theta}_{j_0,i+1}(n))_{n\in[N]}$ is C_{i+1}-rational and has period at most C_{i+1}.

For $j \ne j_0$ we do not modify the factorization of the step i; that is, we let

$$\boldsymbol{\eta}_{j,i+1} := \boldsymbol{\eta}_{j,i}, \quad \boldsymbol{\alpha}_{j,i+1} := \boldsymbol{\alpha}_{j,i}, \quad \boldsymbol{\theta}_{j,i+1} := \boldsymbol{\theta}_{j,i}, \quad \text{for } j \ne j_0.$$

Since $C_{i+1} \ge C_i$, we have that $(\boldsymbol{\eta}_{j,i+1}(n))_{n\in[N]}$ is C_{i+1}-smooth and $(\boldsymbol{\theta}_{j,i+1}(n))_{n\in[N]}$ is C_{i+1}-rational and has period at most C_{i+1}. We have thus produced for $j = 1, \dots, d$ factorizations similar to (6.15), with $i + 1$ substituted for i that satisfy Properties (a)–(c).

[12]Note that this is necessarily the case when all the tori $T_{j,i}$ are trivial.

The output. At each step, the dimension of exactly one of the tori $T_{j,i}$ decreases by one, and thus the iteration stops after $k \leq dm$ steps at which point the above described test has a positive outcome. For $M := C_k$, we obtain the factorizations (6.14), satisfying the required properties (i), (ii), (iii). Moreover, for $j = 1, \ldots, d$, the torus $T_{j,k}$ is M-rational and thus belongs to the family $\mathcal{F}_2(M)$. Finally, at each step, C_{i+1} is bounded by a quantity that depends only on C_i and on ω_2, and thus $M \leq M_2$ for some $M_2 := M_2(\omega_2)$. This completes the proof. □

6.5. **Quantitative equidistribution of product sequences on the torus.** Given a "sufficiently" totally equidistributed sequence $(\mathbf{g}(n))_{n\in[N]}$ on some torus $\mathbb{T}^m$, we study here equidistribution properties of the product sequence $(\mathbf{g}(pn), \mathbf{g}(qn))_{n\in[N]}$ on $\mathbb{T}^m \times \mathbb{T}^m$ where p, q are distinct positive integers. Our goal is to show that the product sequence is sufficiently equidistibuted on a sub-torus of $\mathbb{T}^m \times \mathbb{T}^m$ that contains an ample supply of "interesting" elements; for instance, we show that this sub-torus contains non-diagonal elements of $\mathbb{T}^m \times \mathbb{T}^m$. We start with a simple but key observation.

Lemma 6.4. *Let $\varepsilon_3 > 0$. There exist $\delta_3 := \delta_3(\varepsilon_3)$ and $N_3 := N_3(\varepsilon_3)$ such that the following holds: Let $N \geq N_3$, and for $j = 1, \ldots, d$, let $\boldsymbol{\alpha}_j \in \mathbb{T}^m$ and suppose that the sequence $(n^j\boldsymbol{\alpha}_j)_{n\in[N]}$ takes values on a sub-torus T_j of $\mathbb{T}^m$ and is totally δ_3-equidistributed on T_j. Then*

(i) *The sequence $(n\boldsymbol{\alpha}_1 + \cdots + n^d\boldsymbol{\alpha}_d)_{n\in[N]}$ is totally ε_3-equidistributed on the torus $T := T_1 + \cdots + T_d$.*
(ii) *For $j = 1, \ldots, d$, the sequence $(n^j(p^j\boldsymbol{\alpha}_j, q^j\boldsymbol{\alpha}_j))_{n\in[N]}$ is totally ε_3-equidistributed on the sub-torus $T_{j,p,q} := \{(p^j\mathbf{x}, q^j\mathbf{x}) : \mathbf{x} \in T_j\}$ of $\mathbb{T}^{2m} = \mathbb{T}^m \times \mathbb{T}^m$.*

Proof. We prove (i). Let $\mathbf{g}(n) := n\boldsymbol{\alpha}_1 + \cdots + n^d\boldsymbol{\alpha}_d$, and suppose that the sequence $(\mathbf{g}(n))_{n\in[N]}$ is not totally ε_3-equidistributed on the torus T. Applying Theorem 5.2 on this torus we get that there exists a constant $C_1 := C_1(\varepsilon_3)$ and $\mathbf{k} \in \mathbb{Z}^m$ with $\|\mathbf{k}\| \leq C_1$ such that $\mathbf{k} \cdot \mathbf{x} \neq 0$ for some $\mathbf{x} \in T$ and

$$\|\mathbf{k} \cdot \boldsymbol{\alpha}_j\| \leq C_1/N^j \quad \text{for} \quad j = 1, \ldots, d. \tag{6.16}$$

Then, for some $j \in \{1, \ldots, d\}$ and some $\mathbf{x} \in T_j$ we have $\mathbf{k} \cdot \mathbf{x} \neq 0$. Using this, using relation (6.16), and applying Theorem 5.3 for the torus T_j, we get that for some $\delta'_j := \delta'_j(\varepsilon_3) > 0$ the sequence $(n^j\boldsymbol{\alpha}_j)_{n\in[N]}$ is not totally δ'_j-equidistributed on T_j. Hence, for $\delta' := \min\{\delta'_1, \ldots, \delta'_d\}$ (in place of δ_3) Property (i) is satisfied.

Now we prove (ii). Suppose that for some $j \in \{1, \ldots, d\}$, the sequence $(n^j(p^j\boldsymbol{\alpha}_j, q^j\boldsymbol{\alpha}_j))_{n\in[N]}$ is not totally ε_3-equidistributed on the torus $T_{j,p,q}$. Applying Theorem 5.2 on this torus we get that there exists a positive real $C_2 := C_2(\varepsilon_3)$ and $(\mathbf{k}_1, \mathbf{k}_2) \in \mathbb{Z}^m \times \mathbb{Z}^m$ with $\|\mathbf{k}_1\| + \|\mathbf{k}_2\| \leq C_2$, such that

$$\mathbf{k}_1 \cdot \mathbf{x}_1 + \mathbf{k}_2 \cdot \mathbf{x}_2 \neq 0 \quad \text{for some} \quad (\mathbf{x}_1, \mathbf{x}_2) \in T_{j,p,q} \tag{6.17}$$

and that $\|(\mathbf{k}_1, \mathbf{k}_2) \cdot (p^j\boldsymbol{\alpha}_j, q^j\boldsymbol{\alpha}_j)\| \leq C_2/N^j$ or, equivalently,

$$\|(p^j\mathbf{k}_1 + q^j\mathbf{k}_2) \cdot \boldsymbol{\alpha}_j\| \leq C_2/N^j. \tag{6.18}$$

Furthermore, writing $(\mathbf{x}_1, \mathbf{x}_2) = (p^j\mathbf{x}, q^j\mathbf{x})$ for some $\mathbf{x} \in T_j$ we get by (6.17) that $(p^j\mathbf{k}_1 + q^j\mathbf{k}_2) \cdot \mathbf{x} \neq 0$, and thus $p^j\mathbf{k}_1 + q^j\mathbf{k}_2 \neq 0$. On the other hand, $\|p^j\mathbf{k}_1 + q^j\mathbf{k}_2\| \leq (p^d + q^d)C_2$, and (6.18) combined with Theorem 5.3 for the torus $\mathbb{T}^m$ gives that for some $\delta''_j := \delta''_j(\varepsilon_3)$ the sequence $(n^j\alpha_j)_{n\in[N]}$ is not totally δ''_j-equidistributed. Hence, for $\delta'' := \min\{\delta''_1,\dots,\delta''_d\}$ (in place of δ_3) Property (ii) is satisfied.

Letting $\delta_3 := \min\{\delta',\delta''\}$ completes the proof. □

Combining Lemma 6.2, Lemma 6.4, and Theorem 6.3, we prove the following factorization result on the torus that is crucial for our purposes.

Proposition 6.5 (Equidistribution of $\mathbf{h}(n)$ in $\mathbb{T}^m$)**.** *For every $M\in\mathbb{N}$ there exists a finite family $\mathcal{F}_4(M)$ of sub-tori of $\mathbb{T}^m$ that increases with M, and for every function $\omega_4\colon\mathbb{N}\to\mathbb{R}^+$ there exist positive integers $M_4 := M_4(\omega_4)$, $N_4 := N_4(\omega_4)$, and a positive real $\delta_4 := \delta_4(\omega_4)$, such that the following holds: Let $N\geq N_4$ be an integer, $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_d\in\mathbb{T}^m$, and $\mathbf{g}(n)=\boldsymbol{\alpha}_1 n+\cdots+\boldsymbol{\alpha}_d n^d$, and suppose that*

$$\text{(6.19)}\qquad \textit{the sequence } (\mathbf{g}(n))_{n\in[N]} \textit{ is totally } \delta_4\textit{-equidistributed on } \mathbb{T}^m.$$

Then there exist $M\in\mathbb{N}$ with $M\leq M_4$ and sub-tori T_j, $j=1,\dots,d$, of $\mathbb{T}^m$, belonging to the family $\mathcal{F}_4(M)$, such that

$$\text{(6.20)}\qquad T_1+\cdots+T_d=\mathbb{T}^m,$$

and the sequence $(\mathbf{h}(n))_{n\in[N]}$ on $\mathbb{T}^{2m}$ defined by $\mathbf{h}(n) := (\mathbf{g}(pn),\mathbf{g}(qn))$ can be factorized as follows:

$$\mathbf{h}(n)=\boldsymbol{\epsilon}(n)+\mathbf{h}'(n)+\boldsymbol{\gamma}(n),\quad n\in[N],$$

where $\boldsymbol{\epsilon}(n)$, $\mathbf{h}'(n)$, $\boldsymbol{\gamma}(n)$ are polynomial sequences on $\mathbb{T}^{2m}$ such that

(i) *$\boldsymbol{\epsilon}\colon[N]\to\mathbb{T}^{2m}$ is (M,N)-smooth;*

(ii) *$(\mathbf{h}'(n))_{n\in[N]}$ takes values and is totally $\omega_4(M)$-equidistributed on the sub-torus*

$$\text{(6.21)}\quad R_{T_1,\dots,T_d} := \{(p\mathbf{x}_1+\cdots+p^d\mathbf{x}_d\,,\, q\mathbf{x}_1+\cdots+q^d\mathbf{x}_d)\colon \mathbf{x}_j\in T_j \textit{ for } j=1,\dots,d\}$$

of $\mathbb{T}^{2m}$;

(iii) *$\boldsymbol{\gamma}\colon[N]\to\mathbb{T}^{2m}$ is M-rational and has period at most M.*

Proof. Throughout this proof when we write "for every sufficiently large N," we mean for every $N\in\mathbb{N}$ that is larger than a constant that depends on ω_4.

Let $\omega'_4\colon\mathbb{N}\to\mathbb{R}^+$ be a function that will be specified later and depends only on ω_4 (its defining properties are given by (6.24) and (6.25) below).

For every $M\in\mathbb{N}$, Theorem 6.3 applied on $\mathbb{T}^m$ provides a finite family $\mathcal{F}_2(M)$ of sub-tori of $\mathbb{T}^m$ that increases with M, and we define as $\mathcal{F}_4(M)$ the family spanned by $\mathcal{F}_2(M)$, the torus $\mathbb{T}^m$, and is invariant under addition of tori. Applying Theorem 6.3 on $\mathbb{T}^m$ with the function ω'_4, we get a positive integer $M_4 := M_4(\omega'_4)$, such that the following holds: For every $\boldsymbol{\alpha}_1,\dots,\boldsymbol{\alpha}_d\in\mathbb{T}^m$ there exists $M\in\mathbb{N}$ with $M\leq M_4$ and sub-tori $T_1,\dots,T_d$ of $\mathbb{T}^m$ belonging to the family $\mathcal{F}_4(M)$, and for $j=1,\dots,d$, factorizations

$$\text{(6.22)}\qquad n^j\boldsymbol{\alpha}_j=\boldsymbol{\eta}_j(n)+n^j\boldsymbol{\alpha}'_j+\boldsymbol{\theta}_j(n),\quad n\in[N],$$

where

(a) $\boldsymbol{\eta}_j\colon[N]\to\mathbb{T}^m$ is (M,N)-smooth;

(b) $\boldsymbol{\alpha}'_j\in T_j$ and the sequence $(n^j\boldsymbol{\alpha}'_j)_{n\in[N]}$ is totally $\omega'_4(M)$-equidistributed on T_j;

(c) $\boldsymbol{\theta}_j\colon[N]\to\mathbb{T}^m$ is M-rational and has period at most M.

We are going to show that for an appropriate choice of ω'_4 and δ_4 we can use these data to get a factorization for $\mathbf{h}$ that satisfies Properties (i)–(iii) and (6.20). To this end, write

$$\mathbf{h}(n) = \boldsymbol{\epsilon}(n) + \mathbf{h}'(n) + \boldsymbol{\gamma}(n), \quad n \in [N],$$

where

$$\boldsymbol{\epsilon}(n) := (\boldsymbol{\eta}_1(pn), \boldsymbol{\eta}_1(qn)) + (\boldsymbol{\eta}_2(pn), \boldsymbol{\eta}_2(qn)) + \cdots + (\boldsymbol{\eta}_d(pn), \boldsymbol{\eta}_d(qn));$$
$$\mathbf{h}'(n) := n(p\boldsymbol{\alpha}'_1, q\boldsymbol{\alpha}'_1) + n^2(p^2\boldsymbol{\alpha}'_2, q^2\boldsymbol{\alpha}'_2) + \cdots + n^d(p^d\boldsymbol{\alpha}'_d, q^d\boldsymbol{\alpha}'_d);$$
$$\boldsymbol{\gamma}(n) := (\boldsymbol{\theta}_1(pn), \boldsymbol{\theta}_1(qn)) + (\boldsymbol{\theta}_2(pn), \boldsymbol{\theta}_2(qn)) + \cdots + (\boldsymbol{\theta}_d(pn), \boldsymbol{\theta}_d(qn)).$$

Note that $\boldsymbol{\gamma}$ is M^d-rational and has period at most M^d. Replacing M_4 with CM_4^d and M with CM^d for some constant C that depends only on d, m, p, q, we have that the sequence $\boldsymbol{\epsilon}$ is (M, N)-smooth, and Properties (i) and (iii) of the proposition are satisfied.

We move now to Property (ii). Note first that $\mathbf{h}'(n)$ takes values on the torus $R_{T_1,\ldots,T_d}$. By (b), for $j = 1, \ldots, d$ the sequence $(n^j \boldsymbol{\alpha}'_j)_{n\in[N]}$ is totally $\omega'_4(M)$-equidistributed on T_j. Using this property and Part (ii) of Lemma 6.4, we get that for every sufficiently large N, for $j = 1, \ldots, d$, the sequence

$$(n^j(p^j\boldsymbol{\alpha}'_j, q^j\boldsymbol{\alpha}'_j))_{n\in[N]} \tag{6.23}$$

is $\rho'(\omega'_4(M))$-equidistributed on the torus

$$\{(p^j\mathbf{x}, q^j\mathbf{x}) : \mathbf{x} \in T_j\},$$

where $\rho' \colon \mathbb{R}^+ \to \mathbb{R}^+$ is a function that decreases to 0 as $t \to 0^+$ and depends only on d, m, p, q. We are now in position to apply Part (i) of Lemma 6.4 on the torus $\mathbb{T}^{2m}$ for the sequences in (6.23). It gives that the sequence $(\mathbf{h}'(n))_{n\in[N]}$ is $\omega_4(M)$-equidistributed on $R_{T_1,\ldots,T_d}$ as long as the function $\omega'_4 \colon \mathbb{N} \to \mathbb{R}^+$ satisfies

$$\rho'(\omega'_4(M)) \le \delta_3(\omega_4(M)), \quad \text{for every} \;\; M \in \mathbb{N}, \tag{6.24}$$

where δ_3 was defined in Lemma 6.4. As $\delta_3 > 0$ and $\rho'(t) \to 0$ as $t \to 0^+$, such an ω'_4 exists.

It remains to establish (6.20), that is, that $T_1 + \cdots + T_d = \mathbb{T}^m$. To get this, we need to impose two additional conditions, one on ω'_4 and one on the degree of total equidistribution δ_4 of $(g(n))_{n\in[N]}$ that was left unspecified until this point. We choose $\omega'_4 \colon \mathbb{N} \to \mathbb{R}^+$ so that in addition to (6.24) it satisfies

$$\omega'_4(M) \le \min_{S_1,S_2\in\mathcal{F}_4(M)} \{\delta_1(S_1, S_2, 1, dM^d)\}, \quad \text{for every} \;\; M \in \mathbb{N}, \tag{6.25}$$

where $\delta_1(S_1, S_2, 1, M^2)$ was defined on Lemma 6.2. Now M_4 is well defined, and we let

$$\delta_4 := \min_{1\le M\le M_4} \{\omega'_4(M)\}. \tag{6.26}$$

We assume that Property (6.19) holds for this value of δ_4 and Properties (a), (b), (c) hold for some $M \le M_4$.

Next, note that we have two factorizations for the sequence $(\mathbf{g}(n))_{n\in[N]}$. The first is the trivial one: $\mathbf{g}(n) = 0 + \mathbf{g}(n) + 0$ where $\mathbf{g}(n)$ takes values in $\mathbb{T}^m$ and is totally $\omega'_4(M)$-equidistributed on $\mathbb{T}^m$ (by (6.26)). The second is given by (6.22),

$$\mathbf{g}(n) = \boldsymbol{\eta}(n) + \mathbf{g}'(n) + \boldsymbol{\theta}(n), \quad n \in [N],$$

where

$$\boldsymbol{\eta}(n) := \boldsymbol{\eta}_1(n) + \cdots + \boldsymbol{\eta}_d(n) \text{ is } (dM, N)\text{-smooth} \quad \text{(by (a))};$$

$$\mathbf{g}'(n) := n\boldsymbol{\alpha}'_1 + \cdots + n^d\boldsymbol{\alpha}'^d_d \text{ takes values in } T_1 + \cdots + T_d \quad \text{(by (b))};$$

$$\boldsymbol{\theta}(n) := \boldsymbol{\theta}_1(n) + \cdots + \boldsymbol{\theta}_d(n) \text{ is } M^d\text{-rational and has period at most } M^d \quad \text{(by (c))}.$$

Since by assumption $\mathcal{F}_4(M)$ contains $\mathbb{T}^m$, $T_1, \ldots, T_d$, and is closed under addition of tori, we have $\mathbb{T}^m$, $T_1 + \cdots + T_d \in \mathcal{F}_4(M)$. Furthermore, by (6.25) we have $\omega'_4(M) \le \delta_1(\mathbb{T}^m, T_1 + \cdots + T_d, 1, dM^d)$. Hence, Lemma 6.2 is applicable and gives that for every sufficiently large N we have $\mathbb{T}^m \subset T_1 + \cdots + T_d$. It follows that $\mathbb{T}^m = T_1 + \cdots + T_d$, and the proof is complete. □

6.6. **A key algebraic fact.** Our goal is to establish the following key property.

Proposition 6.6. *Let G be an s-step nilpotent group and H be a subgroup of $G \times G$. Suppose that there exist normal subgroups $G^1, \ldots, G^d$ of G such that*

(i) $G = G^1 \cdots G^d$;

(ii) $\{(g_1^p \cdots g_d^{p^d}, g_1^q \cdots g_d^{q^d}) \colon g_1 \in G^1, \ldots, g_d \in G^d\} \subset H \cdot (G_2 \times G_2)$.

Then the set $U := \{g \in G_s \colon (g^{p^j}, g^{q^j}) \in H \text{ for some } j \in \mathbb{N}\}$ generates G_s.

Proof. For $j = 1, \ldots, d$, let

$$H^j := H \cap \{(g^{p^j} u, g^{q^j} u') \colon g \in G^j, \ u, u' \in G_2\}.$$

Then H^j is a normal subgroup of H. For $k \in \mathbb{N}$ and $\vec{j} = (j_1, \ldots, j_k) \in \{1, \ldots, d\}^k$, we write

$$G_{\vec{j}} := [\cdots[[G^{j_1}, G^{j_2}], G^{j_3}]\cdots] \quad \text{and} \quad H_{\vec{j}} := [\cdots[[H^{j_1}, H^{j_2}], H^{j_3}]\cdots]$$

for the iterated commutator groups.

Since the subgroups G^j of G are normal, every group $G_{\vec{j}}$ is normal. Moreover, since $G = G^1 \cdots G^d$, for every $k \in \{1, \ldots, d\}$ the group G_k is the product of the groups $G_{\vec{j}}$ for $\vec{j} \in \{1, \ldots, d\}^k$.

Claim 1. *Let $\vec{j} = (j_1, \ldots, j_k)$ with coordinates in $\{1, \ldots, d\}$ and $j := j_1 + \cdots + j_k$. For every $g \in G_{\vec{j}}$ there exist $u, u' \in G_{k+1}$ such that $(g^{p^j} u, g^{q^j} u') \in H_{\vec{j}}$.*

We prove the claim by induction on k. If $k = 1$, then $\vec{j} = (j_1)$ so $G_{\vec{j}} = G^{j_1}$, $H_{\vec{j}} = H^{j_1}$, and the announced property follows immediately from the definition of the group H^{j_1} and the hypothesis. Let $k > 1$, and suppose that the result holds for $k - 1$; we are going to show that it holds for k. We let

$$A := \{g \in G_{\vec{j}} \colon \exists u, u' \in G_{k+1}, \ (g^{p^j} u, g^{q^j} u') \in H_{\vec{j}}\},$$

and we have to prove that $A = G_{\vec{j}}$. We claim first that A is a subgroup of $G_{\vec{j}}$. Indeed, let $g, h \in A$, then $g, h \in G_{\vec{j}}$, and there exist $u, u', v, v' \in G_{k+1}$ such that $(g^{p^j} u, g^{q^j} u')$ and $(h^{p^j} v, h^{q^j} v')$ belong to $H_{\vec{j}}$. Then $(g^{p^j} u h^{p^j} v, g^{q^j} u' h^{q^j} v') \in H_{\vec{j}}$ and furthermore

$$g^{p^j} u h^{p^j} v = (gh)^{p^j} \bmod G_{k+1}, \quad g^{q^j} u' h^{q^j} v' = (gh)^{q^j} \bmod G_{k+1}.$$

Hence, $gh \in A$. Furthermore, $(u^{-1} g^{-p^j}, u'^{-1} g^{-q^j}) = (g^{-p^j} u_1, g^{-q^j} u_2) \in H_{\vec{j}}$ for some $u_1, u_2 \in G_{k+1}$. Hence, $g^{-1} \in A$. It follows that A is a group.

We let $\vec{i} := (j_1, \dots, j_{k-1})$ and $i := j_1 + \cdots + j_{k-1}$. Then $G_{\vec{j}}$ is the group spanned by elements $[h, z]$ with $h \in G_{\vec{i}}$ and $z \in G^{j_k}$, and as A is a group, it suffices to prove that each element of this form belongs to A. By the induction hypothesis, there exist $u, u' \in G_k$ with $(h^{p^i}u, h^{q^i}u') \in H_{\vec{i}}$ and by the first step there exist $v, v' \in G_2$ with $(z^{p^{j_k}}v, z^{q^{j_k}}v') \in H^{j_k}$. The commutator $\big([h^{p^i}u, z^{p^{j_k}}v]\,,\,[h^{q^i}u', z^{q^{j_k}}v']\big)$ of these two elements belongs to $H_{\vec{j}}$. Furthermore,

$$[h^{p^i}u, z^{p^{j_k}}v] = [h,z]^{p^j} \bmod G_{k+1}, \quad [h^{q^i}u', z^{q^{j_k}}v'] = [h,z]^{q^j} \bmod G_{k+1}.$$

Hence, $[h, z] \in H_{\vec{j}}$. This completes the proof of Claim 1.

Taking $k = s$ and using that G_{s+1} is trivial, we get the following claim.

Claim 2. *Let $\vec{j} := (j_1, \dots, j_s)$ with coordinates in $\{1, \dots, d\}$ and $j := j_1 + \cdots + j_s$. Then for every $g \in G_{\vec{j}}$ we have $(g^{p^j}, g^{q^j}) \in H_s$.*

We are now ready to show that the set U generates G_s. As already noticed, G_s is the product of the groups $G_{\vec{j}}$ for $\vec{j} \in \{1, \dots, d\}^s$. Hence, it suffices to show that the set U contains all these groups. So let $\vec{j} := (j_1, \dots, j_s)$ with $j_i \in \{1, \dots, d\}$, $i = 1, \dots, s$, and suppose that $g \in G_{\vec{j}}$. By Claim 2 we have $(g^{p^j}, g^{q^j}) \in H_s \subset H$ for $j = j_1 + \cdots + j_s$, which proves that $g \in U$. This completes the proof of Proposition 6.6. □

7. Minor arc nilsequences-proof of the discorrelation estimate

Our goal in this section is to prove Theorem 6.1. Suppose that the group G is s-step nilpotent. The proof goes by induction on s. We assume either that $s = 1$ or that $s \geq 2$, and the result holds for $(s-1)$-step nilmanifolds. We are going to show that it holds for s-step nilmanifolds.

7.1. Reduction to the case of a nilcharacter.

We start with some reductions, similar to those made in the proof of [30, Lemma 3.7]. We let $r := \dim(G_s)$ and $t := \dim(X)$. Suppose that (6.1) holds.

There exists a constant $A_1 := A_1(X, \tau)$ and a function Φ' on X with

$$\|\Phi - \Phi'\|_\infty \leq \frac{\tau}{2}, \quad \int_X \Phi'\, dm_X = 0, \quad \text{and} \quad \|\Phi'\|_{\mathcal{C}^{2t}(X)} \leq A_1.$$

Then $|\mathbb{E}_{n\in[N]} \mathbf{1}_P(n) f(n) \Phi'(g(n+k)\cdot e_X)| > \tau/2$. Therefore, substituting Φ' for Φ, up to a change in the constants, we can assume that

$$\|\Phi\|_{\mathcal{C}^{2t}(X)} \leq 1.$$

We proceed now to a "vertical Fourier decomposition." Using the Mal'cev basis of G, we identify the vertical torus $G_s/(G_s \cap \Gamma)$ with $\mathbb{T}^r$ and its dual group with $\mathbb{Z}^r$. For $\mathbf{h} \in \mathbb{Z}^r$ let

$$\Phi_{\mathbf{h}}(x) := \int_{\mathbb{T}^r} \mathrm{e}(-\mathbf{h}\cdot\mathbf{u})\, \Phi(\mathbf{u}\cdot x)\, dm_{\mathbb{T}^r}(\mathbf{u}).$$

We have

$$\|\Phi_{\mathbf{h}}\|_{\mathrm{Lip}(X)} \leq 1, \qquad \int_X \Phi_{\mathbf{h}}\, dm_X = 0,$$

and $\Phi_{\mathbf{h}}$ is a nilcharacter of frequency $\mathbf{h}$, that is,

$$\Phi_{\mathbf{h}}(\mathbf{v}\cdot x) = \mathrm{e}(\mathbf{h}\cdot\mathbf{v})\, \Phi_{\mathbf{h}}(x) \quad \text{for every } \mathbf{v} \in \mathbb{T}^r \text{ and every } x \in X.$$

Moreover, since $\|\Phi\|_{\mathcal{C}^{2t}(X)} \leq 1$, we have

$$\|\Phi_{\mathbf{h}}\|_\infty \leq A_2(1+\|\mathbf{h}\|)^{-2t}$$

for some constant $A_2 := A_2(X)$ and

$$\Phi(x) = \sum_{\mathbf{h}\in\mathbb{Z}^r} \Phi_{\mathbf{h}}(x) \quad \text{for every} \quad x \in X.$$

It follows from (6.1) that there exists $\mathbf{h} \in \mathbb{Z}^r$, with $\|\mathbf{h}\| \leq A_3$ and

$$|\mathbb{E}_{n\in[N]} \mathbf{1}_P(n)\, f(n)\, \Phi_{\mathbf{h}}(g(n+k)\cdot e_X)| > \tau_1 \tag{7.1}$$

for some positive reals $\tau_1 := \tau_1(X,\tau)$ and $A_3 := A_3(X,\tau)$. Therefore, we can assume that (6.1) holds with τ_1 in place of τ for some nilcharacter Φ of frequency $\mathbf{h}$ and $\|\mathbf{h}\| \leq A_3$.

7.2. **Reduction to the case of a non-zero frequency.** First, suppose that $s = 1$ and $\mathbf{h} = 0$. We have $G = G_s$, and $\Phi_{\mathbf{0}}$ is constant. Since the integral of Φ is equal to zero, Φ is identically zero, and we have a contradiction by (6.1).

Suppose now that $s \geq 2$ and that $\mathbf{h} = 0$. As in Section 4.4, we let $\widetilde{G} := G/G_s$ and $\widetilde{\Gamma} := \Gamma/(\Gamma\cap G_s)$. Then the $(s-1)$-step nilmanifold $\widetilde{X} := \widetilde{G}/\widetilde{\Gamma}$ is identified with the quotient of X under the action of the vertical torus $\mathbb{T}^r$. Let $\pi\colon X \to \widetilde{X}$ be the natural projection. Since Φ is a nilcharacter with frequency 0, it can be written as $\Phi := \widetilde{\Phi}\circ\pi$ for some function $\widetilde{\Phi}$ on $\widetilde{X}$, and we have $\int_{\widetilde{X}} \widetilde{\Phi}\, dm_{\widetilde{X}} = 0$ and $\|\widetilde{\Phi}\|_{\mathrm{Lip}(\widetilde{X})} \leq A_4$ for some constant $A_4 := A_4(X)$. Let $\widetilde{g}$ be the image of the polynomial sequence g in $\widetilde{G}$ under the natural projection. Then $\widetilde{g} \in \mathrm{poly}(\widetilde{G}_\bullet)$ where $\widetilde{G}^{(j)} := (G^{(j)}G_s)/G_s$ for every j. We have

$$|\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)f(n)\widetilde{\Phi}(\widetilde{g}(n+k)\cdot e_{\widetilde{X}})| = |\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)f(n)\Phi(g(n+k)\cdot e_X)| > \tau_1.$$

Assuming that N is sufficiently large, depending on X and on τ, by the induction hypothesis we get that the sequence $(\widetilde{g}(n)\cdot e_{\widetilde{X}})_{n\in[N]}$ is not totally σ'-equidistributed for some $\sigma' := \sigma'(X,\tau)$. This implies a similar property for the sequence $(g(n)\cdot e_X)_{n\in[N]}$ and completes the induction in the case where $\mathbf{h} = 0$.

We can therefore assume that the frequency $\mathbf{h}$ of the nilcharacter Φ is non-zero.

7.3. **Reduction to the case where $k = 0$ and $g(0) = 1_G$.** Suppose that the conclusion (6.2) holds for some N_0 and σ, under the stronger assumption that the hypothesis (6.1) holds for $k = 0$ and for sequences that satisfy $g(0) = 1_G$. We are going to show that it holds without these assumptions. Let $\tau > 0$ and $N \geq N_0$. Let $F \subset G$ be a bounded fundamental domain of the projection $G \to X$ (we assume that F is fixed given X). By the first statement of Lemma 4.1 there exists a constant $C_1 > 0$ such that

$$C_1^{-1} d_X(x,x') \leq d_X(g\cdot x, g\cdot x') \leq C_1 d_X(x,x') \ \text{ for every } g\in F \text{ and } x,x'\in X. \tag{7.2}$$

Let the sequence $g \in \mathrm{poly}(G_\bullet)$ be given as above and $k \in \mathbb{N}$. We write

$$g(k) = a_k\gamma_k \quad \text{where} \quad a_k \in F \quad \text{and} \quad \gamma_k \in \Gamma.$$

Let $\widetilde{g}\colon [N] \to G$ be defined by

$$\widetilde{g}(n) := a_k^{-1} g(n+k)\gamma_k^{-1}.$$

Then $\widetilde{g}(0) = 1_G$. By the first definition in Section 5.1 and since the subgroups $G^{(i)}$ of G are normal, we have $\widetilde{g} \in \mathrm{poly}(G_\bullet)$ and for every $n \in \mathbb{N}$ we have

$$g(n+k) \cdot e_X = a_k\, \widetilde{g}(n) \cdot e_X.$$

We let

$$\Phi_k(x) := \Phi(a_k \cdot x).$$

Since Φ is a nilcharacter with non-zero frequency $\mathbf{h} \in \mathbb{Z}^r$, for every $k \in \mathbb{N}$, Φ_k is also a nilcharacter with the same frequency. Since a_k belongs to F for every $k \in \mathbb{N}$ and $\|\Phi\|_{\mathrm{Lip}(X)} \leq 1$, we get by (7.2) that $\|\Phi_k\|_{\mathrm{Lip}(X)} \leq C_1$. We let $\widetilde{\Phi}_k := \Phi_k / C_1$. Then $\|\widetilde{\Phi}_k\|_{\mathrm{Lip}(X)} \leq 1$, $\int \widetilde{\Phi}_k \, dm_X = 0$, and estimate (7.1) implies that

$$|\mathbb{E}_{n \in [N]} \mathbf{1}_P(n)\, f(n)\, \widetilde{\Phi}_k(\widetilde{g}(n) \cdot e_X)| \geq \tau_2$$

for some $\tau_2 := \tau_2(X, \tau) > 0$. We are now in a situation where the additional hypotheses are satisfied, that is, $k = 0$ and $\widetilde{g}(0) = 1_G$. We deduce that the sequence $(\widetilde{g}(n) \cdot e_X)_{n \in [N]}$ is not totally σ_1-equidistributed in X for some $\sigma_1 > 0$. Let η be the horizontal character provided by Theorem 5.2. We have that $\eta(\widetilde{g}(n)) = \eta(g(k))^{-1} \eta(g(n+k))$. Applying Lemma 5.1 with $\phi(n) := \eta(g(n+k))$ and $\psi(n) := \eta(g(n))$ and then applying Lemma 5.3, we deduce that there exist a positive integer N_0' and a positive real σ_2, such that if $N \geq N_0'$, then the sequence $(g(n) \cdot e_X)_{n \in [N]}$ is not totally σ_2-equidistributed in X. Hence, in establishing Theorem 6.1, we can assume that $k = 0$ and $g(0) = 1_G$.

Therefore, in the rest of this proof, we can and will assume that $g(0) = 1_G$, and that

$$|\mathbb{E}_{n \in [N]} \mathbf{1}_P(n)\, f(n)\, \Phi(g(n) \cdot e_X)| \geq \tau_2 \tag{7.3}$$

for some $\tau_2 := \tau_2(X, \tau) > 0$, where

$$\Phi \ \text{ is a nil-character with non-zero frequency.}$$

7.4. **Using the orthogonality criterion of Kátai.** Combining the lower bound (7.3) and Lemma 3.1 we get that there exists a positive integer $K := K(X, \tau_2) = K(X, \tau)$, primes p, q with $p < q < K$, and a positive real $\tau_3 := \tau_3(X, \tau_2) = \tau_3(X, \tau)$, such that

$$|\mathbb{E}_{n \in [N]} \mathbf{1}_{[N/q]}(n) \mathbf{1}_P(pn)\, \mathbf{1}_P(qn)\, \Phi(g(pn) \cdot e_X) \cdot \overline{\Phi}(g(qn) \cdot e_X)| \geq \tau_3.$$

Let $P_1 \subset [N]$ be an arithmetic progression such that $\mathbf{1}_{[N/q]}(n) \mathbf{1}_P(pn) \mathbf{1}_P(qn) = \mathbf{1}_{P_1}(n)$. Then the last inequality can be rewritten as

$$|\mathbb{E}_{n \in [N]} \mathbf{1}_{P_1}(n)\, \Phi(g(pn) \cdot e_X) \cdot \overline{\Phi}(g(qn) \cdot e_X)| \geq \tau_3. \tag{7.4}$$

We remark that the pairs (p, q) with $p, q < K$ belong to some finite set that depends only on X and τ. Therefore, from this point on we can and will assume that p and q are fixed distinct primes. Almost all parameters defined below will depend on p and q, and in order to ease our notation a bit, this dependence is going to be left implicit.

7.5. **Using the factorization theorem on $X \times X$.** Recall that $g \in \text{poly}(G_\bullet)$. Let $G \times G$ be endowed with the product filtration and the product Mal'cev basis. We define the sequence $h\colon [N] \to G \times G$ by

$$h(n) := (g(pn), g(qn)), \quad n \in [N]. \tag{7.5}$$

For every $i \in \mathbb{N}$ and $k_1, \ldots, k_i \in \mathbb{Z}$ we have (recall that $\partial_h g(n) := g(n+h)g(n)^{-1}$)

$$\partial_{k_i} \cdots \partial_{k_1} h(n) = ((\partial_{pk_i} \cdots \partial_{pk_1} g)(pn), (\partial_{qk_i} \cdots \partial_{pk_1} g)(qn)) \in G^{(j)} \times G^{(j)} = (G \times G)^{(j)},$$

and thus $h \in \text{poly}((G \times G)_\bullet)$. We can rewrite (7.4) as

$$|\mathbb{E}_{n \in [N]} \mathbf{1}_{P_1}(n)\, (\Phi \otimes \overline{\Phi})(h(n) \cdot e_{X \times X})| \geq \tau_3. \tag{7.6}$$

For the sequence $h \in \text{poly}((G \times G)_\bullet)$, given by (7.5), we apply Theorem 5.6 for a function $\omega_5\colon \mathbb{N} \to \mathbb{R}^+$ that will be determined shortly (its defining relation is (7.10)) and depends only on X, τ. We get families $\mathcal{F}_5(M)$, $M \in \mathbb{N}$, of sub-nilmanifolds of $X \times X$ (which do not depend on ω_5), that increase with M, a constant $M_5 := M_5(X, \omega_5)$, an integer M^* with $1 \leq M^* \leq M_5$, a closed and connected rational subgroup H of $G \times G$, a nilmanifold $Y := H/(H \cap (\Gamma \times \Gamma))$ belonging to the family $\mathcal{F}_5(M^*)$, and a factorization

$$h(n) = \epsilon(n)\, h'(n)\, \gamma(n), \quad n \in [N],$$

where

(i) $\epsilon\colon [N] \to G \times G$ is (M^*, N)-smooth;
(ii) $h' \in \text{poly}(H_\bullet)$ and $(h'(n) \cdot e_{X \times X})_{n \in [N]}$ is totally $\omega_5(M^*)$-equidistributed in Y with the metric d_Y induced by the filtration $H_\bullet$;
(iii) $\gamma\colon [N] \to G \times G$ is M^*-rational and $(\gamma(n) \cdot e_{X \times X})_{n \in [N]}$ has period at most M^*.

We start from (7.6) and use the argument of Section 5.5 with $X \times X$ and $G \times G$ substituted for X and G, respectively. We get a rational element α belonging to the finite subset $\Sigma(M^*)$ of $G \times G$, an integer $n_0 \in [N]$, and an arithmetic progression $P_2 \subset P_1$, such that the function $(\Phi \otimes \overline{\Phi})_\alpha$ and the sequence h'_α, defined by

$$(\Phi \otimes \overline{\Phi})_\alpha(y) := (\Phi \otimes \overline{\Phi})(\epsilon(n_0)\alpha \cdot y) \ \text{ for } y \in Y; \tag{7.7}$$
$$h'_\alpha(n) := \alpha^{-1} h'(n) \alpha \ \text{ for } n \in [N],$$

satisfy

$$|\mathbb{E}_{n \in [N]} \mathbf{1}_{P_2}(n)\, (\Phi \otimes \overline{\Phi})_\alpha(h'_\alpha(n) \cdot e_{X \times X})| \geq \tau_4(M^*), \tag{7.8}$$

where

$$\tau_4(M^*) := \tau_3^2 / (64\, H_1(M^*)^2 {M^*}^2)$$

and $H_1(M^*)$ is the constant defined as $H(M^*)$ in Section 5.5. We proceed as in Section 5.5 with H in place of G' and Y in place of X', and we let

$$H_\alpha := \alpha^{-1} H \alpha \ \text{ and } \ Y_\alpha := H_\alpha \cdot e_Y.$$

We have that $h'_\alpha \in \text{poly}(H_{\alpha\bullet})$, and by Property (v) of Section 5.5.1, if N is large enough, depending on $X, M^*, \omega_5(M^*)$, then for every $\alpha \in \Sigma(M^*)$ and $Y \in \mathcal{F}_5(M^*)$ we have

$$(h'_\alpha(n) \cdot e_Y)_{n \in [N]} \text{ is totally } \rho_{X \times X}(M^*, \omega_5(M^*))\text{-equidistributed in } Y_\alpha. \tag{7.9}$$

Note that since $\|\Phi\|_{\text{Lip}(X)} \leq 1$, we have $\|\Phi \otimes \overline{\Phi}\|_{\text{Lip}(X \times X)} \leq C_2$ for some positive real $C_2 := C_2(X)$.

We can now define the function ω_5. Recall that for every fixed $M \in \mathbb{N}$ we have $\rho_{X\times X}(M,t) \to 0$ as $t \to 0^+$. Therefore, there exists $\omega_5 \colon \mathbb{N} \to \mathbb{R}^+$ such that

$$\rho_{X\times X}(M, \omega_5(M)) < \tau_4(M) H_1(M)^{-2} C_2^{-1} \quad \text{for every } M \in \mathbb{N}. \tag{7.10}$$

Note that there is no circularity in defining these parameters, as $\mathcal{F}_5(M)$, τ_4, H_1, C_2 do not depend on ω_5. Furthermore, the function ω_5 and the integer M_5 depend only on X and on τ; hence there exists $N_5 := N_5(X, \tau)$ such that

$$\text{if } N \geq N_5, \text{ then (7.9) holds for } M^* \in \mathbb{N} \text{ with } M^* \leq M_5 \text{ defined as above.} \tag{7.11}$$

7.6. **Reduction to a zero mean property.** We work with the value of M^* found in the previous subsection and assume that $N \geq N_5$. Suppose for the moment that

$$\int_{Y_\alpha} (\Phi \otimes \overline{\Phi})_\alpha \, dm_{Y_\alpha} = 0. \tag{7.12}$$

Since $M^* \leq M_5$, by (7.11), $(h'_\alpha(n) \cdot e_{X\times X})_{n\in[N]}$ is totally $\rho_{X\times X}(M^*, \omega_5(M^*))$-equidistributed in Y_α. Furthermore, it follows from (5.12) that $\|(\Phi \otimes \overline{\Phi})_\alpha\|_{\mathrm{Lip}(X\times X)} \leq C_2 H_1(M^*)$, and using Property (iv) of Section 5.5.1 we get $\|(\Phi \otimes \overline{\Phi})_\alpha|_{Y_\alpha}\|_{\mathrm{Lip}(Y_\alpha)} \leq C_2 H_1(M^*)^2$. It follows that

$$|\mathbb{E}_{n\in[N]} \mathbf{1}_{P_2}(n)\, (\Phi \otimes \overline{\Phi})_\alpha (h'_\alpha(n) \cdot e_{X\times X})| \leq \rho_{X\times X}(M^*, \omega_5(M^*)) \cdot C_2 H_1(M^*)^2 < \tau_4(M^*)$$

by (7.10), contradicting (7.8). Hence, (7.12) cannot hold.

Therefore, in order to complete the proof of Theorem 6.1 it remains to show that there exist a positive real σ and a positive integer N_6, both depending on X and τ only, such that if for some $N \geq N_6$ the sequence $(g(n) \cdot e_X)_{n\in[N]}$ is totally σ-equidistributed in X, then (7.12) holds, where Φ, α, Y_α are as before. The values of σ and N_6 are going to be determined in Section 7.8.1.

7.7. **Reduction to an algebraic property.** Suppose for the moment that the group H defined in Section 7.5 satisfies the following property:

$$\text{the set } U := \{u \in G_s \colon (u^{p^j}, u^{q^j}) \in H \text{ for some } j \in \mathbb{N}\} \text{ generates } G_s, \tag{7.13}$$

where we use multiplicative notation for G_s. We claim that then (7.12) holds. To see this, notice first that since Φ is a nilcharacter of X with non-zero frequency, there exists a non-trivial multiplicative character $\theta \colon G_s \to \mathbb{T}$ such that

$$\Phi(u \cdot y) = \theta(u) \cdot \Phi(y), \quad \text{for every } y \in Y, \ u \in G_s.$$

Since θ is non-trivial, there exists $u \in G_s$ such that $\theta(u)$ is irrational. By (7.13), there exists u with $\theta(u)$ irrational and such that $(u^{p^j}, u^{q^j}) \in H$ for some $j \in \mathbb{N}$. Then $\theta(u^{p^j - q^j}) \neq 1$ and

$$(\Phi \otimes \overline{\Phi})_\alpha((u^{p^j}, u^{q^j}) \cdot y) = \theta(u^{p^j - q^j}) \cdot (\Phi \otimes \overline{\Phi})_\alpha(y), \tag{7.14}$$

where $(\Phi \otimes \overline{\Phi})_\alpha$ is defined in (7.7) and we used that (u^{p^j}, u^{q^j}) belongs in the center of $G \times G$ and hence commutes with the element α. Left multiplication by (u^{p^j}, u^{q^j}) is a measure preserving transformation on Y_α and thus, after integrating the last relation on this set, we obtain (7.12).

Thus, at this point we have reduced matters to showing that there exist a positive real σ and a positive integer N_6, both depending on X and τ only, such that if for some $N \geq N_6$ the sequence $(g(n) \cdot e_X)_{n\in[N]}$ is totally σ-equidistributed in X, then Property (7.13) holds. We show this in the final part of this section using the tools developed in Section 6.

7.8. **Our plan and definition of parameters.** There are two key ingredients involved in the proof of Property (7.13). The first is Proposition 6.5 that gives information about the action of the sequence $(h(n) \cdot e_{X\times X})_{n\in[N]}$ on the horizontal torus of $X \times X$. This is the place where we use our assumption that the sequence $(g(n) \cdot e_X)_{n\in[N]}$ is σ-equidistributed in X for σ suitably small and N sufficiently large. Using Proposition 6.5 one can then extract information about the group $H \cdot (G_2 \times G_2)$, and our second key ingredient is the purely algebraic Proposition 6.6 that utilizes this information in order to prove Property (7.13).

7.8.1. *Some notation.* To facilitate reading, before proceeding to the main body of the proof of Property (7.13), we introduce some notation and we organize some data and parameters that are spread out in this and the previous section. These parameters are going to be used in the definition of σ and the range of eligible N's used in the statement of Theorem 6.1.

Recall that $X := G/\Gamma$ and that G is endowed with the rational filtration $G_\bullet$. We denote by d the degree of $G_\bullet$, and thus all polynomial sequences under consideration have degree at most d. The distinct primes p, q were introduced in Section 7.4 and are bounded by a constant that depends on X and τ only.

We write $Z := G/(G_2\Gamma)$ for the horizontal torus of X and m for its dimension. We identify $Z = \mathbb{T}^m$, the identification being given by the Mal'cev basis of G. We write $\pi_Z \colon G \to Z = \mathbb{T}^m$ for the natural projection and let $\pi_{Z\times Z} := \pi_Z \times \pi_Z$. Let $A_5 := A_5(X)$ be a Lipschitz constant for the maps π_Z and $\pi_{Z\times Z}$.

In Lemma 6.2, for all $L_1, L_2 \in \mathbb{N}$ and sub-tori S_1, S_2 of $\mathbb{T}^m$, we defined a positive real $\delta_1 := \delta_1(S_1, S_2, L_1, L_2)$ and a positive integer $N_1 := N_1(S_1, S_2, L_1, L_2)$.

Proposition 6.5 defines finite families $\mathcal{F}_4(M)$, $M \in \mathbb{N}$, of sub-tori of $\mathbb{T}^m$ (which do not depend on ω_4), as well as integers $M_4 := M_4(\omega_4)$, $N_4 := N_4(\omega_4)$, and a positive real $\delta_4 := \delta_4(\omega_4)$. Also, if $T_1, \dots, T_d$ are sub-tori of $\mathbb{T}^m$, the sub-torus $R_{T_1,\dots,T_d}$ of T^{2m} was defined by (6.21).

In Section 7.5 we defined positive integers $M_5 := M_5(X, \tau)$, $N_5 := N_5(X, \tau)$, and, for $1 \leq M \leq M_5$, finite families $\mathcal{F}_5(M)$ of rational sub-nilmanifolds of $X \times X$ that increase with M. Then the family $\mathcal{F}_5(M_5)$ is the largest, and to each nilmanifold $Y \in \mathcal{F}_5(M_5)$ we assigned a rational subgroup H of $G\times G$. We let $\mathcal{F}_5'(M_5)$ denote the family of the corresponding subgroups of $G \times G$. Note that for every $H \in \mathcal{F}_5'(M_5)$, $\pi_{Z\times Z}(H)$ is a sub-torus of $\mathbb{T}^{2m}$.

7.8.2. *Defining σ and the range of eligible N.* We define the function $\omega_4 \colon \mathbb{N} \to \mathbb{R}^+$ by

$$\omega_4(M) := \min_{T_1,\dots,T_d\in\mathcal{F}_4(M);\, H\in\mathcal{F}_5'(M_5)} \bigl\{\delta_1\bigl(R_{T_1,\dots,T_d}, \pi_{Z\times Z}(H), M, A_5M_5\bigr)\bigr\}, \quad M \in \mathbb{N}, \tag{7.15}$$

and we let

$$\sigma := A_5^{-1}\delta_4(\omega_4). \tag{7.16}$$

Finally, we let

$$\widetilde{N}_1 := \max_{T_1,\dots,T_d\in\mathcal{F}_4(M_4);\, 1\leq M\leq M_4;\, H\in\mathcal{F}_5'(M_5)} \bigl\{N_1\bigl(R_{T_1,\dots,T_d}, \pi_{Z\times Z}(H), M, A_5M_5\bigr)\bigr\} \tag{7.17}$$

and

$$N_6 := \max\{\widetilde{N}_1, N_4, N_5\}. \tag{7.18}$$

Note that all the above defined parameters and the function $\omega_4\colon \mathbb{N} \to \mathbb{R}^+$ depend only on X and τ. Henceforth, we assume that

$$N \geq N_6 \ \text{ and the sequence } \ (g(n) \cdot e_X)_{n\in[N]} \ \text{ is totally } \ \sigma\text{-equidistributed in } X. \tag{7.19}$$

Under this assumption we plan to establish Property (7.13). Recall that Property (7.13) implies Property (7.12), and this in turn suffices to complete the proof of Theorem 6.1.

7.9. **Proof of the algebraic property.** Our first goal is to extract some information about the group $H \cdot (G_2 \times G_2)$. The idea is to compare two factorizations of the projection of the sequence h to $Z \times Z$. The first is the one we get by projecting the factorization of h given in Section 7.5 to $Z \times Z$. The second is the one we get after imposing an equidistribution assumption on the projection of the sequence g on Z and using Proposition 6.5. The two factorizations involve total equidistribution properties on the subtori $\pi_{Z\times Z}(H)$ and $R_{T_1,\dots,T_d}$. Assuming that we have "sufficient" total equidistribution in the second case, we are going to show using Lemma 6.2 that $R_{T_1,\dots,T_d} \subset \pi_{Z\times Z}(H)$. This then easily implies that the group H satisfies the hypothesis of Proposition 6.6, and the conclusion of this proposition then enables us to deduce Property (7.13). We proceed now to the details.

The sequence

$$\mathbf{g}(n) := \pi_Z(g(n)), \quad n \in \mathbb{N},$$

is a polynomial sequence of degree at most d in $Z = \mathbb{T}^m$ with $\pi_Z(g(0)) = 0$. As explained in Section 5.1, we can write this sequence as

$$\mathbf{g}(n) = \pi_Z(g(n)) = \boldsymbol{\alpha}_1 n + \cdots + \boldsymbol{\alpha}_d n^d$$

for some $\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_d \in \mathbb{T}^m$. Recall also that

$$h(n) = (g(pn), g(qn)), \quad n \in [N].$$

7.9.1. *First factorization of* $\mathbf{h}$. Recall that in Section 7.5 we defined an integer $M^* \leq M_5$, a nilmanifold $Y = H/(H \cap (\Gamma \times \Gamma))$ belonging to a family $\mathcal{F}_5(M^*) \subset \mathcal{F}_5(M_5)$ with $H \in \mathcal{F}'_5(M_5)$, and for $N \geq N_5$ a factorization

$$h(n) = \epsilon(n) h'(n) \gamma(n), \quad n \in [N],$$

that satisfies Properties (i)–(iii) stated in Section 7.5. Projecting both sides of this identity to $Z \times Z = \mathbb{T}^{2m}$ we get a factorization

$$\mathbf{h}(n) = \boldsymbol{\epsilon}_1(n) + \mathbf{h}_1(n) + \boldsymbol{\gamma}_1(n), \quad n \in [N], \tag{7.20}$$

where

$$\mathbf{h}(n) := \pi_{Z\times Z}(h(n)),$$
$$\boldsymbol{\epsilon}_1(n) := \pi_{Z\times Z}(\epsilon(n)),\ \mathbf{h}_1(n) := \pi_{Z\times Z}(h'(n)),\ \boldsymbol{\gamma}_1(n) := \pi_{Z\times Z}(\gamma(n)).$$

For $N \geq N_6$ (then $N \geq N_5$ by (7.18)) we have

(i) $\boldsymbol{\epsilon}_1\colon [N] \to \mathbb{T}^{2m}$ is $(A_5 M_5, N)$-smooth;
(ii) $\mathbf{h}_1\colon [N] \to \pi_{Z\times Z}(H)$ is a polynomial sequence of degree d;
(iii) $\boldsymbol{\gamma}_1\colon [N] \to \mathbb{T}^{2m}$ is M_5-rational and has period at most M_5,

where we used that the Lipschitz constant of $\pi_{Z\times Z}\colon G\times G\to Z\times Z$ is at most A_5. Here we do not use the equidistribution properties of the sequences h' and $\mathbf{h}_1$; what is important is that h' takes values in H, and thus $\mathbf{h}_1$ takes values in the sub-torus $\pi_{Z\times Z}(H)$ of $\mathbb{T}^{2m}$.

7.9.2. *Second factorization of* $\mathbf{h}$. Recall that $\mathbf{g}(n)=\pi_Z(g(n))=\boldsymbol{\alpha}_1 n+\cdots+\boldsymbol{\alpha}_d n^d$ and $\mathbf{h}(n)=\pi_{Z\times Z}(h(n))$. We have

$$\mathbf{h}(n)=(\mathbf{g}(pn),\mathbf{g}(qn)).$$

Our assumption (7.19) and the defining property of σ, given in (7.16), imply that the sequence

$$(\mathbf{g}(n))_{n\in[N]} \text{ is totally } \delta_4(\omega_4)\text{-equidistributed in } \mathbb{T}^m,$$

where we used the fact that the Lipschitz constant of $\pi_Z\colon G\to Z$ is at most A_5.

Hence, Proposition 6.5 applies. Recall that $M_4:=M_4(\omega_4)$, $N_4:=N_4(\omega_4)$ were defined by this proposition and that $N_6\geq N_4$ by (7.18). Therefore, by Proposition 6.5, for every $N\geq N_6$, there exists a positive integer $M^{**}\leq M_4$, and sub-tori $T_1,\dots,T_d\in\mathcal{F}_4(M^{**})$ of $\mathbb{T}^m$ such that

$$T_1+\cdots+T_d=\mathbb{T}^m, \tag{7.21}$$

and the sequence $(\mathbf{h}(n))_{n\in[N]}$ can be factorized as follows:

$$\mathbf{h}(n)=\boldsymbol{\epsilon}_2(n)+\mathbf{h}_2(n)+\boldsymbol{\gamma}_2(n),\quad n\in[N], \tag{7.22}$$

where $\boldsymbol{\epsilon}_2(n)$, $\mathbf{h}_2(n)$, $\boldsymbol{\gamma}_2(n)$ are polynomial sequences on $\mathbb{T}^{2m}$ such that

(i) $\boldsymbol{\epsilon}_2(n)$ is (M^{**},N)-smooth;

(ii) $\mathbf{h}_2(n)$ takes values and is totally $\omega_4(M^{**})$-equidistributed in the sub-torus

$$R_{T_1,\dots,T_d}:=\bigl\{(p\mathbf{x}_1+\cdots+p^d\mathbf{x}_d\,,\,q\mathbf{x}_1+\cdots+q^d\mathbf{x}_d)\colon \mathbf{x}_j\in T_j \text{ for } j=1,\dots,d\bigr\} \tag{7.23}$$

of $\mathbb{T}^{2m}$;

(iii) $\boldsymbol{\gamma}_2(n)$ is M^{**}-rational and has period at most M^{**}.

7.9.3. *Using the two factorizations.* For $N\geq N_6$, in Sections 7.9.1 and 7.9.2 we have defined the factorizations $\mathbf{h}(n)=\boldsymbol{\epsilon}_1(n)+\mathbf{h}_1(n)+\boldsymbol{\gamma}_1(n)$ and $\mathbf{h}(n)=\boldsymbol{\epsilon}_2(n)+\mathbf{h}_2(n)+\boldsymbol{\gamma}_2(n)$ and the integer $M^{**}\leq M_4$. By the defining property of ω_4, given in (7.15), and since $T_1,\dots,T_d\in\mathcal{F}_4(M^{**})$ and $H\in\mathcal{F}_5'(M_5)$, for $N\geq N_6$ we have

$$\omega_4(M^{**})\leq\delta_1(R_{T_1,\dots,T_d},\pi_{Z\times Z}(H),M^{**},A_5M_5).$$

We have $N\geq N_6\geq\widetilde{N}_1$ by (7.18). Furthermore, since $M^{**}\leq M_4$ we have $\mathcal{F}_4(M^{**})\subset\mathcal{F}_4(M_4)$; thus $T_1,\dots,T_d\in\mathcal{F}_4(M^{**})$, and by (7.17) we obtain that $N\geq\widetilde{N}_1$ which is greater than $N_1(R_{T_1,\dots,T_d},\pi_{Z\times Z}(H),M^{**},A_5M_5)$. Hence, Lemma 6.2 applies and gives that

$$R_{T_1,\dots,T_d}\subset\pi_{Z\times Z}(H). \tag{7.24}$$

7.9.4. *End of the proof.* For $j = 1, \dots, d$, let

$$G^j := \pi_Z^{-1}(T_j).$$

Note that for $j = 1, \dots, d$ the group G^j contains G_2, and thus G^j is a normal subgroup of G. It follows immediately from (7.21) that

$$G^1 \cdots G^d = G.$$

Let

$$W := \{(g_1^p \cdots g_d^{p^d}, g_1^q \cdots g_d^{q^d}) \colon g_1 \in G^1, \dots, g_d \in G^d\}.$$

Since for $j = 1, \dots, d$ we have $\pi_Z(G^j) = T_j$, it follows that $\pi_{Z\times Z}(W)$ is included in the torus $R_{T_1,\dots,T_d}$ given by (7.23). Hence, (7.24) gives that $\pi_{Z\times Z}(W) \subset \pi_{Z\times Z}(H)$ which implies that

$$W \subset H \cdot (G_2 \times G_2).$$

We have just established that if (7.19) holds, then the group H satisfies the hypothesis of Proposition 6.6. We deduce that H satisfies Property (7.13), and as explained in Section 7.7, this completes the proof of Theorem 6.1.

8. The U^s-structure theorems

In this section, our main goal is to prove Theorems 1.1 and 2.1. The proof of the second result is based on the following more informative variant of the first result.

Theorem 8.1 (Structure theorem for multiplicative functions I′)**.** *Let $s \geq 2$ and $\varepsilon > 0$. There exists $\theta_0 := \theta_0(s, \varepsilon)$ such that for $0 < \theta < \theta_0$ there exist positive integers N_0, Q, R, depending on s, ε, θ only, such that the following holds: For every $N \geq N_0$ and every $f \in \mathcal{M}$, the function f_N admits the decomposition*

$$f_N(n) = f_{N,\mathrm{st}}(n) + f_{N,\mathrm{un}}(n) \quad \textit{for every } n \in \mathbb{Z}_{\widetilde{N}},$$

where the functions $f_{N,\mathrm{st}}$ and $f_{N,\mathrm{un}}$ satisfy the following:

(i) *$f_{N,\mathrm{st}} = f_N * \phi_{N,\theta}$, where $\phi_{N,\theta}$ is the kernel on $\mathbb{Z}_{\widetilde{N}}$ defined by (3.8), is independent of f, and the convolution product is defined in $\mathbb{Z}_{\widetilde{N}}$;*

(ii) *If $\xi \in \mathbb{Z}_{\widetilde{N}}$ satisfies $\widehat{f}_{N,\mathrm{st}}(\xi) \neq 0$, then $\left|\frac{\xi}{\widetilde{N}} - \frac{p}{Q}\right| \leq \frac{R}{\widetilde{N}}$ for some $p \in \{0, \dots, Q-1\}$;*

(iii) *$|f_{N,\mathrm{st}}(n+Q) - f_{N,\mathrm{st}}(n)| \leq \frac{R}{\widetilde{N}}$ for every $n \in \mathbb{Z}_{\widetilde{N}}$, where $n+Q$ is taken $\bmod \widetilde{N}$;*

(iv) *$\|f_{N,\mathrm{un}}\|_{U^s(\mathbb{Z}_{\widetilde{N}})} \leq \varepsilon$.*

We stress the fact that the values of θ_0, Q, R do not depend on $f \in \mathcal{M}$ and $N \in \mathbb{N}$, N_0 does not depend on $f \in \mathcal{M}$, and these values are not the same as the ones given in Theorem 3.3. Recall that $\widetilde{N}$ is any prime between N and ℓN, where ℓ is a positive integer that is fixed throughout this argument.

The proof of Theorem 8.1 is given in Sections 8.2–8.9. The proof of Theorem 2.1 is given in Section 8.10. Before proceeding to the details we sketch the proof strategy for Theorem 8.1.

8.1. **Some preliminary remarks and proof strategy.** Our proof strategy follows in part the general ideas of an argument of Green and Tao from [26, 31] where U^s-uniformity of the Möbius function was established. In our case, we are faced with some important additional difficulties. The first is the need to establish discorrelation estimates for general multiplicative functions not just the Möbius, and it is important for applications to establish estimates with implied constants independent of the elements of $\mathcal{M}$. Another difficulty is that we cannot simply hope to prove that all multiplicative functions with zero mean are U^s-uniform, not even for $s = 2$ (see the examples in Section 2.1.3). To compensate for the lack of U^2-uniformity of a normalized multiplicative function f, we subtract from it a suitable "structured component" f_{st} given by Theorem 3.3, so that $f_{\text{un}} := f - f_{\text{st}}$ has extremely small U^2-norm. Our goal is then to show that f_{un} has small U^s-norm (see Proposition 8.2). In view of the U^s-inverse theorem (see Theorem 4.3), this would follow if we show that f_{un} has very small correlation with all $(s-1)$-step nilsequences of bounded complexity. This then becomes our main goal (see Proposition 8.3).

The factorization theorem for polynomial sequences (Theorem 5.6) practically allows us to treat correlation with major arc and minor arc nilsequences separately. Orthogonality to major arc (approximately periodic) nilsequences can be deduced from the U^2-uniformity of f_{un}. To handle the much more difficult case of minor arc (totally equidistributed) nilsequences, Theorem 6.1 comes to the rescue as it shows that such sequences are asymptotically orthogonal to all multiplicative functions. The function f_{un} is not multiplicative though, but this can be taken care of by the fact that $f_{\text{un}} = f - f_{\text{st}}$ and the fact that f_{st} can be recovered from f by taking a convolution product with a kernel. Using these properties it is possible to transfer estimates from f to f_{un}. Combining the above, we get the needed orthogonality of f_{un} to all $(s-1)$-step nilsequences of bounded complexity. Furthermore, a close inspection of the argument shows that all implied constants are independent of f. This suffices to complete the proof of Theorem 8.1.

Although the previous sketch communicates the basic ideas behind the proof of Theorem 8.1, the various results needed to implement this plan come with a significant number of parameters that one has to juggle with, making the bookkeeping rather cumbersome. We use Section 8.4 to organize some of these data.

We start the proof with two successive reductions. The first one uses the $s = 2$ case of Theorem 8.1 established in Theorem 3.3. The second uses the inverse theorem for the U^s-norms (see Theorem 4.3).

8.2. **Using the U^2-structure theorem.** An immediate consequence of Theorem 3.3 is that in order to prove Theorem 8.1 it suffices to prove the following result.

Proposition 8.2. *Let $s \in \mathbb{N}$ and $\varepsilon > 0$. There exists $\theta_0 > 0$ such that for every θ with $0 < \theta \leq \theta_0$, and every sufficiently large N, the decomposition $f_N = f_{N,\text{st}} + f_{N,\text{un}}$ associated to θ by Theorem* 3.3 *satisfies Properties* (i) *and* (ii) *of this theorem, and also*

$$\|f_{N,\text{un}}\|_{U^s(\mathbb{Z}_{\widetilde{N}})} \leq \varepsilon.$$

So our next goal becomes to prove Proposition 8.2.

8.3. **Using the inverse theorem for the U^s-norms.** An immediate consequence of the U^s-inverse theorem stated in Theorem 4.3 is that in order to prove Proposition 8.2 it suffices to prove the following result.

Proposition 8.3. *Let $X := G/\Gamma$ be a nilmanifold with the natural filtration and $\delta > 0$. There exists $\theta_0 > 0$ such that for every θ with $0 < \theta < \theta_0$ and every sufficiently large N, the decomposition $f_N = f_{N,\mathrm{st}} + f_{N,\mathrm{un}}$ associated to θ by Theorem* 3.3 *satisfies Properties* (i) *and* (ii) *of this theorem, and also*

$$\sup_{f,g,\Phi} \big| \mathbb{E}_{n\in[\widetilde{N}]} f_{N,\mathrm{un}}(n)\, \Phi(g^n \cdot e_X) \big| \leq \delta, \tag{8.1}$$

where f ranges over $\mathcal{M}$, g over G, and $\Phi\colon X \to \mathbb{C}$ over all functions with $\|\Phi\|_{\mathrm{Lip}(X)} \leq 1$.

We are going to prove Proposition 8.3 and thus finish the proof of Theorems 1.1 and 8.1 in Sections 8.4–8.9.

8.4. **Setting up the stage.** In this subsection, we define and organize some data that will be used in the proof of Proposition 8.3. We take some extra care to do this before the main body of its proof in order to make sure that there is no circularity in the admittedly complicated collection of choices involved.

As $X := G/\Gamma$ is going to be a fixed nilmanifold throughout the argument (recall that it was determined in Section 8.3 and depends only on s and ε), in order to ease notation:

Henceforth, we leave the dependence on X implicit.

Remember also that G is endowed with the natural filtration.

We first define several objects that depend on a positive real δ, and a positive integer parameter M that we consider for the moment as a free variable. The explicit choice of M takes place in Section 8.6 and depends on various other choices that will be made subsequently; what is important though is that it is bounded by a positive constant that depends only on δ (and, following our convention, on X). The families $\mathcal{F}(M)$, $M \in \mathbb{N}$, of nilmanifolds appearing in Theorem 5.6 play a central role in our constructions, and it is important to remark that they do not depend on the choice of the function ω in the same theorem, allowing us to postpone the definition of this function.

8.4.1. *Building families of nilmanifolds.* For the nilmanifold $X := G/\Gamma$, with the natural filtration, Theorem 5.6 defines for every $M \in \mathbb{N}$ a finite family $\mathcal{F}(M)$ of sub-nilmanifolds of X. In Section 5.5 we defined a finite subset $\Sigma(M)$ of G, and for every nilmanifold $X' := G'/(G' \cap \Gamma) \in \mathcal{F}(M)$ and every $\alpha \in \Sigma(M)$ we defined the sub-nilmanifold $X'_\alpha := G'_\alpha/(G'_\alpha \cap \Gamma)$ of X where $G'_\alpha := \alpha^{-1}G'\alpha$, and as usual, we consider the induced filtration in G'_α (which is not the natural filtration in G'_α). We let

$$\mathcal{F}'(M) := \{X'_\alpha \colon X' \in \mathcal{F}(M),\ \alpha \in \Sigma(M)\}.$$

Furthermore, let

$$H(M) \ \text{ and } \ \rho\colon \mathbb{N} \times \mathbb{R}_+ \to \mathbb{R}_+$$

be so that Properties (i)–(v) of Section 5.5.1 are satisfied.

8.4.2. *Restating the discorrelation estimates.* The following claim follows immediately from Theorem 6.1 applied to each nilmanifold in the finite family $\mathcal{F}'(M)$ and is the central ingredient in the proof of Proposition 8.3.

Claim 3. *Let $M \in \mathbb{N}$ and $\tau > 0$. Then there exist $\sigma := \sigma(M,\tau) > 0$ and $N_1 := N_1(M,\tau) \in \mathbb{N}$ such that for every $N \geq N_1$ the following property holds: Let $X'_\alpha \in \mathcal{F}'(M)$ and $h \in \mathrm{poly}(G'_{\alpha\bullet})$ be such that*

$$\big|\mathbb{E}_{n\in[N]} \mathbf{1}_P(n)\, f(n)\, \Phi(h(n+k)\cdot e_X)\big| \geq \tau,$$

for some $k \in \mathbb{N}$ with $|k| \leq N$, arithmetic progression $P \subset [N]$, $f \in \mathcal{M}$, and function $\Phi\colon X'_\alpha \to \mathbb{C}$ with $\|\Phi\|_{\mathrm{Lip}(X'_\alpha)} \leq 1$ and $\int_{X'_\alpha} \Phi\, dm_{X'_\alpha} = 0$. Then the sequence $(h(n)\cdot e_X)_{n\in[N]}$ is not totally σ-equidistributed in X'_α.

Remark. We stress that although the filtration in G is the natural one, the induced filtration in G'_α is not necessarily the natural one, and this is why in proving Theorem 6.1 we treat the more difficult case of arbitrary filtrations.

8.4.3. *Parameters related to the factorization theorem: ω and M_1.* We let

$$\lambda(M) := \frac{\delta^2}{128\, C_1 H(M)^2 M^2}, \tag{8.2}$$

where C_1 is the universal constant defined by Lemma A.6. We define also

$$\widetilde{\sigma}(M) := \sigma\big(M, \frac{C_1\lambda(M)}{4\, H(M)^2}\big), \tag{8.3}$$

where σ is the function defined in Claim 3 above.

Since $\rho(M,t)$ defined in (v) of Section 5.5.1 decreases to 0 as $t \to 0^+$ and M is fixed, there exists a function $\omega\colon \mathbb{N} \to \mathbb{R}_+$ that satisfies

$$\rho(M,\omega(M)) \leq \widetilde{\sigma}(M) \quad \text{for every} \quad M \in \mathbb{N}. \tag{8.4}$$

For this choice of ω, Theorem 5.6 associates to X a positive integer M_1. Note that as ω depends only on δ, the value of M_1 depends only on δ. We let

$$N_2 := \max_{1\leq M\leq M_1} \Big\{ N_1\Big(M, \frac{C_1\lambda(M)}{4\, H(M)^2}\Big)\Big\} \quad \text{where} \quad N_1(M,\tau) \quad \text{is defined in Claim 3;}$$

$$N_3 := \max_{1\leq M\leq M_1} \Big\{ \frac{16\, M^2 H(M)^2}{\delta} \Big\}.$$

We remark that these numbers too depend only on δ.

8.4.4. *Defining θ_0 and the eligible range of N.* Let $\delta > 0$. We let

$$\theta_0 := \min_{1\leq M\leq M_1} \lambda(M), \tag{8.5}$$

where the function λ is given by (8.2) and M_1 is defined in Section 8.4.3. Note that θ_0 depends on δ only.

Let θ now be such that

$$0 < \theta \leq \theta_0. \tag{8.6}$$

We define

$$N_4(\delta,\theta) := \max\{N_0(\theta), N_2(\delta), N_3(\delta)\}, \tag{8.7}$$

where $N_0(\theta)$ is the integer in the statement of Theorem 3.3 and N_2, N_3 were defined above and depend only on δ. Henceforth, we assume that

$$\theta \text{ satisfies (8.6) and } N \geq N_4 \text{ where } N_4 \text{ satisfies (8.7).} \tag{8.8}$$

8.5. **Our goal restated.** After setting up the stage we are now ready to enter the main body of the proof of Proposition 8.3. We argue by contradiction. For a fixed nilmanifold $X := G/\Gamma$ and $\delta > 0$ we let θ_0, θ, N_4 satisfy (8.5), (8.6), (8.7), respectively. Suppose that

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} f_{N,\mathrm{un}}(n)\,\Phi(g^n\cdot e_X)\big| > \delta, \tag{8.9}$$

for some integer $N \geq N_4$, $f \in \mathcal{M}$, $g \in G$, and function Φ with $\|\Phi\|_{\mathrm{Lip}(X)} \leq 1$. We are going to derive a contradiction.

8.6. **Using the factorization theorem.** Recall that G is endowed with its natural filtration. The sequence $(g^n)_{n\in[\widetilde{N}]}$ is a polynomial sequence in G, and thus there exists an integer M^* with

$$1 \leq M^* \leq M_1,$$

where M_1 is defined in Section 8.4.3, such that the sequence admits a factorization

$$g^n = \epsilon(n)g'(n)\gamma(n), \quad n \in [N],$$

as in Theorem 5.6; the sequence ϵ is $(M^*, \widetilde{N})$ smooth, the polynomial sequence g' takes values in G', $(g'(n)\cdot e_X)_{n\in[\widetilde{N}]}$ is totally $\omega(M^*)$-equidistributed in $X' \in \mathcal{F}(M^*)$, and the sequence γ is M^*-rational and $(\gamma(n)\cdot e_X)_{n\in[N]}$ has period at most M^*.

8.7. **Eliminating the smooth and periodic components.** From this point on, we work with the value of M^* and the factorization given in the previous subsection. We use the objects, notation, and estimates associated to this factorization in Sections 5.5.1 and 5.5.2 with $\widetilde{N}$ substituted for N.

By (5.12) we have $\|\Phi'\|_{\mathrm{Lip}(X)} \leq H(M^*)$, and Property (iv) of Section 5.5.1 gives that

$$\|\Phi'|_{X'_\alpha}\|_{\mathrm{Lip}(X'_\alpha)} \leq H(M^*)^2. \tag{8.10}$$

Furthermore, since the sequence $(g'(n)\cdot e_X)_{n\in[\widetilde{N}]}$ is totally $\omega(M^*)$-equidistributed in X', it follows from Property (v) of Section 5.5.1 that the sequence $(g'_\alpha(n)\cdot e_X)_{n\in[\widetilde{N}]}$ is totally $\rho(M^*,\omega(M^*))$-equidistributed in X'_α where $g'_\alpha \in \mathrm{poly}(G'_{\alpha\bullet})$. By (8.4), we get

(8.11) the sequence $(g'_\alpha(n)\cdot e_X)_{n\in[\widetilde{N}]}$ is totally $\widetilde{\sigma}(M^*)$-equidistributed in X'_α.

Since $\widetilde{N} \geq N_4 \geq 16\,H(M^*)^2 {M^*}^2/\delta$, following the argument in Section 5.5.2 we get that

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_{P_1}(n)\, f_{N,\mathrm{un}}(n)\,\Phi'(g'_\alpha(n)\cdot e_X)\big| \geq \frac{\delta^2}{64\,H(M^*)^2{M^*}^2} = 2C_1\lambda(M^*), \tag{8.12}$$

where $P_1 \subset [\widetilde{N}]$ is an arithmetic progression and g'_α, Φ' are defined in (5.10) and (5.11). The main advantage now is that the sequence g'_α is "totally equidistributed" on some sub-nilmanifold of X.

8.8. **Reducing to the zero integral case.** Our goal is to show that upon replacing Φ' with $\Phi' - z$, where z is some constant, we get a bound similar to (8.12). To this end, we make crucial use of the fact that the U^2-norm of $f_{N,\mathrm{un}}$ is suitably small, in fact, this is the step that determined our choice of the degree of U^2-uniformity θ_0 of $f_{N,\mathrm{un}}$ in Section 8.4.4. Recall Theorem 3.3 gives that $\|f_{N,\mathrm{un}}\|_{U^2(\mathbb{Z}_{\widetilde{N}})} \leq \theta \leq \theta_0$. We let

$$z := \int_{X'_\alpha} \Phi' \, dm_{X'_\alpha} \quad \text{and} \quad \Phi'_0 := \Phi' - z.$$

Then of course $\int_{X'_\alpha} \Phi'_0 \, dm_{X'_\alpha} = 0$.

Combining Lemma A.6 in the Appendix, Theorem 3.3, the definition (8.5) of θ_0, and that $1 \leq M^* \leq M_1$, we get

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_{P_1}(n)\, z\, f_{N,\mathrm{un}}(n)\big| \leq C_1 \|f_{N,\mathrm{un}}\|_{U^2(\mathbb{Z}_{\widetilde{N}})} \leq C_1\theta_0 \leq C_1\lambda(M^*).$$

From this estimate and (8.12) we deduce that

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_{P_1}(n)\, f_{N,\mathrm{un}}(n)\, \Phi'_0(g'_\alpha(n)\cdot e_X)\big| \geq C_1\lambda(M^*). \tag{8.13}$$

Moreover, the bound (8.10) remains valid with Φ'_0 substituted for Φ'.

8.9. **End of proof of Theorem 8.1.** We are now very close to completing the proof of Proposition 8.3 and hence of Theorem 8.1. To this end, we are going to combine the correlation estimate (8.13), the equidistribution result (8.11), and Claim 3 to deduce a contradiction.

Recall that $f_{N,\mathrm{st}} = f_N * \phi$ (the convolution is taken in $\mathbb{Z}_{\widetilde{N}}$) where ϕ is a kernel in $\mathbb{Z}_{\widetilde{N}}$, meaning a non-negative function with $\mathbb{E}_{n\in\mathbb{Z}_{\widetilde{N}}}\phi(n) = 1$. Since $f_{N,\mathrm{un}} = f_N - f_{N,\mathrm{st}}$, we can write $f_{N,\mathrm{un}} = f_N * \psi$, where the function ψ on $\mathbb{Z}_{\widetilde{N}}$ is given by

$$\psi(n) := \begin{cases} -\phi(n) & \text{if } n \neq 0 \mod \widetilde{N}; \\ \widetilde{N} - \phi(0) & \text{if } n = 0 \bmod \widetilde{N}, \end{cases}$$

and satisfies $\mathbb{E}_{n\in\mathbb{Z}_{\widetilde{N}}}|\psi(n)| \leq 2$.

We deduce from (8.13) that there exists an integer q with $0 \leq q < \widetilde{N}$ such that

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_{P_1}(n + q \bmod \widetilde{N})\, f_N(n)\, \Phi'_0(g'_\alpha(n + q \bmod \widetilde{N})\cdot e_X)\big| \geq \frac{C_1\lambda(M^*)}{2}, \tag{8.14}$$

where the residue class $n+q \bmod \widetilde{N}$ is taken in $[\widetilde{N}]$ instead of the more commonly used interval $[0, \widetilde{N})$. It follows that

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_{P_1}(n+k)\, \mathbf{1}_J(n)\, \mathbf{1}_{[N]}(n)\, f(n)\Phi'_0(g'_\alpha(n+k)\cdot e_X)\big| \geq \frac{C_1\lambda(M^*)}{4},$$

where either J is the interval $[\widetilde{N} - q]$ and $k := q$ or J is the interval $(\widetilde{N} - q, \widetilde{N}]$ and $k := q - \widetilde{N}$. In either case we have $|k| \leq \widetilde{N}$ and $\mathbf{1}_P(n+k)\, \mathbf{1}_J(n)\, \mathbf{1}_{[N]}(n) = \mathbf{1}_{P_2}(n)$ for some arithmetic progression $P_2 \subset [N]$. Thus, for some $k \in \mathbb{N}$ with $|k| \leq \widetilde{N}$ we have

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_{P_2}(n)\, f(n)\, \Phi'_0(g'_\alpha(n+k)\cdot e_X)\big| \geq \frac{C_1\lambda(M^*)}{4}. \tag{8.15}$$

Recall that $\int_{X'_\alpha} \Phi'_0\, dm_{X'_\alpha} = 0$, $\|\Phi'_0|_{X'_\alpha}\|_{\mathrm{Lip}(X'_\alpha)} \le H(M^*)^2$, and the nilmanifold X'_α belongs to the family $\mathcal{F}'$. By (8.8) and since $g'_\alpha \in \mathrm{poly}(G'_{\alpha\bullet})$ we can use Claim 3 with $\tau := C_1\lambda(M^*)/4\,H(M^*)^2$. We deduce that the sequence $(g'_\alpha(n)\cdot e_X)_{n\in[\widetilde{N}]}$ is not totally $\sigma(M^*, C_1\lambda(M^*)/4\,H(M^*)^2)$-equidistributed in X'_α. But, by definition (8.3) we have $\widetilde{\sigma}(M^*) = \sigma\big(M^*, C_1\lambda(M^*)/4\,H(M^*)^2\big)$ which contradicts (8.11).

Hence, our hypothesis (8.9) cannot hold, and as a consequence Proposition 8.3 is verified. This completes the proof of Theorem 8.1 and thus of Theorem 1.1. □

8.10. **End of proof of Theorem 2.1.** We are going to deduce Theorem 2.1 from Theorem 8.1 using an iterative argument of energy increment. To do this, we will use explicit properties of the kernels introduced in Section 3.3 and used to define the structured part $f_{N,\mathrm{st}}$ in Theorem 8.1. In particular, the following monotonicity of the Fourier coefficients of the kernels $\phi_{N,\theta}$ is key:

(8.16) if $\theta \ge \theta' > 0$ and $N \ge \max\{N_0(\theta), N_0(\theta')\}$,

$$\text{then for every } \xi \in \mathbb{Z}_{\widetilde{N}},\ \widehat{\phi_{N,\theta'}}(\xi) \ge \widehat{\phi_{N,\theta}}(\xi) \ge 0,$$

where N_0 is given by Theorem 3.3.

We fix a function $F\colon \mathbb{N}\times\mathbb{N}\times\mathbb{R}^+ \to \mathbb{R}^+$, an $\varepsilon > 0$, and a probability measure ν on the compact space $\mathcal{M}$ of multiplicative functions.

We define inductively a sequence (θ_j) of positive reals and sequences (N_j), (Q_j), (R_j) of positive integers as follows. We let $\theta_1 = N_1 = Q_1 = R_1 = 1$. Suppose that $j \ge 1$ and that the first j terms of the sequences are defined. We apply Theorem 8.1 with

$$\frac{1}{F(Q_j, R_j, \varepsilon)} \ \text{ substituted for } \ \varepsilon.$$

Theorem 8.1 provides a real $\theta_0 := \theta_0(j) > 0$, and we define

$$\theta_{j+1} := \min\{\theta_0, \theta_j\}.$$

Then Theorem 8.1 with θ_{j+1} substituted for θ provides integers N_0, Q, R, and we let $Q_{j+1} := Q$, $R_{j+1} := R$, and $N_{j+1} := \max\{N_0, N_j\}$. For every $N \ge N_{j+1}$ the kernel $\phi_{N,\theta_{j+1}}$ can be defined, and the functions

$$f_{j+1,N,\mathrm{st}} := f_N * \phi_{N,\theta_{j+1}} \quad \text{and} \quad f_{j+1,N,\mathrm{un}} := f_N - f_{j+1,N,\mathrm{st}}$$

satisfy Property (ii) of Theorem 3.3 and the conclusion of Theorem 8.1; that is,

$$|f_{j+1,N,\mathrm{st}}(n+Q_{j+1}) - f_{j+1,N,\mathrm{st}}(n)| \le \frac{R_{j+1}}{\widetilde{N}} \quad \text{for every } \ n \in \mathbb{Z}_{\widetilde{N}}; \tag{8.17}$$

$$\|f_{j+1,N,\mathrm{un}}\|_{U^s(\mathbb{Z}_{\widetilde{N}})} \le \frac{1}{F(Q_j, R_j, \varepsilon)}. \tag{8.18}$$

By construction, the sequence (N_j) increases and the sequence (θ_j) decreases with j. Let

$$J := 1 + \lceil 2\varepsilon^{-2} \rceil \quad \text{and} \quad N_0' := N_{J+1}.$$

For every $N \geq N_0'$ we have

$$\sum_{j=2}^{J} \int_{\mathcal{M}} \|f_{j+1,N,\mathrm{st}} - f_{j,N,\mathrm{st}}\|^2_{L^2(\mathbb{Z}_{\widetilde{N}})} \, d\nu(f) =$$

$$\int_{\mathcal{M}} \sum_{\xi \in \mathbb{Z}_{\widetilde{N}}} |\widehat{f_N}(\xi)|^2 \sum_{j=2}^{J} |\widehat{\phi_{N,\theta_{j+1}}}(\xi) - \widehat{\phi_{N,\theta_j}}(\xi)|^2 \, d\nu(f) \leq$$

$$2 \int_{\mathcal{M}} \sum_{\xi \in \mathbb{Z}_{\widetilde{N}}} |\widehat{f_N}(\xi)|^2 \sum_{j=2}^{J} \big(\widehat{\phi_{N,\theta_{j+1}}}(\xi) - \widehat{\phi_{N,\theta_j}}(\xi)\big) \, d\nu(f),$$

where to get the last estimate we used that $\theta_{j+1} \leq \theta_j$ and thus $\widehat{\phi_{N,\theta_{j+1}}}(\xi) \geq \widehat{\phi_{N,\theta_j}}(\xi) \geq 0$ for every $\xi \in \mathbb{Z}_{\widetilde{N}}$ by (8.16). Since $|\widehat{\phi_{N,\theta}}(\xi)| \leq 1$, the last quantity in the estimate is at most

$$2 \int_{\mathcal{M}} \sum_{\xi \in \mathbb{Z}_{\widetilde{N}}} |\widehat{f_N}(\xi)|^2 \, d\nu(f) \leq 2.$$

Therefore, for every $N \geq N_0'$ there exists $j_0 := j_0(F, N, \varepsilon, \nu)$ with

$$2 \leq j_0 \leq J \tag{8.19}$$

such that

$$\int_{\mathcal{M}} \|f_{j_0+1,N,\mathrm{st}} - f_{j_0,N,\mathrm{st}}\|^2_{L^2(\mathbb{Z}_{\widetilde{N}})} \, d\nu(f) \leq \frac{2}{J-1} \leq \varepsilon^2. \tag{8.20}$$

For $N \geq N_0'$, we let

$$\psi_{N,1} := \phi_{N,\theta_{j_0}}; \quad \psi_{N,2} := \phi_{N,\theta_{j_0+1}};$$
$$f_{N,\mathrm{st}} := f_N * \psi_{N,1} = f_{j_0,N,\mathrm{st}}; \quad f_{N,\mathrm{un}} := f_N - f_N * \psi_{N,2} = f_{j_0+1,N,\mathrm{un}};$$
$$f_{N,\mathrm{er}} := f_N * (\psi_{N,2} - \psi_{N,1}) = f_{j_0+1,N,\mathrm{st}} - f_{j_0,N,\mathrm{st}};$$
$$Q := Q_{j_0} \quad \text{and} \quad R := R_{j_0}.$$

Then we have the decomposition

$$f_N = f_{N,\mathrm{st}} + f_{N,\mathrm{un}} + f_{N,\mathrm{er}}.$$

Furthermore, Property (ii) of Theorem 2.1 follows from (8.17) (applied for $j := j_0 - 1$), Property (iv) follows from (8.18) (applied for $j := j_0$), and Property (v) follows from (8.20) and the Cauchy-Schwarz inequality. Last, it follows from (8.19) that the integers N_0', Q, R are bounded by a constant that depends on F and ε only. Thus, all the announced properties are satisfied, completing the proof of Theorem 2.1. □

9. Aperiodic multiplicative functions

In this section our goal is to prove the main results regarding aperiodic multiplicative functions, that is, Theorems 2.5 and 2.6.

9.1. Proof of Proposition 2.4.

We prove here the equivalence between the four characterizations of aperiodic multiplicative functions given in Proposition 2.4.

The equivalence of the first three properties of Proposition 2.4 is easy. The equivalence of (i) and (ii) is simple. Furthermore, (ii) immediately implies that for every periodic function a we have $\mathbb{E}_{n\in[N]}f(n)a(n)\to 0$ as $N\to+\infty$, and this in turn implies (iii). Going from (iii) to (ii) is also simple and standard: for $p,q\in\mathbb{N}$ we have

$$\mathbb{E}_{n\in[N]}f(n)\mathrm{e}(np/q)=\frac{1}{N}\sum_{d|q^r \text{ for some } r} f(d)\sum_{1\le n\le \frac{N}{d},(n,q)=1} f(n)\mathrm{e}(dnp/q),$$

and the function $\mathbf{1}_{\{k\colon (k,q)=1\}}(n)\mathrm{e}(dnp/q)$ has period q. The implication follows from the fact that any periodic function with period q that is supported in the set $\{k\colon (k,q)=1\}$ can be expressed as a finite linear combination of Dirichlet characters with period q.

The equivalence of (iii) and (iv) follows from Theorem 2.3. □

9.2. U^2 norm and aperiodic multiplicative functions.

We establish here two preliminary results used in the proof of Theorem 2.5.

Lemma 9.1. *Let $\varepsilon>0$. There exist $\delta:=\delta(\varepsilon)>0$ and $Q:=Q(\varepsilon)\in\mathbb{N}$ such that the following holds: If $f\in\mathcal{M}$ is a multiplicative function and $\limsup_{N\to+\infty}\|f\|_{U^2[N]}\ge\varepsilon$, then there exists $p\in\mathbb{N}$ with $0\le p<Q$ such that*

$$\limsup_{N\to+\infty}\Big|\mathbb{E}_{n\in[N]}f(n)\,\mathrm{e}\big(n\frac{p}{Q}\big)\Big|\ge\delta.$$

Remark. Note that the implication fails for arbitrary bounded sequences; consider, for example, a sequence of the form $(\mathrm{e}(n\alpha))_{n\in\mathbb{N}}$ where α is irrational in place of $(f(n))_{n\in\mathbb{N}}$.

Proof. Let $\varepsilon>0$ and $f\in\mathcal{M}$ be such that $\|f\|_{U^2[N]}\ge\varepsilon$ for infinitely many values of $N\in\mathbb{N}$. For these values of N let $\widetilde{N}$ be the smallest prime in the interval $(2N,4N]$. Since $\widetilde{N}\le 4N$, by Definition 2.1.1 of the norm $U^2[N]$ and Lemma A.3, we deduce that for these values of N we have

$$\|f_N\|_{U^2(\mathbb{Z}_{\widetilde{N}})}\ge\frac{\varepsilon}{4}. \tag{9.1}$$

We apply Corollary 3.2, with $\varepsilon^2/2^4$ in place of θ, for $\ell=4$ (see notation in Section 2.1.3) and for $\widetilde{N}$ chosen as above. We get some positive integers N_0, Q, V that depend only on ε such that (3.1) holds. Henceforth, we assume that $N\ge N_0$ is such that (9.1) holds.

Combining (2.4) and (9.1) we deduce that there exists $\xi\in\mathbb{Z}_{\widetilde{N}}$ such that

$$\big|\mathbb{E}_{n\in[\widetilde{N}]}f_N(n)\,\mathrm{e}\big(n\frac{\xi}{\widetilde{N}}\big)\big|\ge\frac{\varepsilon^2}{2^4}. \tag{9.2}$$

By implication (3.1) of Corollary 3.2, there exists $p\in\mathbb{N}$ with $0\le p\le Q$ such that

$$\Big|\frac{\xi}{\widetilde{N}}-\frac{p}{Q}\Big|\le\frac{V}{\widetilde{N}}. \tag{9.3}$$

We deduce that there exist infinitely many $N \in \mathbb{N}$ for which the above estimate holds for the same value of p. Henceforth, we further restrict ourselves to these values of N.

Let $N_1 := \lfloor \varepsilon^2 N/2^4 \rfloor$. Since $\widetilde{N} \geq 2N$, it follows from (9.2) that

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_{[N_1,N]}(n)\, f(n)\, \mathrm{e}\big(n\frac{\xi}{\widetilde{N}}\big)\big| \geq \frac{\varepsilon^2}{2^5}.$$

We let

$$\delta := \frac{\varepsilon^2}{2^{10}\,\pi V} \quad \text{and} \quad L := \Big\lfloor \delta N \Big\rfloor$$

and suppose that $N \in \mathbb{N}$ is sufficiently large so that $L \geq 2$. We partition the interval $[N_1, N]$ into intervals of length between L and $2L$. The number of these intervals is bounded by N/L, and thus one of them, say J, satisfies

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_J(n)\, f(n)\, \mathrm{e}\Big(n\frac{\xi}{\widetilde{N}}\Big)\big| \geq \frac{L}{N}\,\frac{\varepsilon^2}{2^5} \geq \frac{\delta\varepsilon^2}{2^6}. \tag{9.4}$$

Let n_0 be the first term of the interval. For every $n \in J$ we have

$$\Big|\mathrm{e}\big(n\big(\frac{\xi}{\widetilde{N}} - \frac{p}{Q}\big)\big) - \mathrm{e}\big(n_0\big(\frac{\xi}{\widetilde{N}} - \frac{p}{Q}\big)\big)\Big| \leq (n - n_0) 2\pi \Big|\frac{\xi}{\widetilde{N}} - \frac{p}{Q}\Big| \leq 4\pi V\delta,$$

where the last estimate follows from (9.3) and the fact that the length of J is at most $2L$. Combining this estimate with (9.4), and using again the fact that the length of J is at most $2L$, we get

$$\big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_J(n)\, f(n)\, \mathrm{e}\big(n\frac{p}{Q}\big)\big| \geq \frac{\delta\varepsilon^2}{2^6} - \frac{2L}{\widetilde{N}}\, 4\pi V\delta \geq \frac{\varepsilon^4}{2^{17}\,\pi V}.$$

Writing $J = (N_2, N_3]$ where $N_1 \leq N_2 \leq N_3 \leq N$ we have $\mathbf{1}_J = \mathbf{1}_{[N_3]} - \mathbf{1}_{[N_2]}$. For $N_4 := N_2$ or $N_4 := N_3$ we have

$$\big|\mathbb{E}_{n\in[N_4]} f(n)\, \mathrm{e}\big(n\frac{p}{Q}\big)\big| \geq \big|\mathbb{E}_{n\in[\widetilde{N}]} \mathbf{1}_{[N_4]}(n)\, f(n)\, \mathrm{e}\big(n\frac{p}{Q}\big)\big| \geq \frac{\varepsilon^4}{2^{18}\,\pi V}.$$

Since $N_4 \geq N_1 \geq \varepsilon^2 N/2^6$, we have that $N_4 \to +\infty$ as $N \to +\infty$ and we deduce that

$$\limsup_{N\to+\infty} \big|\mathbb{E}_{n\in[N]} f(n)\, \mathrm{e}\big(n\frac{p}{Q}\big)\big| > 0.$$

This completes the proof. □

Corollary 9.2. *If $f \in \mathcal{M}$, then f is aperiodic if and only if $\|f\|_{U^2[N]} \to 0$ as $N \to +\infty$.*

Proof. The necessity of the condition follows immediately from Lemma 9.1 and the characterization (ii) of aperiodic functions given in Proposition 2.4. The sufficiency follows from Lemma A.4, Lemma A.7, and the same characterization. □

9.3. **Proof of Theorem 2.5.** We move now to the proof of Theorem 2.5 which makes essential use of Theorem 8.1.

Proof of Theorem 2.5. For $N \in \mathbb{N}$ let $\widetilde{N}$ be the smallest prime in the interval $(2N, 4N]$. Suppose that $\|f\|_{U^s[N]}$ does not converge to 0 as $N \to +\infty$. Since $\widetilde{N} \leq 4N$, by Definition 2.1.1 of the norm $U^s[N]$ and Lemma A.3, we have that $\|f_N\|_{U^s(\mathbb{Z}_{\widetilde{N}})}$ does not tend to zero. As a consequence, there exists $\varepsilon > 0$ such that $\|f_N\|_{U^s(\mathbb{Z}_{\widetilde{N}})} \geq 2\varepsilon$ for infinitely many $N \in \mathbb{N}$ for which Theorem 8.1 applies (for $\ell = 4$). Let $Q, R \in \mathbb{N}$, $f_{N,\mathrm{st}}$, and $f_{N,\mathrm{un}}$ be given by Theorem 8.1 for this value of ε and these values of N. Then Property (iv) of Theorem 8.1 implies that $\|f_{N,\mathrm{st}}\|_{U^s(\mathbb{Z}_{\widetilde{N}})} \geq \varepsilon$.

By Property (ii) of Theorem 8.1, the cardinality of the spectrum of $f_{N,\mathrm{st}}$ (that is, the set of $\xi \in \mathbb{Z}_{\widetilde{N}}$ such that $\widehat{f_{N,\mathrm{st}}}(\xi) \neq 0$) is bounded by a positive real S that depends only on ε and s. Since the $U^s(\mathbb{Z}_{\widetilde{N}})$-norm of each function $\mathrm{e}(n\xi/\widetilde{N})$ is equal to 1 for every $s \geq 2$, it follows that there exist $\xi \in \mathbb{Z}_{\widetilde{N}}$ (depending on N) such that $|\widehat{f_{N,\mathrm{st}}}(\xi)| \geq \varepsilon/S$. By Property (i) of Theorem 8.1, we have $\widehat{f_{N,\mathrm{st}}}(\xi) = \widehat{\phi_N}(\xi)\widehat{f_N}(\xi)$, and since $|\widehat{f_{N,\mathrm{st}}}(\xi)| \geq \varepsilon/S$ and $|\widehat{\phi_N}(\xi)| \leq 1$, it follows that $|\widehat{f_N}(\xi)| \geq \varepsilon/S$. We deduce from (2.4) that $\|f\|_{U^2[N]} \geq \|f_N\|_{U^2(\mathbb{Z}_{\widetilde{N}})} \geq \varepsilon/S$. Hence, $\|f\|_{U^2[N]}$ does not converge to zero as $N \to +\infty$. Corollary 9.2 gives that f is not aperiodic, completing the proof. □

9.4. **Background on quadratic fields.** Our next goal is to prove Theorem 2.6. Here d is a positive integer, and we adopt the notation and refer the reader to Section 2.3.2 for the definition of τ_d, $\mathcal{N}(z)$, and Q_d.

We recall some classical facts about the ring $\mathbb{Z}[\tau_d]$. For every $N \in \mathbb{N}$ we let

$$B_N := \big\{(m,n) \in \mathbb{Z}^2 \colon Q_d(m,n) \leq N^2\big\}. \tag{9.5}$$

Recall that Q_d is a positive definite quadratic form, and in this case it is not hard to see that there exist constants $R_d \in \mathbb{N}$ and $c_d > 0$ such that

$$\Big[-\frac{N}{R_d}, \frac{N}{R_d}\Big]^2 \subset B_N \subset [-R_d N, R_d N]^2; \tag{9.6}$$

$$\lim_{x\to+\infty} \frac{1}{x}\big|\{z \in \mathbb{Z}[\tau_d] \colon \mathcal{N}(z) \leq x\}\big| = c_d. \tag{9.7}$$

The units of $\mathbb{Z}[\tau_d]$ are the elements of norm 1; their number is denoted by $N_1(d)$ and is equal to $2, 4$, or 6. In general, the ring $\mathbb{Z}[\tau_d]$ is not a principal ideal domain, but it is always a Dedekind domain, and has the property of unique factorization of ideals into prime ideals (see, for example, [64, Theorem 5.3.6]).

We say that $\alpha \in \mathbb{Z}[\tau_d]$ is a *prime element* if $\alpha \neq 0$ and the ideal (α) spanned by α is a prime ideal. Note that prime elements are irreducible, but the converse is not true in general. To avoid ambiguities, we do not abbreviate the expressions "prime integer," "prime ideal," and "prime element" of $\mathbb{Z}[\tau_d]$. Any two prime elements that generate the same ideal or, equivalently, that can be obtained from one another by multiplication by a unit are called *associates*, and we identify them. Some prime elements of $\mathbb{Z}[\tau_d]$ are prime integers and some others are not; we plan to work with non-integer prime elements only. We write

$$\mathcal{I}_d := \{\text{non-integer prime elements of } \mathbb{Z}[\tau_d]\}, \tag{9.8}$$

and throughout the argument we take into account the aforementioned identification; that is, we assume that no two elements in $\mathcal{I}_d$ are associates.

In the sequel we use the following immediate observations. If $\alpha \in \mathcal{I}_d$, then $\mathcal{N}(\alpha) = |\alpha|^2$ is a prime integer; if $z \in \mathbb{Z}[\tau_d]$ is such that $\mathcal{N}(\alpha)$ divides $\mathcal{N}(z)$, then z is a multiple of α or of $\overline{\alpha}$ (or of both). We have

$$\sum_{\alpha\in\mathcal{I}_d} \frac{1}{\mathcal{N}(\alpha)^2} \leq \sum_{z\in\mathbb{Z}[\tau_d], z\neq 0} \frac{1}{\mathcal{N}(z)^2} < +\infty, \tag{9.9}$$

where the convergence of the second series follows from (9.7). We also need a deeper result. We have

$$\sum_{\alpha \text{ prime element of } \mathbb{Z}[\tau_d]} \frac{1}{\mathcal{N}(\alpha)} = N_1(d) \sum_{\mathfrak{p} \text{ principal prime ideal, } \mathfrak{p}\neq\{0\}} \frac{1}{\mathcal{N}(\mathfrak{p})} = +\infty.$$

The divergence of the last series can be deduced from the Chebotarev density theorem (see, for example, [65, Theorem 13.4]), but also a much more elementary proof can be found, for example, on pages 148 and 149 of [64]. On the other hand, writing as usual $\mathbb{P}$ for the set of prime integers, we have

$$\sum_{p\in\mathbb{P}} \frac{1}{\mathcal{N}(p)} = \sum_{p\in\mathbb{P}} \frac{1}{p^2} < +\infty,$$

and thus we have

$$\sum_{\alpha\in\mathcal{I}_d} \frac{1}{\mathcal{N}(\alpha)} = +\infty. \tag{9.10}$$

9.5. **The Kátai orthogonality criterion for** $\mathbb{Z}[\tau_d]$**.** Next we prove a variant of the orthogonality criterion of Kátai (see Lemma 3.1) that works for the rings $\mathbb{Z}[\tau_d]$. Given the basic information about the rings $\mathbb{Z}[\tau_d]$ recorded above, the proof is a straightforward adaptation of the original argument of Kátai [46]; we give it for completeness.

Lemma 9.3 (Turán-Kubilius for subsets of $\mathbb{Z}[\tau_d]$)**.** *Let $\mathcal{P}$ be a finite subset of $\mathcal{I}_d$ and for $z \in \mathbb{Z}[\tau_d]$ let*

$$\mathcal{A} := \sum_{\alpha\in\mathcal{P}} \frac{1}{\mathcal{N}(\alpha)}, \quad \omega(z) := \sum_{\alpha\in\mathcal{P},\alpha|z} 1.$$

Then for every $x \in \mathbb{N}$ we have

$$\sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le x} |\omega(z) - \mathcal{A}| \ll \sqrt{\mathcal{A}}\cdot x + |\mathcal{P}|\cdot o(x),$$

where the implied constant and the $o(x)$ term depend only on d.

Proof. Using (9.7) and the Cauchy-Schwarz inequality, we see that it suffices to show that

$$\sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le x} (\omega(z) - \mathcal{A})^2 \ll \mathcal{A}\cdot x + |\mathcal{P}|^2\cdot o(x). \tag{9.11}$$

The left hand side is equal to

$$\sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le x} \omega(z)^2 + c_d\mathcal{A}^2\cdot x + o(\mathcal{A}^2 x) - 2\mathcal{A} \sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le x} \omega(z), \tag{9.12}$$

where the constant c_d is defined in (9.7).

As every $z \in \mathbb{Z}[\tau_d]$ that is divisible by some $\alpha \in \mathcal{P}$ is of the form αw for some $w \in \mathbb{Z}[\tau_d]$, using (9.7) we see that

$$\sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le x,\ \alpha|z} 1 = \left|\left\{w \in \mathbb{Z}[\tau_d]\colon \mathcal{N}(w) \le \frac{x}{\mathcal{N}(\alpha)}\right\}\right| = c_d\frac{x}{\mathcal{N}(\alpha)} + o(x).$$

Summing over $\alpha \in \mathcal{P}$ we find that

$$\sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le x} \omega(z) = c_d\mathcal{A}\cdot x + |\mathcal{P}|\cdot o(x). \tag{9.13}$$

In the same way,

$$\sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le x} \omega(z)^2 = \sum_{\alpha\neq\beta\in\mathcal{P}} \sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le x,\ \alpha\beta|z} 1 + \sum_{\alpha\in\mathcal{P}} \sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le x,\ \alpha|z} 1$$

$$= \sum_{\alpha,\beta\in\mathcal{P}} \Big(c_d\frac{x}{\mathcal{N}(\alpha)\mathcal{N}(\beta)} + o(x)\Big) - \sum_{\alpha\in\mathcal{P}} \Big(c_d\frac{x}{\mathcal{N}(\alpha)^2} + o(x)\Big) + \sum_{\alpha\in\mathcal{P}} \Big(c_d\frac{x}{\mathcal{N}(\alpha)} + o(x)\Big)$$

$$= c_d\mathcal{A}^2\cdot x + \mathcal{A}\cdot O(x) + |\mathcal{P}|^2\cdot o(x)$$

by (9.9). Combining this formula with (9.12) and (9.13) we get (9.11), completing the proof. □

Definition. We say that a function $f\colon \mathbb{Z}[\tau_d] \to \mathbb{C}$ is *multiplicative* if $f(zz') = f(z)f(z')$ whenever $\mathcal{N}(z)$ and $\mathcal{N}(z')$ are relatively prime. We denote by $\mathcal{M}_d$ the family of multiplicative functions $f\colon \mathbb{Z}[\tau_d] \to \mathbb{C}$ with modulus at most 1. We remark that for $g \in \mathcal{M}$ and $r \in \mathbb{N}$ the function $f\colon z \mapsto g(\mathcal{N}(z)^r)$ belongs to $\mathcal{M}_d$.

Lemma 9.4 (Kátai estimate for $\mathbb{Z}[\tau_d]$)**.** *Let $f \in \mathcal{M}_d$ be a multiplicative function and $h\colon \mathbb{Z}[\tau_d] \to \mathbb{C}$ be an arbitrary function of modulus at most* 1*. For $x \in \mathbb{N}$, let also*

$$S(x) := \sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le x} f(z)\, h(z),$$

$$C(x) := \sum_{\alpha,\beta\in\mathcal{P},\alpha\neq\beta} \Big| \sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le\min\{x/\mathcal{N}(\alpha),\ x/\mathcal{N}(\beta)\}} f(\alpha z)\cdot\overline{f(\beta z)}\Big|,$$

where $\mathcal{P}$ is a finite subset of $\mathcal{I}_d$. Then we have the estimate

$$\Big|\frac{S(x)}{x}\Big|^2 \ll \frac{1}{\mathcal{A}} + \frac{1}{\mathcal{A}^2} + \frac{|\mathcal{P}|^2}{\mathcal{A}^2}\, o(1) + \frac{1}{\mathcal{A}^2}\frac{C(x)}{x},$$

where $\mathcal{A} := \sum_{\alpha\in\mathcal{P}} \mathcal{N}(\alpha)^{-1}$, the implied constant, and the $o(1)$ term depend only on d.

Proof. Let $\omega(z)$ be defined as in Lemma 9.3 and

$$S'(x) := \sum_{w\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(w)\le x} f(w)\, h(w)\, \omega(w).$$

From Lemma 9.3 we deduce that

$$|S'(x) - \mathcal{A}\cdot S(x)| \ll \sqrt{\mathcal{A}}\cdot x + |\mathcal{P}|\cdot o(x).$$

The formula defining $S'(x)$ can be rewritten as

$$S'(x) = \sum_{w\in\mathbb{Z}[\tau_d],\alpha\in\mathcal{P}\colon \alpha|w \text{ and } \mathcal{N}(w)\le x} f(w)\, h(w) = \sum_{z\in\mathbb{Z}[\tau_d],\ \alpha\in\mathcal{P}\colon \mathcal{N}(\alpha z)\le x} f(\alpha z)\, h(\alpha z).$$

In this sum, the term corresponding to a pair (z, α) is equal to $f(\alpha)f(z)h(\alpha z)$ except if $\mathcal{N}(\alpha)$ and $\mathcal{N}(z)$ are not relatively prime. Since $\mathcal{N}(\alpha)$ is a prime integer, this holds only if $\mathcal{N}(\alpha)$ divides $\mathcal{N}(z)$, that is, if α or $\overline{\alpha}$ divides z. We let

$$S''(x) := \sum_{z\in\mathbb{Z}[\tau_d],\ \alpha\in\mathcal{P}\colon \mathcal{N}(\alpha z)\le x} f(\alpha)\, f(z)\, h(\alpha z).$$

Since $|f| \leq 1$ and $|h| \leq 1$, it follows that

$$\begin{aligned}|S(x) - S''(x)| &\leq 2\,\big|\big\{\big(z,\alpha\big) \in \mathbb{Z}[\tau_d] \times \mathcal{P}\colon \alpha \text{ or } \overline{\alpha} \text{ divides } z \text{ and } \mathcal{N}(\alpha z) \leq x\big\}\big| \\ &\leq 4 \sum_{\alpha\in\mathcal{P}} \big|\big\{z \in \mathbb{Z}[\tau_d]\colon \mathcal{N}(z) \leq x\mathcal{N}(\alpha)^{-2}\big\}\big| \leq 4x \sum_{\alpha\in\mathcal{P}} \mathcal{N}(\alpha)^{-2} \ll x\end{aligned}$$

by (9.9), where the implied constant depends only on d.

We rewrite $S''(x)$ as

$$S''(x) = \sum_{z\in\mathbb{Z}[\tau_d]\colon |\mathcal{N}(z)|\leq x} f(z) \sum_{\alpha\in\mathcal{P}\colon \mathcal{N}(\alpha)\leq x/\mathcal{N}(z)} f(\alpha)\, h(\alpha z).$$

Using (9.7) and the Cauchy-Schwarz inequality we deduce that

$$|S''(x)|^2 \ll x \sum_{z\in\mathbb{Z}[\tau_d]\colon |\mathcal{N}(z)|\leq x} \Big| \sum_{\alpha\in\mathcal{P}\colon \mathcal{N}(\alpha)\leq x/|\mathcal{N}(z)|} f(\alpha)\, h(\alpha z)\Big|^2.$$

Expanding the square, we get that this last expression is equal to

$$x \sum_{\alpha,\beta\in\mathcal{P}} \sum_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\leq x/\mathcal{N}(\alpha),\ \mathcal{N}(z)\leq x/\mathcal{N}(\beta)} f(\alpha)\, \overline{f(\beta)}\, h(\alpha z)\, \overline{h(\beta z)}.$$

By (9.7), the contribution of the diagonal terms with $\alpha = \beta$ is at most

$$x \sum_{z\in\mathbb{Z}[\tau_d],\ \alpha\in\mathcal{P}\colon \mathcal{N}(z)\leq x/\mathcal{N}(\alpha)} 1 \ll \mathcal{A}\cdot x^2,$$

and the contribution of the off diagonal terms is bounded by $x\,C(x)$, where $C(x)$ was defined in the statement.

Combining the previous estimates we get that

$$\begin{aligned}|\mathcal{A}\cdot S(x)|^2 &\ll |S'(x) - \mathcal{A}\cdot S(x)|^2 + |S'(x) - S''(x)|^2 + |S''(x)|^2 \\ &\ll \mathcal{A}\cdot x^2 + |\mathcal{P}|^2\, o(x^2) + x^2 + \mathcal{A}\, x^2 + x\, C(x).\end{aligned}$$

The asserted estimates follows upon dividing by $\mathcal{A}^2 \cdot x^2$. □

Using the previous estimate and (9.7) we get as an immediate corollary the following orthogonality criterion.

Proposition 9.5 (Kátai orthogonality criterion for $\mathbb{Z}[\tau_d]$)**.** *Let $d \in \mathbb{N}$, τ_d, and $\mathcal{I}_d$ be as above, and let $\mathcal{P}$ be a subset of $\mathcal{I}_d$ such that*

$$\sum_{\alpha\in\mathcal{P}} \frac{1}{\mathcal{N}(\alpha)} = +\infty.$$

Let $h_x\colon \mathbb{Z}[\tau_d] \to \mathbb{C}$, $x \in \mathbb{N}$, be arbitrary functions of modulus at most 1 such that

$$\lim_{x\to+\infty} \mathbb{E}_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\leq x}\, h_x(\alpha z)\cdot \overline{h_x(\beta z)} = 0$$

for every $\alpha, \beta \in \mathcal{P}$ with $\alpha \neq \beta$. Then

$$\lim_{x\to+\infty} \sup_{f\in\mathcal{M}_d} \big|\mathbb{E}_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\leq x}\, f(z)\, h_x(z)\big| = 0.$$

Remark. By (9.10) the assumption $\sum_{\alpha\in\mathcal{P}} \mathcal{N}(\alpha)^{-1} = +\infty$ is satisfied for $\mathcal{P} := \mathcal{I}_d$ and also for any set $\mathcal{P}$ obtained by removing finitely many elements from $\mathcal{I}_d$.

9.6. **Estimates involving Gowers norms.** The following lemma will be crucial in the proof of Theorem 2.6. The method of proof is classical (see, for example, [29, Proof of Proposition 7.1]), and we summarize it for completeness. Lemma 10.7 below is proved in a similar fashion.

Lemma 9.6. *Let $s \in \mathbb{N}$ and $L_j(m,n)$, $j = 1,\ldots,s$, be linear forms with integer coefficients and suppose that either $s = 1$ or $s > 1$ and the linear forms L_1, L_j are linearly independent for $j = 2,\ldots,s$. For $j = 1,\ldots,s$, let $h_j\colon \mathbb{Z} \to \mathbb{C}$ be bounded functions. Suppose that h_1 is an even function and $\|h_1\|_{U^{s'}[N]} \to 0$ as $N \to +\infty$ where $s' := \max\{s-1, 2\}$. Then*

$$\lim_{N\to+\infty} \mathbb{E}_{1\le m,n\le N} \mathbf{1}_{K_N}(m,n) \prod_{j=1}^{s} h_j(L_j(m,n)) = 0, \tag{9.14}$$

where K_N, $N \in \mathbb{N}$, are arbitrary convex subsets of $[-N,N]^2$.

Proof. Without loss of generality we can assume that $|h_j| \le 1$ for $j = 1,\ldots,s$. Furthermore, we can assume that the linear forms L_j are pairwise independent. Indeed, if for some distinct i,j the linear forms L_i and L_j are integer multiples of the same linear form L, then by hypothesis, i and j must both be greater than 1, and we can write $h_i(L_i(m,n))h_j(L_j(m,n)) = h(L(m,n))$ for some bounded function h.

Next, as we want to have Fourier analysis tools available, we reduce matters to averages on a cyclic group. For $j = 1,\ldots,s$, we write $L_j(m,n) = \kappa_j m + \lambda_j n$ where $\kappa_j, \lambda_j \in \mathbb{Z}$. For every $N \in \mathbb{N}$, let $\widetilde{N}$ be the smallest prime such that $\widetilde{N} > 2N$, $\widetilde{N} > 2|L_j(m,n)|$ for $m,n \in [N]$ and $j = 1,\ldots,s$, and $\widetilde{N} > |\kappa_i\lambda_j - \lambda_i\kappa_j|$ for $i,j = 1,\ldots,s$. Then $\widetilde{N}/N$ is bounded by a constant that depends on the linear forms $L_1,\ldots,L_s$ only.

For $j = 1,\ldots,s$, let $\widetilde{h}_j\colon \mathbb{N} \to \mathbb{C}$ be periodic of period $\widetilde{N}$ and equal to h_j on the interval $\big[-\lceil \widetilde{N}/2\rceil, \lfloor \widetilde{N}/2\rfloor\big)$. For $(m,n) \in K_N$, since $|L_j(m,n)| < \widetilde{N}/2$, we have $h_j(L_j(m,n)) = \widetilde{h}_j(L_i(m,n))$. Hence,

$$\mathbb{E}_{m,n\in[N]} \mathbf{1}_{K_N}(m,n) \prod_{j=1}^{s} h_j(L_j(m,n)) = \Big(\frac{\widetilde{N}}{N}\Big)^2 \mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}} \mathbf{1}_{K_N}(m,n) \prod_{j=1}^{s} \widetilde{h}_j(L_j(m,n)). \tag{9.15}$$

Henceforth, we work with the right hand side and assume that the linear forms L_j and the functions $\widetilde{h}_j$ are defined on $\mathbb{Z}_{\widetilde{N}}$. We first prove that the right hand side of (9.15) converges to 0 as $N \to +\infty$ under the assumption that

$$\lim_{N\to+\infty} \|\widetilde{h}_1\|_{U^{s'}(\mathbb{Z}_{\widetilde{N}})} = 0, \tag{9.16}$$

and then we verify that this assumption is satisfied (the assumption that h_1 is even is only needed here).

Our next goal is to remove the cutoff $\mathbf{1}_{K_N}(m,n)$. One can follow the exact same method as in [29, Proposition 7.1]; we skip the details and only give a sketch. The idea is to imbed $\mathbb{Z}_{\widetilde{N}}^2$ in the torus $\mathbb{T}^2$ in the natural way and to represent K_N as the intersection of $\mathbb{Z}_{\widetilde{N}}^2$ with a "convex" subset of $\mathbb{T}^2$. This convex set is then approximated by a sufficiently regular function on $\mathbb{T}^2$ which is in turn approximated by a trigonometric polynomial with bounded coefficients and spectrum of bounded cardinality (with respect to N). After these reductions, we are left with showing that

$$\lim_{N\to+\infty}\max_{\eta,\xi\in\mathbb{Z}_{\widetilde{N}}}\Big|\mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}}\mathrm{e}\big(m\frac{\eta}{\widetilde{N}}+n\frac{\xi}{\widetilde{N}}\big)\prod_{j=1}^{s}\widetilde{h}_j(L_j(m,n))\Big|=0.$$

We now show that we can restrict ourselves to the case where $\xi=\eta=0$. Indeed, if the linear form $m\eta+n\xi$ does not belong to the linear span of the forms L_j, then the average vanishes. On the other hand, if the linear form $m\eta+n\xi$ belongs to the linear span of the forms L_j, then we can remove the exponential term by multiplying each function h_j by a complex exponential of the form $\mathrm{e}(\theta_j n/\widetilde{N})$ for some $\theta_j\in\mathbb{Z}_{\widetilde{N}}$, and since $s'\geq 2$, this modification does not change the $U^{s'}(\mathbb{Z}_{\widetilde{N}})$-norm of the functions. Note that now h_1 is not necessarily even, but we no longer need this assumption. Therefore, we are reduced to proving that

$$\lim_{N\to+\infty}\mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}}\prod_{j=1}^{s}\widetilde{h}_j(L_j(m,n))=0. \tag{9.17}$$

The pairwise independence of the linear forms L_j, combined with the last condition on $\widetilde{N}$, implies that the forms L_j on $\mathbb{Z}_{\widetilde{N}}\times\mathbb{Z}_{\widetilde{N}}$ are pairwise linearly independent over $\mathbb{Z}_{\widetilde{N}}$. Using this and an iteration of the Cauchy-Schwarz inequality (see, for example, [72, Theorem 3.1]), we get

$$\big|\mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}}\prod_{j=1}^{s}\widetilde{h}_j(L_j(m,n))\big|\leq\|\widetilde{h}_1\|_{U^{s'}(\mathbb{Z}_{\widetilde{N}})}.$$

From this estimate and (9.16) it follows that (9.17) holds.

It thus remains to verify that (9.16) holds. We write

$$\mathbb{Z}_{\widetilde{N}}=I\cup J\cup\{0\}\quad\text{where}\quad I:=[1,\lfloor\widetilde{N}/2\rfloor)\quad\text{and}\quad J:=[\lfloor N/2\rfloor,N).$$

Since $\|\mathbf{1}_{\{0\}}\cdot\widetilde{h}_1\|_{U^{s'}(\mathbb{Z}_{\widetilde{N}})}\to 0$ when $N\to+\infty$, it suffices to show that $\|\mathbf{1}_I\cdot h_1\|_{U^{s'}(\mathbb{Z}_{\widetilde{N}})}\to 0$ and $\|\mathbf{1}_J\cdot h_1\|_{U^{s'}(\mathbb{Z}_{\widetilde{N}})}\to 0$. By hypothesis, $\|h_1\|_{U^{s'}[\widetilde{N}]}\to 0$ and by Lemma A.4 in the Appendix, $\|h_1\|_{U^{s'}(\mathbb{Z}_{\widetilde{N}})}\to 0$. By Lemma A.1 in the Appendix, $\|\mathbf{1}_I\cdot h_1\|_{U^{s'}(\mathbb{Z}_{\widetilde{N}})}\to 0$. Since $\widetilde{h}_1$ and h_1 coincide on I, we have $\|\mathbf{1}_I\cdot\widetilde{h}_1\|_{U^{s'}(\mathbb{Z}_{\widetilde{N}})}\to 0$. By assumption, h_1 is an even function; hence, for $n\in J$ we have $\widetilde{h}_1(n)=\widetilde{h}_1(n-N)=h_1(n-N)=h_1(N-n)$. The map $n\mapsto N-n$ maps the interval J onto the interval $J':=[1,\lceil\widetilde{N}/2\rceil]$, and thus $\|\mathbf{1}_J\cdot\widetilde{h}_1\|_{U^{s'}(\mathbb{Z}_{\widetilde{N}})}=\|\mathbf{1}_{J'}\cdot h_1\|_{U^{s'}(\mathbb{Z}_{\widetilde{N}})}$. The last quantity tends to 0 as $N\to+\infty$ by the same argument as above. This completes the proof. □

9.7. **Proof of Theorem 2.6.** Theorem 2.6 follows from the following stronger result.

Theorem 9.7. *For $s\in\mathbb{N}$ let the linear forms $L_1,\dots,L_s$ and the quadratic form Q be as in the statement of Theorem* 2.6. *Furthermore, let $g\in\mathcal{M}$ be arbitrary, let $f_1\in\mathcal{M}$ be aperiodic, and suppose that both multiplicative functions are extended to even functions on $\mathbb{Z}$. If $s\geq 2$, let also $f_2,\dots,f_s\colon\mathbb{Z}\to\mathbb{C}$ be arbitrary bounded functions. Then*

$$\lim_{N\to+\infty}\mathbb{E}_{1\leq m,n\leq N}\,\mathbf{1}_{K_N}(m,n)\,g(Q(m,n))\prod_{j=1}^{s}f_j(L_j(m,n))=0,$$

where K_N, $N\in\mathbb{N}$, are arbitrary convex subsets of $[-N,N]^2$.

Remark. We deduce that if $Q, L_1, \ldots, L_s$ are as above, $r \in \mathbb{N}$, $R_j \in \mathbb{Z}[t]$ are arbitrary polynomials, and $f \in \mathcal{M}$ is an aperiodic completely multiplicative function, then for $P(m,n) = Q(m,n)^r L_1(m,n) \prod_{j=2}^s R_j(L_j(m,n))$ we have

$$\lim_{N\to+\infty} \mathbb{E}_{1\le m,n\le N} f(P(m,n)) = 0.$$

Proof. The proof proceeds in several steps.

9.7.1. *Reduction to the quadratic form Q_d.* By assumption, the bilinear form Q can be written as $Q(m,n) = Q_d(F(m,n))$ where $d \in \mathbb{N}$ and $F\colon \mathbb{Z}^2 \to \mathbb{Z}^2$ is given by a 2×2 matrix with integer entries and determinant ± 1.

For $j = 1, \ldots, s$ we let $L'_j(m,n) := L_j(F^{-1}(m,n))$. For every $N \in \mathbb{N}$, let $K'_N := F([N]\times[N])$ and $K''_N := F(K_N)$. There exists $R \in \mathbb{N}$, which depends only on the linear map F, such that $K'_N \subset [-RN, RN]^2$ for every $N \in \mathbb{N}$, and thus by (9.6) we have $K''_N \subset K'_N \subset B_{RR_dN}$. The average in the statement can be rewritten as

$$\frac{|B_{RR_dN}|}{N^2} \mathbb{E}_{(m,n)\in B_{RR_dN}} \mathbf{1}_{K''_N}(m,n)\, g(Q_d(m,n)) \prod_{j=1}^s f_j(L'_j(m,n)),$$

where B_N is defined in (9.5). By (9.6) we have $|B_{RR_dN}| = O(N^2)$ and substituting L'_j for L_j for $j = 1, \ldots, s$ and K''_N for K_N we have reduced matters to showing that

$$\lim_{N\to+\infty} \mathbb{E}_{(m,n)\in B_N} \mathbf{1}_{K_N}(m,n)\, g(Q_d(m,n)) \prod_{j=1}^s f_j(L_j(m,n)) = 0, \tag{9.18}$$

where K_N are convex subsets of B_N for $N \in \mathbb{N}$.

9.7.2. *Applying the Kátai orthogonality criterion.* In the rest of the proof, we identify $\mathbb{Z}^2$ with $\mathbb{Z}[\tau_d]$, by mapping $(m,n) \in \mathbb{Z}^2$ to $m + n\tau_d$.

For $j = 1, \ldots, s$, we write $L_j(m,n) = \kappa_j m + \lambda_j n$ and let

$$\zeta_j := \begin{cases} 2(\lambda_j - \kappa_j + \kappa_j\tau_d) & \text{if } d = 1 \bmod 4; \\ \lambda_j + \kappa_j\tau_d & \text{otherwise.} \end{cases}$$

For every $(m,n) \in \mathbb{Z}^2$ we have

$$L_j(m + n\tau_d) = \frac{1}{\sqrt{-d}} \operatorname{Im}(\zeta_j(m + n\tau_d)), \tag{9.19}$$

and thus we are reduced to proving that

$$\lim_{N\to+\infty} \mathbb{E}_{z\in\mathbb{Z}[\tau_d]\colon \mathcal{N}(z)\le N^2}\, g(\mathcal{N}(z))\, \mathbf{1}_{K_N}(z) \prod_{i=1}^s f_i\Big(\frac{1}{\sqrt{-d}} \operatorname{Im}(\zeta_j z)\Big) = 0. \tag{9.20}$$

We use Proposition 9.5 with

$$\mathcal{P} := \big\{\alpha \in \mathcal{I}_d,\ \alpha \text{ and } \overline{\alpha} \text{ are non-associates and do not divide } \zeta_j \text{ for } j = 1, \ldots, s.\big\} \tag{9.21}$$

There are clearly only finitely many $\alpha \in \mathcal{I}_d$ that do not satisfy the second condition. Moreover, since we work on a quadratic number field, α and $\overline{\alpha}$ are associates only if the ideal (α) ramifies, and this can happen for finitely many $\alpha \in \mathcal{I}_d$ by Dedekind's

Theorem (see, for example, [64, Section 5.4] or [65, Proposition 8.4]). By (9.10), we have that

$$\sum_{\alpha\in\mathcal{P}} \frac{1}{\mathcal{N}(\alpha)} = +\infty. \tag{9.22}$$

Note that $z \mapsto g(\mathcal{N}(z))$ defines a multiplicative function on $\mathbb{Z}[\tau_d]$. We apply Proposition 9.5, and we are left with showing the following claim.

Claim 4. *If α and β are distinct elements of $\mathcal{P}$, then*

(9.23)

$$\lim_{N\to+\infty} \mathbb{E}_{z\in\mathbb{Z}[\tau_d],\ \mathcal{N}(z)\le N^2} \mathbf{1}_{K_N^*}(z) \prod_{j=1}^{s} f_j\big(\frac{1}{\sqrt{-d}}\ \mathrm{Im}(\zeta_j\alpha z)\big)\,\overline{f}_j\big(\frac{1}{\sqrt{-d}}\ \mathrm{Im}(\zeta_j\beta z)\big) = 0,$$

where

$$K_N^* := \alpha^{-1}K_N \cap \beta^{-1}K_N \subset [-N,N]^2. \tag{9.24}$$

9.7.3. *Independence of the linear forms.* For $j = 1,\dots,s$, we let

$$\Lambda_j(z) := \frac{1}{\sqrt{-d}}\,\mathrm{Im}(\zeta_j\alpha z) \ \text{ and } \ \Lambda'_j(z) := \frac{1}{\sqrt{-d}}\,\mathrm{Im}(\zeta_j\beta z).$$

Identifying $\mathbb{Z}[\tau_d]$ with $\mathbb{Z}^2$, these $2s$ maps can be thought of as linear forms with integer coefficients.

Claim 5. *The linear form Λ_1 is linearly independent of each of the forms Λ_j for $j = 2,\dots,s$ and of each of the forms Λ'_j for $j = 1,\dots,s$.*

Proof of Claim 5. Suppose that for some non-zero $a, b \in \mathbb{Z}$, some $j \in \{1,\dots,s\}$, and some $\alpha' \in \{\alpha,\beta\}$ with $(1,\alpha) \neq (j,\alpha')$ we have that $a\,\mathrm{Im}(\zeta_1\alpha z) = b\,\mathrm{Im}(\zeta_j\alpha' z)$ for every $z \in \mathbb{Z}[\tau_d]$. Using this relation with $z := 1$ and $z := \tau_d$ we get that

$$a\,\zeta_1\,\alpha = b\,\zeta_j\,\alpha'. \tag{9.25}$$

We consider the following three cases:

(i) Suppose that $j = 1$ and $\alpha' = \beta$. In this case, Equation (9.25) gives that α/β is a rational which is impossible because α and β are distinct non-integer prime elements of $\mathbb{Z}[\tau_d]$ (recall that no two elements of $\mathcal{I}_d$ are associates).

(ii) Suppose that $j > 1$ and $\alpha' = \alpha$. In this case we have that $\zeta_1/\zeta_j \in \mathbb{Q}$, and by (9.19) the linear forms ϕ_1 and ϕ_j are linearly dependent, contradicting the hypothesis.

(iii) It remains to consider the case where $j > 1$ and $\alpha' = \beta$. Let r_a be the exponent of (α) in the factorization of the ideal (a) into prime ideals of $\mathbb{Z}[\tau_d]$. Since α is a non-integer prime element of $\mathbb{Z}[\tau_d]$, $\overline{\alpha}$ is also a non-integer prime element and, since a is real, the exponent of $(\overline{\alpha})$ in the factorization of (a) is also equal to r_a. Since by hypothesis α and $\overline{\alpha}$ are non-associate prime elements, it follows that a can be written as $a = \alpha^{r_a}\overline{\alpha}^{r_a}c$ for some $c \in \mathbb{Z}[\tau_d]$ not divisible by α or $\overline{\alpha}$. In the same way, $b = \alpha^{r_b}\overline{\alpha}^{r_b}d$ for some non-negative integer r_b and some $d \in \mathbb{Z}[\tau_d]$ not divisible by α or $\overline{\alpha}$. Equation (9.25) gives

$$\alpha^{r_a+1}\,\overline{\alpha}^{r_a}\,c\,\zeta_1 = \alpha^{r_b}\,\overline{\alpha}^{r_b}\,d\,\zeta_j\,\beta.$$

By hypothesis, α and $\overline{\alpha}$ are non-associate prime elements, β is a prime element non-associate to α, and d and ζ_j are not divisible by α. It follows that $r_b \ge r_a + 1$. Similarly, c and ζ_1 are not divisible by $\overline{\alpha}$, and it follows that $r_a \ge r_b$, a contradiction. This completes the proof of Claim 5. ∎

9.7.4. *End of the proof.* In order to prove Claim 4 we return to the coordinates (m,n) of a point of $\mathbb{Z}^2$ which is identified with $\mathbb{Z}[\tau_d]$. Denoting the linear forms defined above by $\Lambda_j(m,n)$ and $\Lambda'_j(m,n)$, it remains to show that

$$\lim_{N\to+\infty} \mathbb{E}_{(m,n)\in B_N}\, \mathbf{1}_{K_N^*}(m,n) \prod_{j=1}^{s} f_j(\Lambda_j(m,n))\, \overline{f}_j(\Lambda'_j(m,n)) = 0,$$

where the convex sets $K_N^* \subset [-N,N]^2$ were defined in (9.24). By (9.6), it suffices to show that

$$\lim_{N\to+\infty} \mathbb{E}_{-N\le m,n\le N}\, \mathbf{1}_{K_N^*}(m,n) \prod_{j=1}^{s} f_j(\Lambda_j(m,n))\, \overline{f}_j(\Lambda'_j(m,n)) = 0. \tag{9.26}$$

It follows from Claim 5 that the form Λ_1 is linearly independent of each of the forms Λ_j for $j = 2,\dots,s$ and of each of the forms Λ'_j for $j = 1,\dots,s$. By hypothesis, the function f_1 is an even aperiodic multiplicative function, and thus $\|f_1\|_{U^{2s-1}[N]} \to 0$ as $N \to +\infty$ by Theorem 2.5. All the hypotheses of Lemma 9.6 are satisfied and (9.26) follows. This completes the proof of Theorem 9.7 and hence of Theorem 2.6. □

10. Partition regularity results

The goal of this section is to prove Theorems 2.12 (which implies Theorem 2.7) and 2.13. For notational convenience we prove Theorem 2.12 and indicate at the end of this section the modifications needed to prove the more general Theorem 2.13.

Recall that our goal is to show that given admissible integers $\ell_0,\dots,\ell_4$ (see Section 2.4.2), on every partition of $\mathbb{N}$ into finitely many cells there exist $k,m,n \in \mathbb{Z}$ such that the integers

$$x := k\ell_0(m+\ell_1 n)(m+\ell_2 n), \qquad y := k\ell_0(m+\ell_3 n)(m+\ell_4 n), \tag{10.1}$$

are positive, distinct, and belong to the same cell. We start with some successive reformulations of the problem that culminate in the analytic statement of Proposition 10.5. We then prove this result using the structural result of Theorem 2.1.

10.1. Reduction to a density regularity result.

We first recast Theorem 2.12 as a density regularity statement for dilation invariant densities on the integers.

We write $\mathbb{Q}^+$ for the multiplicative group of positive rationals. Let $p_1, p_2, \dots,$ be the sequence of primes. Then the sequence $(\Phi_M)_{M\in\mathbb{N}}$ of finite subsets of $\mathbb{N}$ defined by

$$\Phi_M := \big\{n \colon n \,|\, (p_1p_2\dots p_M)^M\big\} = \{p_1^{k_1}\cdots p_M^{k_M} : 0 \le k_1,\dots,k_M \le M\}$$

is a *multiplicative Følner sequence*. This means that, for every $r \in \mathbb{Q}^+$, we have

$$\lim_{M\to+\infty} \frac{1}{|\Phi_M|} |r^{-1}\Phi_M \triangle \Phi_M| = 0, \tag{10.2}$$

where for every subset A of $\mathbb{N}$ and for every $r \in \mathbb{Q}^+$, we write

$$r^{-1}A := \{x \in \mathbb{N} \colon rx \in A\} = \{r^{-1}y \colon y \in A\} \cap \mathbb{N}.$$

To this multiplicative Følner sequence we associate a notion of multiplicative density as follows.

Definition (Multiplicative density)**.** The *(upper) multiplicative density* $d_{\text{mult}}(E)$ of a subset E of $\mathbb{N}$ is defined as

$$d_{\text{mult}}(E) := \limsup_{M\to+\infty} \frac{|E\cap\Phi_M|}{|\Phi_M|}.$$

We remark that the multiplicative density and the additive density are non-comparable measures of largeness. For instance, the set of odd numbers has zero multiplicative density with respect to any multiplicative Følner sequence, as has any set that omits all multiples of some positive integer. On the other hand, it is not hard to construct sets with multiplicative density 1 that have additive density 0 (see, for instance, [2]).

An important property of the multiplicative density, and the reason we work with this notion of largeness, is its invariance under dilations. Indeed, for every $E\subset\mathbb{N}$ and every $r\in\mathbb{Q}^+$, it follows from (10.2) that

$$d_{\text{mult}}(E) = d_{\text{mult}}(r^{-1}E).$$

Since any multiplicative density is clearly subadditive, any finite partition of $\mathbb{N}$ has at least one cell with positive multiplicative density. Hence, Theorem 2.12 follows from the following stronger result.

Theorem 10.1 (Density regularity)**.** *Let $\ell_0,\dots,\ell_4\in\mathbb{Z}$ be admissible. Then every set $E\subset\mathbb{N}$ with $d_{\text{mult}}(E)>0$ contains distinct $x,y\in\mathbb{N}$ of the form* (10.1)*.*

Since

$$d_{\text{mult}}(x^{-1}E\cap y^{-1}E) = d_{\text{mult}}\big(\{k\in\mathbb{N}\colon \{kx,ky\}\subset E\}\big),$$

in order to prove Theorem 10.1 it suffices to prove the following result.

Proposition 10.2. *Let $\ell_0,\dots,\ell_4\in\mathbb{Z}$ be admissible. Then every $E\subset\mathbb{N}$ with $d_{\text{mult}}(E)>0$ contains distinct $x,y\in\mathbb{N}$ of the form* (10.1) *such that*

$$d_{\text{mult}}(x^{-1}E\cap y^{-1}E)>0. \tag{10.3}$$

In the next two sections we are going to reinterpret Proposition 10.2 as a more convenient to prove analytic statement.

10.2. **Integral formulation.** We first reformulate Proposition 10.2 using an integral representation result of positive definite sequences on $\mathbb{Q}^+$. Recall that a function $f\colon\mathbb{N}\to\mathbb{C}$ is completely multiplicative if $f(xy)=f(x)f(y)$ for every $x,y\in\mathbb{N}$.

Definition. We denote by $\mathcal{M}_1^c$ the set of completely multiplicative functions of modulus exactly 1.

A completely multiplicative function is uniquely determined by its values on the primes. Every $f\in\mathcal{M}_1^c$ can be extended to a multiplicative function on $\mathbb{Q}^+$, also denoted by f, by letting for every $x,y\in\mathbb{N}$

$$f(xy^{-1}) := f(x)\overline{f}(y). \tag{10.4}$$

Endowed with the pointwise multiplication and the topology of pointwise convergence, the family $\mathcal{M}_1^c$ of completely multiplicative functions is a compact (metrizable) Abelian group, with unit element the constant function $\mathbf{1}$. This group is the dual group of $\mathbb{Q}^+$, the duality being given by (10.4).

Let $E \subset \mathbb{N}$ be a set with $d_{\text{mult}}(E) > 0$. There exists a sequence (M_j) of integers, tending to infinity, such that

$$\lim_{j\to+\infty} \frac{|E \cap \Phi_{M_j}|}{|\Phi_{M_j}|} = d_{\text{mult}}(E); \tag{10.5}$$

$$\text{and} \quad \rho(r) := \lim_{j\to+\infty} \frac{|E \cap (r^{-1}E) \cap \Phi_{M_j}|}{|\Phi_{M_j}|} \quad \text{exists for every} \quad r \in \mathbb{Q}^+. \tag{10.6}$$

Then the function $\rho\colon \mathbb{Q}^+ \to \mathbb{C}$ is positive definite; that is, for every $n \in \mathbb{N}$, all $r_1, \dots, r_n \in \mathbb{Q}^+$, and all $\lambda_1, \dots, \lambda_n \in \mathbb{C}$, we have

$$\sum_{i,j=1}^{n} \lambda_i \overline{\lambda_j}\, \rho(r_i\, r_j^{-1}) \geq 0.$$

By Bochner's theorem, there exists a unique positive finite measure ν on the compact Abelian group $\mathcal{M}_1^c$ with a Fourier-Stieltjes transform $\widehat{\nu}$ equal to the function ρ. This means that

$$\text{for every } r \in \mathbb{Q}^+, \ \int_{\mathcal{M}_1^c} f(r)\, d\nu(f) = \widehat{\nu}(r) = \rho(r) = \lim_{j\to+\infty} \frac{|E \cap (r^{-1}E) \cap \Phi_{M_j}|}{|\Phi_{M_j}|}. \tag{10.7}$$

We collect the properties of the measure ν used in the sequel.

Claim. Let the set E and the measure ν be as before and $\delta := d_{\text{mult}}(E)$. Then

$$\int_{\mathcal{M}_1^c} f(x)\, \overline{f}(y)\, d\nu(f) \geq 0 \quad \text{for every } x, y \in \mathbb{N}; \tag{10.8}$$

$$\nu(\{\mathbf{1}\}) \geq \delta^2. \tag{10.9}$$

Proof of the Claim. Property (10.8) follows from (10.7) with $r := xy^{-1}$. The proof of (10.9) is classical, but we give it for completeness.

For $x \in \mathbb{N}$, let $f(x) := \mathbf{1}_E(x) - \delta$. The averages on Φ_{M_j} of the function f tend to 0 as $j \to +\infty$, and it follows from (10.2) that

$$\text{for every } r \in \mathbb{Q}^+, \ \lim_{j\to+\infty} \mathbb{E}_{x\in\Phi_{M_j}} f(rx) = 0.$$

Thus,

$$\widehat{\nu}(r) = \lim_{j\to+\infty} \frac{|E \cap r^{-1}E \cap \Phi_{M_j}|}{|\Phi_{M_j}|} = \lim_{j\to+\infty} \mathbb{E}_{x\in\Phi_{M_j}} \mathbf{1}_E(x)\mathbf{1}_E(rx) = \delta^2 + \psi(r),$$

where

$$\psi(r) := \lim_{j\to+\infty} \mathbb{E}_{x\in\Phi_{M_j}} f(x) f(rx)$$

and the limit exists for every $r \in \mathbb{Q}^+$ by (10.6). The function $\psi\colon \mathbb{Q}^+ \to \mathbb{C}$ is positive definite and by Bochner's theorem $\psi = \widehat{\sigma}$ for some positive finite measure σ on $\mathcal{M}_1^c$. Since $\mathbf{1}$ is the unit element of the group $\mathcal{M}_1^c$ and (Φ_N) is a Følner sequence in $\mathbb{Q}^+$, the averages of $\widehat{\nu}(r)$ on (Φ_N) converge to $\nu(\{\mathbf{1}\})$ and the averages of $\widehat{\sigma}(r)$ to $\sigma(\{\mathbf{1}\})$. Therefore

$$\nu(\{\mathbf{1}\}) = \lim_{M\to+\infty} \mathbb{E}_{r\in\Phi_M} \widehat{\nu}(r) = \delta^2 + \lim_{M\to+\infty} \mathbb{E}_{r\in\Phi_M} \widehat{\sigma}(r) = \delta^2 + \sigma(\{\mathbf{1}\}) \geq \delta^2. \quad \square$$

In order to show Proposition 10.2 it suffices to prove the following.

Proposition 10.3 (Analytic formulation)**.** *Let $\ell_1,\dots,\ell_4$ be distinct integers and suppose that $\min\{\ell_1,\dots,\ell_4\}=0$. Let ν be a probability measure on $\mathcal{M}_1^c$ that satisfies Properties* (10.8) *and* (10.9)*. Then there exist $m,n\in\mathbb{Z}$ such that $(m+\ell_1n)(m+\ell_2n)$ and $(m+\ell_3n)(m+\ell_4n)$ are positive, distinct integers, and we have*

$$\int_{\mathcal{M}_1^c} f(m+\ell_1n)\cdot f(m+\ell_2n)\cdot \overline{f}(m+\ell_3n)\cdot\overline{f}(m+\ell_4n)\,d\nu(f)>0. \tag{10.10}$$

We show that Proposition 10.3 implies Proposition 10.2. Without loss of generality we can assume that the measure ν defined in (10.7) is a probability measure. Let $m,n\in\mathbb{Z}$ satisfy (10.10). Letting $x:=(m+\ell_1n)(m+\ell_2n)$ and $y:=(x+\ell_3n)(x+\ell_4n)$ and using (10.7) we get

$$\begin{aligned} d_{\text{mult}}(x^{-1}E\cap y^{-1}E)&=\limsup_{M\to+\infty}\frac{|x^{-1}yE\cap E\cap\Phi_M|}{|\Phi_M|}\geq\lim_{j\to+\infty}\frac{|x^{-1}yE\cap E\cap\Phi_{M_j}|}{|\Phi_{M_j}|}\\ &=\widehat{\nu}(xy^{-1})=\int_{\mathcal{M}_1^c}f(xy^{-1})\,d\nu(f)=\int_{\mathcal{M}_1^c}f(x)\,\overline{f}(y)\,d\nu(f)>0.\end{aligned}$$

This proves Proposition 10.2 in the case where the integers $\ell_1,\dots,\ell_4$ are distinct and $\min\{\ell_1,\dots,\ell_4\}=0$.

Let $\underline{\ell}:=\min\{\ell_1,\dots,\ell_4\}\neq 0$, by replacing ℓ_j with $\ell_j-\underline{\ell}$ for $j=1,\dots,4$, and making the change of variables $m\mapsto m-\underline{\ell}n$, we reduce matters to the case that $\underline{\ell}=0$.

It remains to consider the degenerate cases where ℓ_1 or ℓ_2 is equal to ℓ_3 or ℓ_4. Suppose that $\ell_1=\ell_3$, and the other cases are similar. Since $\ell_0,\dots,\ell_4$ are admissible, $\ell_2\neq\ell_4$. We can assume that $\ell_2<\ell_4$, and the other case is similar. As before, we see that it suffices to show that there exist $m,n\in\mathbb{Z}$ such that the integers $m+\ell_2n$ and $m+\ell_4n$ are positive and satisfy

$$\int_{\mathcal{M}_1^c} f(m+\ell_2n)\cdot\overline{f}(m+\ell_4n)\,d\nu(f)>0.$$

After making the change of variables $m\mapsto m-\ell_2n$ we see that it suffices to show that there exist $m,n\in\mathbb{N}$ such that

$$\int_{\mathcal{M}_1^c} f(m)\cdot\overline{f}(m+(\ell_4-\ell_2)n)\,d\nu(f)>0.$$

Since the averages of $\widehat{\nu}$ on the Følner sequence (Φ_M) converge as $M\to+\infty$ to $\nu(\{\mathbf{1}\})$ which is positive by (10.9), there exists $n_0\in\mathbb{N}$ such that $\widehat{\nu}(n_0+1)>0$. Taking $m:=\ell_4-\ell_2$ and $n:=n_0$ we have

$$\int_{\mathcal{M}_1^c} f(m)\cdot\overline{f}(m+(\ell_4-\ell_2)n)\,d\nu(f)=\widehat{\nu}(n_0+1)>0.$$

Convention. In the rest of the proof we assume that $\ell_1,\dots,\ell_4$ are distinct integers with $\min\{\ell_1,\dots,\ell_4\}=0$. We let

$$\ell:=\ell_1+\ell_2+\ell_3+\ell_4.$$

For every $N\in\mathbb{N}$, we denote by $\widetilde{N}$ the smallest prime in the interval $[2\ell N,4\ell N]$. As usual, for every function ϕ on $\mathbb{N}$, we denote by ϕ_N the function $\mathbf{1}_{[N]}\,\phi$, considered as a function on $\mathbb{Z}_{\widetilde{N}}$.

10.3. **Final analytic formulation.** In order to prove Proposition 10.3, it suffices to establish the stronger fact that there are "many" $m, n \in \mathbb{N}$ such that the integral in this statement is positive.

Proposition 10.4 (Averaged analytic formulation)**.** *Let* $\ell_1, \ldots, \ell_4 \in \mathbb{Z}$ *be distinct with* $\min\{\ell_1, \ldots, \ell_4\} = 0$*, and let* ν *be a probability measure on* $\mathcal{M}_1^c$ *that satisfies Properties* (10.8) *and* (10.9)*. Then*
(10.11)

$$\liminf_{N\to+\infty} \int_{\mathcal{M}_1^c} \mathbb{E}_{(m,n)\in\Theta_N} f(m+\ell_1 n)\cdot f(m+\ell_2 n)\cdot \overline{f}(m+\ell_3 n)\cdot \overline{f}(m+\ell_4 n)\, d\nu(f) > 0,$$

where $\Theta_N := \{(m,n) \in [N]\times[N] \colon 1 \le m+\ell_i n \le N \text{ for } i = 1,2,3,4\}$.

In order to show that Proposition 10.4 implies Proposition 10.3 we remark that for $N \in \mathbb{N}$ sufficiently large we have $|\Theta_N| \ge c_1 N^2$, and the cardinality of the set of pairs $(m,n) \in \Theta_N$ that satisfy $(m+\ell_1 n)(m+\ell_2 n) = (m+\ell_3 n)(m+\ell_4 n)$ is bounded by $C_1 N$ for some constants c_1 and C_1 that depend only on ℓ. Therefore, Property (10.11) implies that there exist $m, n \in \mathbb{N}$ such that $(m+\ell_1 n)(m+\ell_2 n) \neq (m+\ell_3 n)(m+\ell_4 n)$ and $\int_{\mathcal{M}_1^c} f(m+\ell_1 n) f(m+\ell_2 n)\, \overline{f}(m+\ell_3 n)\, \overline{f}(m+\ell_4 n)\, d\nu(f) > 0$. Hence, the conclusion of Proposition 10.3 holds.

Remark. An alternate (and arguably more natural) way to proceed is to replace the additive averages in Proposition 10.4 with multiplicative ones. Upon doing this, one is required to analyze averages of the form

$$\mathbb{E}_{m,n\in\Psi_N} f\big((m+\ell_1 n)(m+\ell_2 n)\big)\, \overline{f}\big((m+\ell_3 n)(m+\ell_4 n)\big),$$

where $(\Psi_N)_{N\in\mathbb{N}}$ is a multiplicative Følner sequence in $\mathbb{N}$ and $f \in \mathcal{M}_1^c$. Unfortunately, we were not able to prove anything useful for these multiplicative averages, although one suspects that a positivity property similar to the one in (10.11) may hold.

Next, for technical reasons, we recast the previous proposition as a positivity property involving averages over the cyclic groups $\mathbb{Z}_{\widetilde{N}}$. This is going to be the final form of the analytic statement that we aim to prove.

Proposition 10.5 (Final analytic formulation)**.** *Let* $\ell_1, \ldots, \ell_4 \in \mathbb{Z}$ *be distinct and suppose that* $\min\{\ell_1, \ldots, \ell_4\} = 0$*. Let* $\delta > 0$ *and* ν *be a probability measure on* $\mathcal{M}_1^c$*, such that*

(i) $\nu(\{\mathbf{1}\}) \ge \delta^2$*;*
(ii) $\int_{\mathcal{M}_1^c} f(x)\, \overline{f}(y)\, d\nu(f) \ge 0$ *for every* $x, y \in \mathbb{N}$.

Then we have

$$\liminf_{N\to+\infty} \int_{\mathcal{M}_1^c} \mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}} \mathbf{1}_{[N]}(n) f_N(m+\ell_1 n) f_N(m+\ell_2 n) \overline{f}_N(m+\ell_3 n) \times \overline{f}_N(m+\ell_4 n)\, d\nu(f) > 0, \tag{10.12}$$

where in the above average the expressions $m+\ell_i n$ *can be considered as elements of* $\mathbb{Z}$ *or* $\mathbb{Z}_{\widetilde{N}}$ *without affecting the value of the average.*

We verify that Proposition 10.5 implies Proposition 10.4. Using the definition of the set Θ_N given in Proposition 10.4, we can rewrite the averages that appear in the statement of Proposition 10.4 as follows:

$$\begin{aligned}(10.13)\quad &\mathbb{E}_{(m,n)\in\Theta_N} f(m+\ell_1 n)\cdot f(m+\ell_2 n)\cdot \overline{f}(m+\ell_3 n)\cdot \overline{f}(m+\ell_4 n)\\ &= \frac{\widetilde{N}^2}{|\Theta_N|}\,\mathbb{E}_{m,n\in[\widetilde{N}]}\mathbf{1}_{[N]}(n)\cdot f_N(m+\ell_1 n)\cdot f_N(m+\ell_2 n)\cdot \overline{f}_N(m+\ell_3 n)\cdot \overline{f}_N(m+\ell_4 n).\end{aligned}$$

To prove this equality, we remark that since $\min\{\ell_1,\dots,\ell_4\}=0$, if m,n are such that $m\in[\widetilde{N}]$, $n\in[N]$, and $m+\ell_j n \bmod \widetilde{N}\in[N]$ for $j=1,\dots,4$, then $m\in[N]$. Thus $1\leq m+\ell_j n\leq(\ell+1)N<\widetilde{N}$; hence $m+\ell_j n=m+\ell_j n \bmod \widetilde{N}\in[N]$ for $j=1,2,3,4$ and every $(m,n)\in\Theta_N$. The sets of pairs (m,n) taken into account in the two averages are identical, and the value of the last expression remains unchanged if we replace each term $m+\ell_i n$ by $m+\ell_i n \bmod \widetilde{N}$.

Using identity (10.13) and the estimate $cN^2\leq|\Theta_N|\leq N^2$ which holds for some positive constant c that depends only on ℓ, we get the asserted implication.

10.4. **A positivity property.**

We derive now a positivity property that will be used in the proof of Proposition 10.5 in the next subsection. Here we make essential use of the positivity Property (ii) of the measure ν given in Proposition 10.5.

Lemma 10.6 (Hidden non-negativity). *Let ν be a positive finite measure on $\mathcal{M}_1^c$ that satisfies Property* (ii) *of Proposition* 10.5. *Let ψ be a non-negative function defined on $\mathbb{Z}_{\widetilde{N}}$. Then*

$$\int_{\mathcal{M}_1^c}(f_N*\psi)(n_1)\cdot(f_N*\psi)(n_2)\cdot(\overline{f}_N*\psi)(n_3)\cdot(\overline{f}_N*\psi)(n_4)\;d\nu(f)\geq 0$$

for every $n_1,n_2,n_3,n_4\in\mathbb{Z}_{\widetilde{N}}$, where the convolution product is taken on $\mathbb{Z}_{\widetilde{N}}$.

Proof. The convolution product $f_N*\psi$ is defined on the group $\mathbb{Z}_{\widetilde{N}}$ by the formula

$$(f_N*\psi)(n)=\mathbb{E}_{k\in\mathbb{Z}_{\widetilde{N}}}\psi(n-k)\cdot f_N(k).$$

It follows that for every $n\in[\widetilde{N}]$ there exists a sequence $(a_n(k))_{k\in\mathbb{Z}_{\widetilde{N}}}$ of non-negative numbers that are independent of f, such that for every $f\in\mathcal{M}_1^c$ we have

$$(f_N*\psi)(n)=\sum_{k\in\mathbb{Z}_{\widetilde{N}}}a_n(k)\,f(k).$$

The left hand side of the expression in the statement is thus equal to

$$\sum_{k_1,k_2,k_3,k_4\in\mathbb{Z}_{\widetilde{N}}}\prod_{i=1}^4 a_{n_i}(k_i)\int_{\mathcal{M}_1^c}f(k_1)\cdot f(k_2)\cdot\overline{f}(k_3)\cdot\overline{f}(k_4)\,d\nu(f)=$$

$$\sum_{k_1,k_2,k_3,k_4\in\mathbb{Z}_{\widetilde{N}}}\prod_{i=1}^4 a_{n_i}(k_i)\int_{\mathcal{M}_1^c}f(k_1k_2)\,\overline{f}(k_3k_4)\,d\nu(f)\geq 0$$

by Property (ii) of Proposition 10.5. □

10.5. **Estimates involving Gowers norms.** Next we establish an elementary estimate that will be crucial in the sequel.

Lemma 10.7 (Uniformity estimates). *Let $s \geq 3$, $\ell_1, \ldots, \ell_s \in \mathbb{Z}$ be distinct, and let $\ell := |\ell_1| + \cdots + |\ell_s|$. Then there exists $C := C(\ell)$ such that for every $N \in \mathbb{N}$ and all functions $a_j \colon \mathbb{Z}_{\widetilde{N}} \to \mathbb{C}$, $j = 1, \ldots, s$, with $|a_j| \leq 1$, we have*

$$\big|\mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}} \mathbf{1}_{[N]}(n) \cdot \prod_{j=1}^{s} a_j(m + \ell_j n)\big| \leq C \min_{j=1,\ldots,s} (\|a_j\|_{U^{s-1}(\mathbb{Z}_{\widetilde{N}})})^{1/2} + \frac{2}{\widetilde{N}},$$

where $\widetilde{N}$ is the smallest prime that is greater than $2\ell N$.

Proof. We first reduce matters to estimating a similar average that does not contain the term $\mathbf{1}_{[N]}(n)$. Let r be an integer that will be specified later and satisfies $0 < r < N/2$. We define the "trapezoid function" ϕ on $\mathbb{Z}_N$ so that $\phi(0) = 0$, ϕ increases linearly from 0 to 1 on the interval $[0, r]$, $\phi(n) = 1$ for $r \leq n \leq N - r$, ϕ decreases linearly from 1 to 0 on $[N - r, N]$, and $\phi(n) = 0$ for $N < n < \widetilde{N}$.

The absolute value of the difference between the average in the statement and

$$\mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}} \phi(n) \cdot \prod_{j=1}^{s} a_j(m + \ell_j n)$$

is bounded by $2r/\widetilde{N}$. Moreover, it is classical that (the argument is sketched in the proof of Lemma A.1 in the Appendix)

$$\|\widehat{\phi}\|_{l^1(\mathbb{Z}_{\widetilde{N}})} \leq \frac{2N}{r} \leq \frac{\widetilde{N}}{r},$$

and thus

$$\Big|\mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}} \phi(n) \cdot \prod_{j=1}^{s} a_j(m + \ell_j n)\Big| \leq \frac{\widetilde{N}}{r} \max_{\xi\in\mathbb{Z}_{\widetilde{N}}} \Big|\mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}} \mathrm{e}(n\xi/\widetilde{N}) \cdot \prod_{j=1}^{s} a_j(m + \ell_j n)\Big|.$$

Since $\ell_1 \neq \ell_2$ and $\widetilde{N} > \ell$, we have $\ell_1 - \ell_2 \neq 0 \bmod \widetilde{N}$, and there exist $\ell^* \in \mathbb{Z}_{\widetilde{N}}$ such that $\ell^*(\ell_1 - \ell_2) = 1 \bmod \widetilde{N}$. Upon replacing $a_1(n)$ with $a_1(n)\mathrm{e}(-\ell^* n\xi/\widetilde{N})$ and $a_2(n)$ with $a_2(n)\mathrm{e}(\ell^* n\xi/\widetilde{N})$, the U^{s-1}-norm of all sequences remains unchanged, and the term $\mathrm{e}(n\xi/\widetilde{N})$ disappears. We are thus left with estimating the average

$$\big|\mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}} \prod_{j=1}^{s} a_j(m + \ell_j n)\big|.$$

Since $\widetilde{N} > 2\ell$, the numbers $\ell_1, \ldots, \ell_s$ are distinct as elements of $\mathbb{Z}_{\widetilde{N}}$. Using this and the fact that $\widetilde{N}$ is a prime, it is possible to show by an iterative use of the Cauchy-Schwarz inequality (see, for example, [72, Theorem 3.1]) that the last average is bounded by

$$U := \min_{1\leq j\leq s} \|a_j\|_{U^{s-1}(\mathbb{Z}_{\widetilde{N}})}.$$

Combining the preceding estimates, we get that the average in the statement is bounded by

$$\frac{2r}{\widetilde{N}} + \frac{2\widetilde{N}}{r} U.$$

Assuming that $U \neq 0$ and choosing $r := \lfloor \sqrt{U}\widetilde{N}/(8\ell) \rfloor + 1$ (then $r \leq \widetilde{N}/(8\ell) \leq N/2$) gives the announced bound. ∎

10.6. **Proof of Proposition 10.5.** We start with a brief sketch of our proof strategy. Roughly speaking, Theorem 2.1 enables us to decompose the restriction of an arbitrary multiplicative function on a finite interval into three terms, a close to periodic term, a "very uniform" term, and an error term. In the course of the proof of Proposition 10.5 we study these three terms separately. The order of the different steps is important as well as the precise properties of the decomposition.

First, we show that the uniform term has a negligible contribution in evaluating the averages in (10.12). To do this we use the uniformity estimates established in Proposition 9.6. It is for this part of the proof that it is very important to work with patterns that factor into products of linear forms in two variables; otherwise, we have no way of controlling the corresponding averages by Gowers uniformity norms. At this point, the error term is shown to have a negligible contribution, and thus can be ignored. Last, the structured term f_{st} is dealt with by restricting the variable n to a suitable sub-progression where each function f_{st} gives approximately the same value to all four linear forms[13]; it then becomes possible to establish the asserted positivity. In fact, the step where we restrict to a sub-progression is rather delicate, as it has to take place before the component f_{er} is eliminated (this also explains why we do not restrict both variables m and n to a sub-progression), and in addition one has to guarantee that the terms left out are non-negative, a property that follows from Lemma 10.6.

We now enter the main body of the proof. Recall that $\ell_1,\dots,\ell_4\in\mathbb{Z}$ are fixed and distinct and that $\ell=|\ell_1|+\cdots+|\ell_4|$. We stress also that in this proof the quantities $m+\ell_i n$ are computed in $\mathbb{Z}_{\widetilde{N}}$, that is, modulo $\widetilde{N}$.

Let ν be a positive finite measure on $\mathcal{M}_1^c$ satisfying Properties (i) and (ii) of Proposition 10.5, and let $\delta>0$ be as in (i). We let

$$\varepsilon:=c_1\delta^2 \qquad\text{and}\qquad F(x,y,z):=C_1^2\,\frac{x^2y^2}{z^4},$$

where c_1 and C_1 are positive constants that will be specified later; what is important is that they depend only on ℓ. Our goal is for all large values of $N\in\mathbb{N}$ (how large will depend only on δ) to bound from below the average

$$A(N):=\int_{\mathcal{M}_1^c}\mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}}\mathbf{1}_{[N]}(n)$$
$$\times\, f_N(m+\ell_1 n)\,f_N(m+\ell_2 n)\,\overline{f}_N(m+\ell_3 n)\,\overline{f}_N(m+\ell_4 n)\;d\nu(f).$$

We start by applying Theorem 2.1 for the U^3-norms, taking as input the measure ν, the number ε, and the function F defined above. Let

$$Q:=Q(F,N,\varepsilon,\nu)=Q(N,\delta,\nu),\quad R:=R(F,N,\varepsilon,\nu)=R(N,\delta,\nu)$$

be the numbers provided by Theorem 2.1. We recall that Q and R are bounded by a constant that depends only on δ. From this point on we assume that $N\in\mathbb{N}$ is sufficiently large, depending only on δ, so that the conclusions of Theorem 2.1 hold. For $f\in\mathcal{M}_1^c$, we have a decomposition

$$f_N(n)=f_{N,\mathrm{st}}(n)+f_{N,\mathrm{un}}(n)+f_{N,\mathrm{er}}(n),\quad n\in\mathbb{Z}_{\widetilde{N}},$$

for the decomposition that satisfies Properties (i)–(v) of Theorem 2.1.

[13]This coincidence of values is very important; not having it is a key technical obstruction for handling equations like $x^2+y^2=n^2$. Restricting the range of both variables m and n does not seem to help either, as this creates problems with controlling the error term in the decomposition.

Next, we use the uniformity estimates of Lemma 10.7 for $s = 4$ in order to eliminate the uniform component f_{un} from the average $A(N)$. We let

$$f_{s,e} := f_{N,\text{st}} + f_{N,\text{er}}$$

and

$$A_1(N) := \int_{\mathcal{M}_1^c} \mathbb{E}_{m,n\in\mathbb{Z}_{\widetilde{N}}} \mathbf{1}_{[N]}(n) f_{s,e}(m+\ell_1 n) f_{s,e}(m+\ell_2 n) \overline{f}_{s,e}(m+\ell_3 n) \overline{f}_{s,e}(m+\ell_4 n)\, d\nu(f).$$

Using Lemma 10.7, Property (iv) of Theorem 2.1, and the estimates $|f_N(n)| \leq 1$, $|f_{s,e}(n)| \leq 1$ for every $n \in \mathbb{Z}_{\widetilde{N}}$, we get that

$$|A(N) - A_1(N)| \leq \frac{4\,C_2}{F(Q,R,\varepsilon)^{\frac{1}{2}}} + \frac{8}{\widetilde{N}}, \tag{10.14}$$

where C_2 is the constant provided by Lemma 10.7 and depends only on ℓ.

Next, we eliminate the error term f_{er}. But before doing this, it is important to first restrict the range of n to a suitable sub-progression; the utility of this maneuver will be clear on our next step when we estimate the contribution of the leftover term f_{st}. We stress that we cannot postpone this restriction on the range of n until after the term f_{er} is eliminated; if we did this, the contribution of the term f_{er} would swamp the positive lower bound we get from the term f_{st}. We let

$$\eta := \frac{\varepsilon}{QR}. \tag{10.15}$$

By Property (i) of Theorem 2.1 and Lemma 10.6, we have the positivity property

$$\int_{\mathcal{M}_1^c} f_{s,e}(n_1) \cdot f_{s,e}(n_2) \cdot \overline{f}_{s,e}(n_3) \cdot \overline{f}_{s,e}(n_4)\, d\nu(f) \geq 0 \tag{10.16}$$

for every $n_1, n_2, n_3, n_4 \in \mathbb{Z}_{\widetilde{N}}$,

Note that the integers Qk, $1 \leq k \leq \eta N$, are distinct elements of the interval $[N]$. It follows from (10.16) that

$$\sum_{m,n\in\mathbb{Z}_{\widetilde{N}}} \int_{\mathcal{M}_1^c} \mathbf{1}_{[N]}(n)\, f_{s,e}(m)\, f_{s,e}(m+\ell_1 n)\, \overline{f}_{s,e}(m+\ell_2 n)\, \overline{f}_{s,e}(m+\ell_3 n)\, d\nu(f)$$

$$\geq \sum_{m\in\mathbb{Z}_{\widetilde{N}}} \sum_{k=1}^{\lfloor \eta N \rfloor} \int_{\mathcal{M}_1^c} f_{s,e}(m+\ell_1 Qk)\, f_{s,e}(m+\ell_2 Qk)\, \overline{f}_{s,e}(m+\ell_3 Qk)\, \overline{f}_{s,e}(m+\ell_4 Qk)\, d\nu(f).$$

Therefore, we have

$$A_1(N) \geq \frac{\lfloor \eta N \rfloor}{\widetilde{N}} A_2(N) \geq \frac{\eta}{40\,\ell} A_2(N) = \varepsilon\, \frac{1}{40\,\ell QR}\, A_2(N), \tag{10.17}$$

where

$$A_2(N) := \int_{\mathcal{M}} \mathbb{E}_{m\in\mathbb{Z}_{\widetilde{N}}}\, \mathbb{E}_{k\in[\lfloor \eta N \rfloor]}$$
$$\times\, f_{s,e}(m+\ell_1 Qk)\, f_{s,e}(m+\ell_2 Qk)\, \overline{f}_{s,e}(m+\ell_3 Qk)\, \overline{f}_{s,e}(m+\ell_4 Qk)\, d\nu(f).$$

We let

$$A_3(N) := \int_{\mathcal{M}_1^c} \mathbb{E}_{m\in\mathbb{Z}_{\widetilde{N}}}\mathbb{E}_{k\in[\lfloor \eta N\rfloor]}$$
$$\times f_{N,\mathrm{st}}(m+\ell_1Qk)f_{N,\mathrm{st}}(m+\ell_2Qk)\overline{f}_{N,\mathrm{st}}(m+\ell_3Qk)\overline{f}_{N,\mathrm{st}}(m+\ell_4Qk)d\nu(f).$$

Since for every $n\in\mathbb{Z}_{\widetilde{N}}$ we have $|f_{N,\mathrm{st}}(n)|\leq 1$, and since $|f_{s,e}(n)| = |f_{N,\mathrm{st}}(n)+f_{N,\mathrm{er}}(n)|\leq 1$, by Property (v) of Theorem 2.1 we deduce that

$$|A_2(N)-A_3(N)| \leq 4\int_{\mathcal{M}_1^c} \mathbb{E}_{m\in\mathbb{Z}_{\widetilde{N}}}|f_{N,\mathrm{er}}(m)|\, d\nu(f) < 4\varepsilon. \tag{10.18}$$

Next, we study the term $A_3(N)$. We utilize Property (ii) of Theorem 2.1, namely

$$|f_{N,\mathrm{st}}(n+Q)-f_{N,\mathrm{st}}(n)| \leq \frac{R}{\widetilde{N}} \quad \text{for every } \ n\in\mathbb{Z}_{\widetilde{N}}.$$

We get for $m\in\mathbb{Z}_{\widetilde{N}}$, $1\leq k\leq \eta N$, and for $i=1,2,3,4$, that

$$|f_{N,\mathrm{st}}(m+\ell_iQk)-f_{N,\mathrm{st}}(m)| \leq \ \ell_i k\,\frac{R}{\widetilde{N}} \leq \ell\eta N\,\frac{R}{\widetilde{N}} \leq \frac{\varepsilon}{Q},$$

where the last estimate follows from (10.15) and the estimate $\widetilde{N}\geq \ell N$. Using this estimate in conjunction with the definition of $A_3(N)$, we get

$$A_3(N) \geq \int_{\mathcal{M}_1^c} \mathbb{E}_{m\in\mathbb{Z}_{\widetilde{N}}}|f_{N,\mathrm{st}}(m)|^4\, d\nu(f) - \frac{3\varepsilon}{Q}.$$

Recall that $\mathbf{1}$ denotes the multiplicative function that is identically equal to 1. By Property (i) of Proposition 10.5 we have $\nu(\{\mathbf{1}\})\geq\delta^2$. Using this, we deduce that

$$\int_{\mathcal{M}_1^c} \mathbb{E}_{m\in\mathbb{Z}_{\widetilde{N}}}|f_{N,\mathrm{st}}(m)|^4\, d\nu(f) \geq \nu(\{\mathbf{1}\})\cdot\mathbb{E}_{m\in\mathbb{Z}_{\widetilde{N}}}|\mathbf{1}_{N,\mathrm{st}}(m)|^4 \geq \delta^2\left|\mathbb{E}_{m\in\mathbb{Z}_{\widetilde{N}}}\mathbf{1}_{N,\mathrm{st}}(m)\right|^4.$$

Since $\mathbf{1}_{N,\mathrm{st}} = \mathbf{1}_N * \psi$ for some kernel ψ on $\mathbb{Z}_{\widetilde{N}}$ and $\widetilde{N}\leq 4\ell N$, we have

$$\mathbb{E}_{m\in\mathbb{Z}_{\widetilde{N}}}\mathbf{1}_{N,\mathrm{st}}(m) = \mathbb{E}_{m\in\mathbb{Z}_{\widetilde{N}}}\mathbb{E}_{k\in\mathbb{Z}_{\widetilde{N}}}\mathbf{1}_N(k)\psi(m-k) = \mathbb{E}_{k\in\mathbb{Z}_{\widetilde{N}}}\mathbf{1}_N(k) = \frac{N}{\widetilde{N}} \geq \frac{1}{4\,\ell}.$$

Combining the above we get

$$A_3(N) \geq \frac{\delta^2}{4^4\,\ell^4} - \frac{3\varepsilon}{Q}. \tag{10.19}$$

Putting (10.14), (10.17), (10.18), and (10.19) together, we get

$$A(N) \geq \varepsilon\,\frac{1}{40\,\ell QR}\Big(\frac{\delta^2}{4^4\,\ell^4} - 7\varepsilon\Big) - \frac{4C_2}{F(Q,R,\varepsilon)^{\frac12}} - \frac{8}{\widetilde{N}}.$$

Recall that $\varepsilon = c_1\delta^2$, for some positive constant c_1 that we left unspecified until now. We choose $c_1<1$, depending only on ℓ, so that

$$\frac{1}{40\,\ell}\Big(\frac{\delta^2}{4^4\,\ell^4} - 7\varepsilon\Big) \geq c_2\delta^2$$

for some positive constant c_2 that depends only on ℓ. Then we have

$$A(N) \geq \delta^2\,\frac{c_2\varepsilon}{QR} - \frac{4C_2}{F(Q,R,\varepsilon)^{\frac12}} - \frac{8}{\widetilde{N}}.$$

Recall that

$$F(Q,R,\varepsilon)=C_1^2\frac{Q^2R^2}{\varepsilon^4},$$

where C_1 was not determined until this point. We choose

$$C_1:=\frac{8c_1C_2}{c_2},$$

and upon recalling that $\varepsilon=c_1\delta^2$, we get

$$A(N)+\frac{8}{\tilde{N}}\geq\delta^2\,\frac{c_2\varepsilon}{QR}-C_2\,\frac{4\varepsilon^2}{C_1QR}=\frac{c_2\delta^2\varepsilon}{2\,QR}=\frac{c_1c_2\delta^4}{2\,QR}>0.$$

Recall that Q and R are bounded by a constant that depends only on δ. Hence, $A(N)$ is greater than a positive constant that depends only on δ, and in particular is independent of N, provided that N is sufficiently large, depending only on δ, as indicated above. This completes the proof of Proposition 10.5. ☐

10.7. **Proof of Theorem 2.13.** The proof of Theorem 2.13 goes along the lines of Theorem 2.12 with small changes only.

As a first step we reduce matters to the case where the coefficient of m in all linear forms is 1. For $i=1,\dots,s$, let the linear forms be given by $L_{1,i}(m,n):=\kappa_im+\lambda_in$, $L_{2,i}(m,n):=\kappa_i'm+\lambda_i'n$, where $\kappa_i,\kappa_i',\lambda_i,\lambda_i'\in\mathbb{Z}$. Let $\ell_0:=\prod_{i=1}^s\kappa_i=\prod_{i=1}^s\kappa_i'$ where the second equality follows by our assumption. We also have $\ell_0\neq 0$ by assumption, and we can assume that $\ell_0>0$; the other case can be treated similarly. Inserting ℓ_0n in place of n and factoring out the coefficients of m we reduce to the case where $\kappa_i=\kappa_i'=1$ for $i=1,\dots,s$. Our assumption gives that $\{\lambda_1,\dots,\lambda_s\}\neq\{\lambda_1',\dots,\lambda_s'\}$.

Theorem 2.13 can be deduced from an analytic statement completely similar to Proposition 10.5. By an induction on s we can reduce to the case that the integers $\lambda_1,\dots,\lambda_s$ and $\lambda_1',\dots,\lambda_s'$ are distinct, and furthermore we can assume that the smallest one is equal to 0. The rest of the argument is identical to the proof of Proposition 10.5 given in this section; the only difference is that in place of Theorem 2.1 for the U^3-norm we use the same result for the U^{2s-1}-norm.

Appendix A. Elementary facts about Gowers norms

In this section we gather some elementary facts about the U^s-norms that we use throughout the main body of the article.

A.1. **Gowers norms and restriction to subintervals.** Our first result shows that if the $U^s(\mathbb{Z}_N)$-norm of a function is sufficiently small, then its restriction to an arbitrary subinterval of $[N]$ is small.

Lemma A.1. *Let $s\geq 2$ be an integer and $\varepsilon>0$. There exists $\delta:=\delta(s,\varepsilon)>0$ and $N_0:=N_0(s,\varepsilon)>0$ such that for every integer $N\geq N_0$, interval $J\subset[N]$, and $a\colon\mathbb{Z}_N\to\mathbb{C}$ with $|a|\leq 1$, the following implication holds:*

$$\text{if}\ \ \|a\|_{U^s(\mathbb{Z}_N)}\leq\delta,\quad\text{then}\ \ \|\mathbf{1}_J\cdot a\|_{U^s(\mathbb{Z}_N)}\leq\varepsilon.$$

Proof. Without loss, we can assume that $0<\|a\|_{U^s(\mathbb{Z}_N)}<\frac{1}{4}$ and that the length of J is an even number, say $2L$. Furthermore, since the $U^s(\mathbb{Z}_N)$-norm is invariant under translations, we can assume that $J=[2L]$.

Let l be an integer with $0<l<L$ that will be defined later. Let $\phi:=\phi(l,L)$ be a "trapezoid function" on $\mathbb{Z}_N$ that increases linearly from 0 to 1 on the interval $[l]$, is equal to 1 between l and $2L-l$, decreases linearly from 1 to 0 between

$2L-l$ and $2L$, and is equal to 0 between $2L$ and N. This function is a variant of the function used in de la Vallée-Poussin sums. Indeed, let ϕ_1 and ϕ_2 be the "triangle functions" of height 1 and of base $[0,2L]$ and $[l,2L-l]$, respectively. These functions are images under some translation of classical Fejer kernels on $\mathbb{Z}_N$ and thus $\|\widehat{\phi}_1\|_{l^1(\mathbb{Z}_N)}=\|\widehat{\phi}_2\|_{l^1(\mathbb{Z}_N)}=1$. Furthermore, for $n\in\mathbb{Z}_N$ we have

$$\phi(n)=\frac{L}{l}\phi_1(n)-\frac{L-l}{l}\phi_2(n),$$

and thus

$$\|\widehat{\phi}\|_{l^1(\mathbb{Z}_N)}\leq\frac{2L}{l}. \tag{A.1}$$

Since the $U^s(\mathbb{Z}_N)$-norm is invariant under multiplication by $\mathrm{e}(n\xi/N)$ for $s\geq 2$, using the triangle inequality for the $U^s(\mathbb{Z}_N)$-norm and (A.1) we get

$$\|\phi\cdot a\|_{U^s(\mathbb{Z}_N)}\leq\frac{2L}{l}\,\|a\|_{U^s(\mathbb{Z}_N)}. \tag{A.2}$$

Furthermore, since $\mathbf{1}_{[2L]}-\phi$ is supported on an interval of length $2l$ and is bounded by 1, it follows that

$$\|\mathbf{1}_{[2L]}\cdot a-\phi\cdot a\|_{U^s(\mathbb{Z}_N)}\leq\Big(\frac{2l}{N}\Big)^{2^{-s}}. \tag{A.3}$$

Using (A.2), (A.3), and the triangle inequality for the $U^s(\mathbb{Z}_N)$-norm, we get

$$\|\mathbf{1}_{[2L]}\cdot a\|_{U^s(\mathbb{Z}_N)}\leq\frac{2L}{l}\,\|a\|_{U^s(\mathbb{Z}_N)}+\Big(\frac{2l}{N}\Big)^{2^{-s}}. \tag{A.4}$$

We choose $l:=\lfloor 2L\cdot\|a\|_{U^s(\mathbb{Z}_N)}^{2^s/(2^s+1)}\rfloor+1$. Since $\|a\|_{U^s(\mathbb{Z}_N)}<\frac{1}{4}$, we get $1\leq l\leq L$, and (A.4) together with the estimate $l\leq N\|a\|_{U^s(\mathbb{Z}_N)}^{2^s/(2^s+1)}+1$ gives the bound

$$\|\mathbf{1}_{[2L]}\cdot a\|_{U^s(\mathbb{Z}_N)}\leq 3\|a\|_{U^s(\mathbb{Z}_N)}^{1/(2^s+1)}+2N^{-2^{-s}}.$$

The asserted result follows at once from this estimate. □

A.2. **Relations between the norms $U^s(\mathbb{Z}_N)$ and $U^s[N]$.** Our next goal is to show that the $U^s(\mathbb{Z}_N)$ and $U^s[N]$ norms (both defined in Section 2.1.1) are equivalent measures of randomness. We make this precise in Lemma A.4 and Proposition A.5. We start with two preliminary lemmas.

Lemma A.2. *Let $s,N,N^*\in\mathbb{N}$ with $s\geq 2$ and $N^*\geq N$, and let $J\subset\mathbb{Z}_N$ be an interval of length smaller than $N/2$. Then, for every function $a\colon[N]\to\mathbb{C}$ we have*

$$\|\mathbf{1}_J\cdot a\|_{U^s(\mathbb{Z}_N)}=\Big(\frac{N^*}{N}\Big)^{(s+1)/2^s}\|\mathbf{1}_J\cdot a\|_{U^s(\mathbb{Z}_{N^*})}.$$

Proof. The proof goes by induction on s. The result is obvious for $s=1$. Suppose that the result holds for $s\geq 1$; we are going to show that it holds for $s+1$. Substituting $\mathbf{1}_J\cdot a$ for a, we can (and will) assume henceforth that a vanishes outside J. Since the Gowers norms are invariant under translation, after shifting the interval J to the left we can assume that $J=[L]$ for some integer L with $0<L\leq N/2$.

For convenience, we identify $\mathbb{Z}_N$ and $\mathbb{Z}_{N^*}$ with the intervals $I_N:=[-\lceil N/2\rceil,\lfloor N/2\rfloor)$ and $I_N^*:=\big[-\lceil N^*/2\rceil,\lfloor N^*/2\rfloor\big)$, respectively. For $t\in\mathbb{Z}_N$ we let $a_t\colon\mathbb{Z}_N\to\mathbb{C}$ be defined by $a_t(n):=a(t+n\bmod N)$, and for $t\in\mathbb{Z}_N^*$ we let $a_t\colon\mathbb{Z}_N\to\mathbb{C}$ be defined by $a_t(n) := a(t+n \bmod N^*)$. Keeping in mind that the function a vanishes outside $[L]$ we see that the following properties hold:

(i) If $t \in I_N^*$ and $t \notin I_N$, then the function $a\overline{a_t^*}$ is identically zero.
(ii) If $t \in I_N$ and $|t| \geq L$, then the functions $a\overline{a_t^*}$ and $a\overline{a_t}$ vanish.
(iii) If $|t| < L$, the functions $a\overline{a_t^*}$ and $a\overline{a_t}$ vanish outside $[L]$ and coincide for $n \in [L]$.

Therefore,

$$\|a\|_{U^{s+1}(\mathbb{Z}_N)}^{2^{s+1}} = \mathbb{E}_{t\in I_N} \|a\,\overline{a_t}\|_{U^s(\mathbb{Z}_N)}^{2^s} = \frac{1}{N}\sum_{|t|<L} \|a\,\overline{a_t}\|_{U^s(\mathbb{Z}_N)}^{2^s},$$

where the last equality follows from Property (ii). Using Property (iii) and the induction hypothesis, we see that the last quantity is equal to

$$\frac{1}{N}\sum_{|t|<L} \|a\,\overline{a_t^*}\|_{U^s(\mathbb{Z}_N)}^{2^s} = \frac{1}{N}\sum_{|t|<L} \Big(\Big(\frac{N^*}{N}\Big)^{(s+1)/2^s} \|a\,\overline{a_t^*}\|_{U^s(\mathbb{Z}_{N^*})}\Big)^{2^s},$$

which in turn, by Properties (i) and (ii), is equal to

$$\Big(\frac{N^*}{N}\Big)^{s+2} \mathbb{E}_{t\in I_{N^*}} \|a\,\overline{a_t^*}\|_{U^s(\mathbb{Z}_{N^*})} = \Big(\frac{N^*}{N}\Big)^{s+2} \|a\|_{U^{s+1}(\mathbb{Z}_{N^*})}^{2^{s+1}}.$$

This completes the induction and the proof. □

Lemma A.3. *If $N^* \geq N$, then for every $s \geq 2$ we have*

$$\|\mathbf{1}_{[N]}\|_{U^s(\mathbb{Z}_{N^*})} \geq \frac{N}{N^*}.$$

Proof. By the monotonicity property (2.2) we have

$$\|\mathbf{1}_{[N]}\|_{U^s(\mathbb{Z}_{N^*})} \geq \|\mathbf{1}_{[N]}\|_{U^1(\mathbb{Z}_{N^*})} = \frac{N}{N^*}$$

as required. □

Lemma A.4. *Let $s \geq 2$ be an integer and $\varepsilon > 0$. There exists $\delta := \delta(s,\varepsilon) > 0$ and $N_0 := N_0(s,\varepsilon) > 0$ such that for every integer $N \geq N_0$ and every function $a\colon [N] \to \mathbb{C}$ with $|a| \leq 1$ we have*

$$\text{if } \|a\|_{U^s[N]} \leq \delta, \text{ then } \|a\|_{U^s(\mathbb{Z}_N)} \leq \varepsilon;$$
$$\text{if } \|a\|_{U^s(\mathbb{Z}_N)} \leq \delta, \text{ then } \|a\|_{U^s[N]} \leq \varepsilon.$$

Proof. Let $\delta := \delta(s,\varepsilon/9)$, $N_0 := N_0(s,\varepsilon/9)$ be defined as in Lemma A.1. Let $N \geq N_0$ be an integer and $a\colon [N] \to \mathbb{C}$ be a function with $|a| \leq 1$.

Suppose first that $\|a\|_{U^s[N]} \leq \delta$. By Lemma A.2 and the definition of the $U^s[N]$-norms we have

$$\|\mathbf{1}_{[N]}\cdot a\|_{U^s(\mathbb{Z}_{3N})} = \|\mathbf{1}_{[N]}\|_{U^s(\mathbb{Z}_{3N})}\cdot \|a\|_{U^s[N]} \leq \|a\|_{U^s[N]} \leq \delta.$$

We partition the interval $[N]$ into three intervals of length less than $N/2$. If J is any of these intervals, by Lemma A.1 applied to the function $\mathbf{1}_{[N]}\cdot a$ and the definition of δ we have $\|\mathbf{1}_J\cdot a\|_{U^s(\mathbb{Z}_{3N})} \leq \varepsilon/9$. By Lemma A.2 we have

$$\|\mathbf{1}_J\cdot a\|_{U^s(\mathbb{Z}_N)} \leq 3^{(s+1)/2^s}\varepsilon/9 \leq \varepsilon/3. \tag{A.5}$$

Taking the sum of these estimates for the three intervals J that partition of $[N]$ we get $\|a\|_{U^s(\mathbb{Z}_N)} \leq \varepsilon$.

Suppose now that $\|a\|_{U^s(\mathbb{Z}_N)} \leq \delta$. As above, we partition the interval $[N]$ into three intervals of length less than $N/2$. If J is any of these intervals, by Lemma A.1 and the definition of δ we have $\|\mathbf{1}_J \cdot a\|_{U^s(\mathbb{Z}_N)} \leq \varepsilon/9$. Hence, by Lemma A.2 we have $\|\mathbf{1}_J \cdot a\|_{U^s(\mathbb{Z}_{3N})} = 3^{-(s+1)/2^s}\|\mathbf{1}_J \cdot a\|_{U^s(\mathbb{Z}_N)} \leq \varepsilon/9$. Last, note that the definition of the $U^s[N]$-norm and Lemma A.3 give $\|\mathbf{1}_J \cdot a\|_{U^s[N]} \leq \varepsilon/3$. Taking the sum of these estimates for the three intervals J that partition $[N]$ we deduce that $\|a\|_{U^s[N]} \leq \varepsilon$. This completes the proof. □

Proposition A.5. *Let $s \geq 2$ and $a\colon \mathbb{N} \to \mathbb{C}$ be bounded. Then the following properties are equivalent:*

(i) $\|a\|_{U^s[N]} \to 0$ *as* $N \to +\infty$;
(ii) $\|a\|_{U^s(\mathbb{Z}_N)} \to 0$ *as* $N \to +\infty$;
(iii) *there exists $C > 1$ and a sequence (N_j) of integers with $N_j < N_{j+1} \leq CN_j$ for every $j \in \mathbb{N}$ such that*

$$\|a\|_{U^s[N_j]} \to 0 \text{ as } j \to +\infty.$$

Proof. The equivalence between (i) and (ii) is given by Lemma A.4.

The implication (i) $\implies$ (iii) is obvious. We show that (iii) $\implies$ (i). Let (N_j) and C be as in the statement. For given $\varepsilon > 0$ let $\delta := \delta(s, \varepsilon/3C)$ and $N_0 := N_0(s, \varepsilon/3C)$ be given by Lemma A.1. Let j_0 be such that $\|a\|_{U^s[N_j]} \leq \delta$ for $j \geq j_0$.

Let $N \geq \max\{N_{j_0}, N_0\}$ be an integer. Let j be the smallest integer such that $N_j \geq N$. By hypothesis, $j \geq j_0$ and $N_j \leq CN$. Let $\widetilde{N}_j := 3N_j$. By Lemma A.2 we have

$$\|1_{[N_j]}a\|_{U^s(\mathbb{Z}_{\widetilde{N}_j})} = \|\mathbf{1}_{[N_j]}\|_{U^s(\mathbb{Z}_{\widetilde{N}_j})} \cdot \|a\|_{U^s[N_j]} \leq \|a\|_{U^s[N_j]} \leq \delta.$$

Thus, by the definition of δ, we have $\|\mathbf{1}_{[N]} \cdot a\|_{U^s(\mathbb{Z}_{\widetilde{N}_j})} \leq \varepsilon/3C$. Combining this and Lemmas A.2 and A.3 we get

$$\|a\|_{U^s[N]} = \|\mathbf{1}_{[N]}\|^{-1}_{U^s(\mathbb{Z}_{\widetilde{N}_j})} \cdot \|\mathbf{1}_{[N]} \cdot a\|_{U^s(\mathbb{Z}_{\widetilde{N}_j})} \leq \frac{\widetilde{N}_j}{N} \cdot \|\mathbf{1}_{[N]} \cdot a\|_{U^s(\mathbb{Z}_{\widetilde{N}_j})} \leq \frac{\widetilde{N}_j}{N} \cdot \frac{\varepsilon}{3C} \leq \varepsilon.$$

Hence, $\limsup_{N\to+\infty}\|a\|_{U^s[N]} \leq \varepsilon$. As ε is arbitrary, we get (i), completing the proof. □

A.3. **Some estimates involving Gowers norms.** We record here two easy estimates that were used in the main text.

Lemma A.6. *There exists a constant $C > 0$ such that for every prime number N, function $a\colon \mathbb{Z}_N \to \mathbb{C}$, and arithmetic progression P contained in the interval $[N]$, we have*

$$\big|\mathbb{E}_{n\in[N]}\mathbf{1}_P(n) \cdot a(n)\big| \leq C\|a\|_{U^2(\mathbb{Z}_N)}.$$

Proof. If N is a prime, since the $U^2(\mathbb{Z}_N)$-norm of a function on $\mathbb{Z}_N$ is invariant under any change of variables of the form $x \mapsto kx + l \bmod N$, where $k, l \in \mathbb{Z}_N$ with $k \neq 0 \bmod \mathbb{Z}_N$, we can reduce matters to the case where $P = \{0, \dots, m\}$ for some $m \in \{0, \dots, N-1\}$, considered as a subset of $\mathbb{Z}_N$. In this case, a direct computation shows that

$$|\widehat{\mathbf{1}_P}(\xi)| \leq \frac{2}{N||\xi/N||} = \frac{2}{\min\{\xi, N-\xi\}} \quad \text{for } \xi = 1, \dots, N-1,$$

and as a consequence

$$\|\widehat{\mathbf{1}_P}(\xi)\|_{l^{4/3}(\mathbb{Z}_N)} \leq C$$

for some universal constant C. Using this estimate, Parseval's identity, Hölder's inequality, and identity (2.3), we deduce that

$$\left|\mathbb{E}_{n\in[N]}\mathbf{1}_P(n)\cdot a(n)\right| = \Big|\sum_{\xi\in[N]} \widehat{\mathbf{1}}_P(\xi)\cdot\widehat{a}(\xi)\Big| \leq C\cdot\Big(\sum_{\xi\in[N]} |\widehat{a}(\xi)|^4\Big)^{1/4} = C\,\|a\|_{U^2(\mathbb{Z}_N)}. \qquad \square$$

Lemma A.7. *There exists a constant $C > 0$ such that for every $N \in \mathbb{N}$ and function $a\colon \mathbb{Z}_N \to \mathbb{C}$ we have*

$$\sup_{t\in\mathbb{R}}\left|\mathbb{E}_{n\in[N]}a(n)\,\mathrm{e}(nt)\right| \leq C\|a\|_{U^2(\mathbb{Z}_N)}.$$

Proof. Writing $\phi_t(n) := \mathrm{e}(nt)$, a direct computation gives that $\|\widehat{\phi_t}\|_{l^{4/3}(\mathbb{Z}_N)} \leq C$ for some universal constant C, and the result follows as above from Parseval's identity, Hölder's inequality, and identity (2.3). $\square$

Appendix B. Rational elements in a nilmanifold

We collect here some properties of rational elements and rational subgroups. Additional relevant material can be found in [30] and in [56].

Let $X := G/\Gamma$ be an s-step nilmanifold of dimension m. As everywhere in this article we assume that G is connected and simply connected, and endowed with a Mal'cev basis. Recall that we write e_X for the image in X of the unit element $\mathbf{1}_G$ of G.

From Properties (iii) and (iv) of Mal'cev bases stated in Section 4.1, we immediately deduce the following lemma.

Lemma B.1. *The group Γ is finitely generated.*

B.1. **Rational elements.** We recall that an element $g \in G$ is Q-rational if $g^n \in \Gamma$ for some $n \in \mathbb{N}$ with $n \leq Q$. We note that all quantities introduced below depend implicitly on the nilmanifold X.

Lemma B.2 ([30, Lemma A.11])**.**

(i) *For every $Q \in \mathbb{N}$ there exists $Q' \in \mathbb{N}$ such that the product of any two Q-rational elements is Q'-rational; it follows that the set of rational elements is a subgroup of G.*

(ii) *For every $Q \in \mathbb{N}$ there exists $Q' \in \mathbb{N}$ such that the Mal'cev coordinates of any Q-rational element are rational with denominators at most Q'; it follows that the set of Q-rational elements is a discrete subset of G.*

(iii) *Conversely, for every $Q' \in \mathbb{N}$ there exists $Q \in \mathbb{N}$ such that, if the Mal'cev coordinates of $g \in G$ are rational with denominators at most Q', then g is Q-rational.*

Corollary B.3. *For every $Q \in \mathbb{N}$ there exists a finite set $\Sigma := \Sigma(Q)$ of Q-rational elements such that all Q-rational elements belong to $\Sigma(Q)\Gamma$.*

Proof. Let K be a compact subset of G such that $G = K\Gamma$.

Let $Q \in \mathbb{N}$. Let Q' be associated to Q by Part (i) of Lemma B.2, and let Σ' be the set of Q'-rational elements of K. By Part (ii) of Lemma B.2, Σ' is finite. Let g be a Q-rational element of G. There exists $\gamma \in \Gamma$ such that $g\gamma^{-1} \in K$. Since γ is obviously Q-rational, $g\gamma^{-1}$ is Q'-rational, and thus it belongs to Σ'. For each element h of Σ' obtained this way we choose a Q-rational point g such that $h \in g\Gamma$. Let $\Sigma := \Sigma(Q)$ be the set consisting of all elements obtained this way. Thus, $\Sigma\Gamma$ contains all Q-rational elements. Furthermore, $|\Sigma| \leq |\Sigma'|$ and so Σ is finite, completing the proof. □

B.2. **Rational subgroups.** We gather here some basic properties of rational subgroups that we use in the main part of the article.

Recall that a *rational subgroup* G' of G is a closed, connected, and simply connected subgroup of G such that $\Gamma' := \Gamma \cap G'$ is co-compact in G. In this case, G'/Γ' is called a *sub-nilmanifold* of X. It can be shown that G' is a rational subgroup of G if and only if its Lie algebra $\mathfrak{g}'$ admits a base that has rational coordinates in the Mal'cev basis of G.

Lemma B.4 ([30, Lemma A.13]). *If G' is a rational subgroup of G and h is a rational element, then $hG'h^{-1}$ is a rational subgroup of G.*

Proof. The conjugacy map $h \mapsto g^{-1}hg$ is a polynomial map with rational coefficients, and thus the linear map Ad_h from $\mathfrak{g}$ to itself has rational coefficients. Since $\mathfrak{g}'$ has a base consisting of vectors with rational coefficients, the same property holds for $\mathrm{Ad}_h\mathfrak{g}$, that is, for the Lie algebra of $hG'h^{-1}$. This proves the claim. □

The argument used to deduce Lemma B.1 shows that the group $\Gamma \cap (hG'h^{-1})$ is finitely generated.

We also need an auxiliary result.

Lemma B.5. *Let Θ be a finitely generated nilpotent group, and let Λ be a subgroup of Θ. Suppose that for every $\gamma \in \Theta$ there exists $n \in \mathbb{N}$ with $\gamma^n \in \Lambda$. Then Λ has a finite index in Θ.*

Proof. The proof goes by induction on the nilpotency degree s of Θ. If $s = 1$, Θ is Abelian and the result is immediate. Suppose that $s > 1$ and that the result holds for $(s-1)$-step nilpotent groups. By the induction hypothesis applied to the Abelian group Θ/Θ_2, the subgroup $(\Lambda\Theta_2)/\Theta_2$ has a finite index in Θ/Θ_2, and thus $\Lambda\Theta_2$ has a finite index in Θ. If $\gamma \in \Theta_2$, then there exists $n \in \mathbb{N}$ with $\gamma^n \in \Lambda$, and hence $\gamma^n \in \Lambda \cap \Theta_2$. Since Θ_2 is a finitely generated $(s-1)$-step nilpotent group, by the induction hypothesis again, $\Lambda \cap \Theta_2$ has a finite index in Θ_2. Thus, Λ has a finite index in $\Lambda\Theta_2$ which has a finite index in Θ. The result follows. □

Lemma B.6 ([10, Theorem 5.29]). *Let $X := G/\Gamma$ be an s-step nilmanifold, $G' \subset G$ be a rational subgroup, $g \in G$ be a rational element, and $\Lambda := \Gamma \cap (g^{-1}\Gamma g) \cap G'$. Then*

(i) *Λ is a subgroup of the finite index of $\Gamma \cap G'$;*
(ii) *Λ is a subgroup of the finite index of $(g^{-1}\Gamma g) \cap G'$.*

Proof. By Part (i) of Lemma B.2, all elements of $g\Gamma g^{-1}$ are rational. Hence, if $\gamma \in \Gamma \cap G'$, then there exists $n \in \mathbb{N}$ with $(g\gamma g^{-1})^n \in \Gamma$, and so we have $\gamma^n \in \Lambda$. Applying Lemma B.1 to G' and $\Gamma \cap G'$, we get the group Λ is finitely generated. By Lemma B.5, Λ has a finite index in $\Gamma \cap G'$. This proves (i). Since $g\Gamma g^{-1}$ is co-compact in G, substituting this group for G and g^{-1} for g in the preceding statement, we get (ii). □

Lemma B.7. *Let $g \in G$ be a rational element and G' a rational subgroup of G. Then $G'g \cdot e_X := \{hg \cdot e_X : h \in G'\}$ is a closed sub-nilmanifold of X.*

Proof. By Lemma B.4, $g^{-1}G'g$ is a rational subgroup of G. Therefore, $\Gamma\cap(g^{-1}G'g)$ is co-compact in $g^{-1}G'g$, and thus $(g\Gamma g^{-1})\cap G'$ is co-compact in G'. Note that $(g\Gamma g^{-1})\cap G'$ is the stabilizer $\{h\in G'\colon hg\cdot e_X = g\cdot e_X\}$ of $g\cdot e_X$ in G', and thus the orbit $G'g\cdot e_X$ is compact and can be identified with the nilmanifold $G'/((g\Gamma g^{-1})\cap G')$. □

Appendix C. Zeros of some homogeneous quadratic forms

We prove Proposition 2.11. We recall the statement for the reader's convenience.

Proposition. *Let the quadratic form p satisfy the hypothesis of Theorem* 2.7*. Then there exist admissible integers $\ell_0,\dots,\ell_4$ (see definition in Section* 2.4.2*), such that for every $k,m,n\in\mathbb{Z}$, the integers $x := k\ell_0(m+\ell_1 n)(m+\ell_2 n)$ and $y := k\ell_0(m+\ell_3 n)(m+\ell_4 n)$ satisfy the equation $p(x,y,z)=0$ for some $z\in\mathbb{Z}$.*

Proof. Let

$$ax^2+by^2+cz^2+dxy+exz+fyz=0 \tag{C.1}$$

be the equation we are interested in solving. Recall that by assumption a,b,c are non-zero integers and that all three integers

$$\Delta_1 := e^2-4ac,\quad \Delta_2 := f^2-4bc,\quad \Delta_3 := (e+f)^2-4c(a+b+d)$$

are non-zero squares.

Step 1. We first reduce to the case where $e=f=0$. Let

$$p'(x,y,z) := p(2cx,2cy,z-ex-fy).$$

Then

$$p'(x,y,z) = c(4ac-e^2)x^2+c(4bc-f^2)y^2+cz^2+2c(2cd-ef)xy.$$

The coefficients of x^2, y^2, z^2 in the quadratic form p' are non-zero by hypothesis. The discriminants of the three quadratic forms $p'(x,0,z)$, $p'(0,y,z)$, $p'(x,x,z)$ are equal to $4c^2\Delta_1$, $4c^2\Delta_2$, $4c^2\Delta_3$, respectively, and thus are non-zero squares by hypothesis. Suppose that the announced result holds for the quadratic form p'. Then there exist admissible integers $\ell_0,\dots,\ell_4$, such that for every $k,m,n\in\mathbb{Z}$, the integers $x' := k\ell_0(m+\ell_1 n)(m+\ell_2 n)$ and $y' := k\ell_0(m+\ell_3 n)(m+\ell_4 n)$ satisfy the equation $p'(x',y',z')=0$ for some $z'\in\mathbb{Z}$. It follows that $x := 2ck\ell_0(m+\ell_1 n)(m+\ell_2 n)$ and $y := 2ck\ell_0(m+\ell_3 n)(m+\ell_4 n)$ satisfy the equation $p(x,y,z)=0$ for $z := z'-ex'-fy'$. If $c>0$, we are done; if $c<0$, we consider the solution $-x,-y,-z$.

Step 2. We consider now the case where $e=f=0$. Then

$$p(x,y,z) = ax^2+by^2+cz^2+dxy.$$

Our hypothesis is that a,b,c are non-zero and the integers $-ac$, $-bc$, $-c(a+b+d)$ are non-zero squares. Without loss, we can restrict to the case where $a>0$ and thus $c<0$. By taking products we get that $ac^2(a+b+d)$ and $bc^2(a+b+d)$ are non-zero squares, and thus $a(a+b+d)$ and $b(a+b+d)$ are non-zero squares. We let

$$\Delta_1' := \sqrt{b(a+b+d)}\ ;\ \Delta_2' := \sqrt{a(a+b+d)}\ ;\ \Delta_3' := \sqrt{-c(a+b+d)},$$

and

$$\ell_0 := -c; \qquad \ell_1 := -(b+\Delta_1'); \qquad \ell_2 := -(b-\Delta_1');$$
$$\ell_3 := -(a+d+\Delta_2')\ ; \qquad \ell_4 := -(a+d-\Delta_2').$$

By direct computation, we check that for every $k, m, n \in \mathbb{Z}$, the integers x, y, z given by

$$x := -kc\big(m^2 - 2bmn - b(a+d)n^2\big) = k\ell_0(m+\ell_1 n)(m+\ell_2 n);$$
$$y := -kc\big(m^2 + 2(a+d)mn + (ad+d^2-ab)n^2\big) = -k\ell_0(m+\ell_3 n)(m+\ell_4 n);$$
$$z := \quad k\Delta_3'(m^2 + dmn + abn^2),$$

satisfy $p(x, y, z) = 0$.

Since $c < 0$, we have $\ell_0 > 0$. Furthermore, since $\Delta_1 \neq 0$, we have $\ell_1 \neq \ell_2$, and since $\Delta_2 \neq 0$, we have $\ell_3 \neq \ell_4$. Last, we verify that $\{\ell_1, \ell_2\} \neq \{\ell_3, \ell_4\}$. Indeed, if these pairs were identical, then the coefficients of mn in x and y would be the same. Hence, $-2b = 2(a+d)$, and thus $a+b+d = 0$, contradicting our hypothesis. We conclude that the integers $\ell_0, \dots, \ell_4$ are admissible. This completes the proof. □

Acknowledgments

The authors would like to thank Wenbo Sun for pointing out a mistake in an earlier version of this article and Nikos Tzanakis for his help with the material in Section 2.4.3. The authors would also like to thank the referee for helpful suggestions and for pointing us to Chebotarev's theorem that helped to strengthen Theorem 2.6.

References

[1] Vitaly Bergelson, *Ergodic Ramsey theory—an update*, Ergodic theory of $\mathbf{Z}^d$ actions (Warwick, 1993), London Math. Soc. Lecture Note Ser., vol. 228, Cambridge Univ. Press, Cambridge, 1996, pp. 1–61, DOI 10.1017/CBO9780511662812.002. MR1411215 (98g:28017)

[2] Vitaly Bergelson, *Multiplicatively large sets and ergodic Ramsey theory*, Israel J. Math. **148** (2005), 23–40, DOI 10.1007/BF02775431. Probability in mathematics. MR2191223 (2008f:37017)

[3] Vitaly Bergelson, Bernard Host, and Bryna Kra, *Multiple recurrence and nilsequences*, Invent. Math. **160** (2005), no. 2, 261–303, DOI 10.1007/s00222-004-0428-6. With an appendix by Imre Ruzsa. MR2138068 (2007i:37009)

[4] V. Bergelson and J. Moreira, *Ergodic theorem involving additive and multiplicative groups of a field and {x+y,xy} patterns*, available at arXiv:1307.6242. To appear in Ergodic Theory Dynam. Systems.

[5] V. Bergelson and J. Moreira, *Measure preserving actions of affine semigroups and {x+y,xy} patterns*, available at arXiv:1509.07574.

[6] V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems*, J. Amer. Math. Soc. **9** (1996), no. 3, 725–753, DOI 10.1090/S0894-0347-96-00194-4. MR1325795 (96j:11013)

[7] J. Bourgain, P. Sarnak, and T. Ziegler, *Disjointness of Moebius from horocycle flows*, From Fourier analysis and number theory to Radon transforms and geometry, Dev. Math., vol. 28, Springer, New York, 2013, pp. 67–83, DOI 10.1007/978-1-4614-4075-8_5. MR2986954

[8] T. D. Browning and S. Prendiville, *A transference approach to a Roth-type theorem in the squares*, available at arXiv:1510.00136.

[9] S. Chowla, *The Riemann hypothesis and Hilbert's tenth problem*, Mathematics and Its Applications, Vol. 4, Gordon and Breach Science Publishers, New York-London-Paris, 1965. MR0177943 (31 #2201)

[10] Lawrence J. Corwin and Frederick P. Greenleaf, *Representations of nilpotent Lie groups and their applications. Part I*, Cambridge Studies in Advanced Mathematics, vol. 18, Cambridge University Press, Cambridge, 1990. Basic theory and examples. MR1070979 (92b:22007)

[11] Hédi Daboussi, *Fonctions multiplicatives presque périodiques B* (French), Journées Arithmétiques de Bordeaux (Conf., Univ. Bordeaux, Bordeaux, 1974), Soc. Math. France, Paris, 1975, pp. 321–324. Astérisque, No. 24–25. D'après un travail commun avec Hubert Delange. MR0374074 (51 #10274)

[12] Hédi Daboussi and Hubert Delange, *Quelques propriétés des fonctions multiplicatives de module au plus égal à* 1 (French), C. R. Acad. Sci. Paris Sér. A **278** (1974), 657–660. MR0332702 (48 #11028)

[13] Hédi Daboussi and Hubert Delange, *On multiplicative arithmetical functions whose modulus does not exceed one*, J. Lond. Math. Soc. (2) **26** (1982), no. 2, 245–264, DOI 10.1112/jlms/s2-26.2.245. MR675168 (83k:10082)

[14] P. Erdős and R. L. Graham, *Old and new problems and results in combinatorial number theory*, Monographies de L'Enseignement Mathématique [Monographs of L'Enseignement Mathématique], vol. 28, Université de Genève, L'Enseignement Mathématique, Geneva, 1980. MR592420 (82j:10001)

[15] G. Faltings, *Endlichkeitssätze für abelsche Varietäten über Zahlkörpern* (German), Invent. Math. **73** (1983), no. 3, 349–366, DOI 10.1007/BF01388432. MR718935 (85g:11026a)

[16] John Friedlander and Henryk Iwaniec, *The polynomial $X^2 + Y^4$ captures its primes*, Ann. of Math. (2) **148** (1998), no. 3, 945–1040, DOI 10.2307/121034. MR1670065 (2000c:11150a)

[17] Harry Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Anal. Math. **31** (1977), 204–256. MR0498471 (58 #16583)

[18] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press, Princeton, N.J., 1981. M. B. Porter Lectures. MR603625 (82j:28010)

[19] W. T. Gowers, *A new proof of Szemerédi's theorem*, Geom. Funct. Anal. **11** (2001), no. 3, 465–588, DOI 10.1007/s00039-001-0332-9. MR1844079 (2002k:11014)

[20] W. T. Gowers, *Decompositions, approximate structure, transference, and the Hahn-Banach theorem*, Bull. Lond. Math. Soc. **42** (2010), no. 4, 573–606, DOI 10.1112/blms/bdq018. MR2669681 (2011k:11015)

[21] W. T. Gowers and J. Wolf, *Linear forms and quadratic uniformity for functions on $\mathbb{Z}_N$*, J. Anal. Math. **115** (2011), 121–186, DOI 10.1007/s11854-011-0026-7. MR2855036 (2012k:11011)

[22] W. T. Gowers and J. Wolf, *Linear forms and higher-degree uniformity for functions on $\mathbb{F}_p^n$*, Geom. Funct. Anal. **21** (2011), no. 1, 36–69, DOI 10.1007/s00039-010-0106-3. MR2773103 (2012f:11024)

[23] Ron Graham, *Old and new problems and results in Ramsey theory*, Horizons of combinatorics, Bolyai Soc. Math. Stud., vol. 17, Springer, Berlin, 2008, pp. 105–118, DOI 10.1007/978-3-540-77200-2_5. MR2432529 (2009h:05207)

[24] Andrew Granville and K. Soundararajan, *Large character sums: pretentious characters and the Pólya-Vinogradov theorem*, J. Amer. Math. Soc. **20** (2007), no. 2, 357–384, DOI 10.1090/S0894-0347-06-00536-4. MR2276774 (2007k:11128)

[25] A. Granville and K. Soundararajan, *The pretentious approach.* To be published.

[26] Ben Green and Terence Tao, *Quadratic uniformity of the Möbius function* (English, with English and French summaries), Ann. Inst. Fourier (Grenoble) **58** (2008), no. 6, 1863–1935. MR2473624 (2010c:11012)

[27] Ben Green and Terence Tao, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proc. Edinb. Math. Soc. (2) **51** (2008), no. 1, 73–153, DOI 10.1017/S0013091505000325. MR2391635 (2009g:11012)

[28] Ben Green and Terence Tao, *An arithmetic regularity lemma, an associated counting lemma, and applications*, An irregular mind, Bolyai Soc. Math. Stud., vol. 21, János Bolyai Math. Soc., Budapest, 2010, pp. 261–334, DOI 10.1007/978-3-642-14444-8_7. MR2815606

[29] Benjamin Green and Terence Tao, *Linear equations in primes*, Ann. of Math. (2) **171** (2010), no. 3, 1753–1850, DOI 10.4007/annals.2010.171.1753. MR2680398 (2011j:11177)

[30] Ben Green and Terence Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Ann. of Math. (2) **175** (2012), no. 2, 465–540, DOI 10.4007/annals.2012.175.2.2. MR2877065

[31] Ben Green and Terence Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. of Math. (2) **175** (2012), no. 2, 541–566, DOI 10.4007/annals.2012.175.2.3. MR2877066

[32] Ben Green, Terence Tao, and Tamar Ziegler, *An inverse theorem for the Gowers $U^{s+1}[N]$-norm*, Ann. of Math. (2) **176** (2012), no. 2, 1231–1372, DOI 10.4007/annals.2012.176.2.11. MR2950773

[33] Katalin Gyarmati and Imre Z. Ruzsa, *A set of squares without arithmetic progressions*, Acta Arith. **155** (2012), no. 1, 109–115, DOI 10.4064/aa155-1-11. MR2982433

[34] G. Halász, *Über die Mittelwerte multiplikativer zahlentheoretischer Funktionen* (German), Acta Math. Acad. Sci. Hungar. **19** (1968), 365–403. MR0230694 (37 #6254)

[35] R. R. Hall, *A sharp inequality of Halász type for the mean value of a multiplicative arithmetic function*, Mathematika **42** (1995), no. 1, 144–157, DOI 10.1112/S0025579300011426. MR1346679 (97a:11152)

[36] D. R. Heath-Brown, *Primes represented by $x^3 + 2y^3$*, Acta Math. **186** (2001), no. 1, 1–84, DOI 10.1007/BF02392715. MR1828372 (2002b:11122)

[37] D. R. Heath-Brown and B. Z. Moroz, *Primes represented by binary cubic forms*, Proc. Lond. Math. Soc. (3) **84** (2002), no. 2, 257–288, DOI 10.1112/plms/84.2.257. MR1881392 (2002k:11157)

[38] D. R. Heath-Brown and B. Z. Moroz, *On the representation of primes by cubic polynomials in two variables*, Proc. Lond. Math. Soc. (3) **88** (2004), no. 2, 289–312, DOI 10.1112/S0024611503014497. MR2032509 (2004m:11147)

[39] Harald Andres Helfgott, *Root numbers and the parity problem*, ProQuest LLC, Ann Arbor, MI, 2003. Thesis (Ph.D.), Princeton University. MR2704329

[40] H. A. Helfgott, *The parity problem for reducible cubic forms*, J. Lond. Math. Soc. (2) **73** (2006), no. 2, 415–435, DOI 10.1112/S0024610706022629. MR2225495 (2007e:11115)

[41] H. Helfgott, *The parity problem for irreducible polynomials*, available at `arXiv:0501177`.

[42] Kevin Henriot, *Logarithmic Bounds for Translation-Invariant Equations in Squares*, Int. Math. Res. Not. IMRN **23** (2015), 12540–12562, DOI 10.1093/imrn/rnv062. MR3431629

[43] K. Henriot, *Additive equations in dense variables, and truncated restriction estimates*, available at `arXiv:1508.05923`.

[44] Bernard Host and Bryna Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. of Math. (2) **161** (2005), no. 1, 397–488, DOI 10.4007/annals.2005.161.397. MR2150389 (2007b:37004)

[45] Bernard Host and Bryna Kra, *Uniformity seminorms on ℓ^∞ and applications*, J. Anal. Math. **108** (2009), 219–276, DOI 10.1007/s11854-009-0024-1. MR2544760 (2010j:11018)

[46] I. Kátai, *A remark on a theorem of H. Daboussi*, Acta Math. Hungar. **47** (1986), no. 1-2, 223–225, DOI 10.1007/BF01949145. MR836415 (87j:11072)

[47] Eugen Keil, *On a diagonal quadric in dense variables*, Glasg. Math. J. **56** (2014), no. 3, 601–628, DOI 10.1017/S0017089514000056. MR3250267

[48] E. Keil, *Some refinements for translation invariant quadratic forms in dense sets*, available at `arXiv:1408.1535`.

[49] Ayman Khalfalah and Endre Szemerédi, *On the number of monochromatic solutions of $x+y = z^2$*, Combin. Probab. Comput. **15** (2006), no. 1-2, 213–227, DOI 10.1017/S0963548305007169. MR2195583 (2006g:11047)

[50] A. Lachand, *Entiers friables et formes binaires.* Ph.D. Thesis (2014), Université de Lorraine, tel-01104211.

[51] A. Lachand, *On the representation of friables by linear forms*, available at `arXiv: hal-01081277`.

[52] E. Landau, *Über die Klassenzahl imaginär-quadratischer Zahlkörper*, Göttinger Nachrichten (1918), 285–295.

[53] A. Leibman, *Polynomial sequences in groups*, J. Algebra **201** (1998), no. 1, 189–206, DOI 10.1006/jabr.1997.7269. MR1608723 (99c:20044)

[54] A. Leibman, *Polynomial mappings of groups*, Israel J. Math. **129** (2002), 29–60, DOI 10.1007/BF02773152. MR1910931 (2003g:20060)

[55] A. Leibman, *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold*, Ergodic Theory Dynam. Systems **25** (2005), no. 1, 201–213, DOI 10.1017/S0143385704000215. MR2122919 (2006j:37004)

[56] A. Leibman, *Rational sub-nilmanifolds of a compact nilmanifold*, Ergodic Theory Dynam. Systems **26** (2006), no. 3, 787–798, DOI 10.1017/S014338570500057X. MR2237471 (2007e:37002)

[57] A. I. Mal′cev, *On a class of homogeneous spaces* (Russian), Izv. Akad. Nauk. SSSR. Ser. Mat. **13** (1949), 9–32. MR0028842 (10,507d)

[58] Lilian Matthiesen, *Linear correlations amongst numbers represented by positive definite binary quadratic forms*, Acta Arith. **154** (2012), no. 3, 235–306, DOI 10.4064/aa154-3-2. MR2949879

[59] L. Matthiesen, *A consequence of the factorisation theorem for polynomial orbits on nilmanifolds. Corrigendum*, Acta Arith. **154** (2012), no. 3, 235–306.

[60] Lilian Matthiesen, *Correlations of the divisor function*, Proc. Lond. Math. Soc. (3) **104** (2012), no. 4, 827–858, DOI 10.1112/plms/pdr046. MR2908784

[61] Lilian Matthiesen, *Correlations of representation functions of binary quadratic forms*, Acta Arith. **158** (2013), no. 3, 245–252, DOI 10.4064/aa158-3-4. MR3040665

[62] L. Matthiesen, *Generalized Fourier coefficients of multiplicative functions*, available at `arXiv:1405.1018`.

[63] H. L. Montgomery and R. C. Vaughan, *Exponential sums with multiplicative coefficients*, Invent. Math. **43** (1977), no. 1, 69–82. MR0457371 (56 #15579)

[64] M. Ram Murty and Jody Esmonde, *Problems in algebraic number theory*, 2nd ed., Graduate Texts in Mathematics, vol. 190, Springer-Verlag, New York, 2005. MR2090972 (2005c:11130)

[65] Jürgen Neukirch, *Algebraic number theory*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 322, Springer-Verlag, Berlin, 1999. Translated from the 1992 German original and with a note by Norbert Schappacher; With a foreword by G. Harder. MR1697859 (2000m:11104)

[66] Richard Rado, *Studien zur Kombinatorik* (German), Math. Z. **36** (1933), no. 1, 424–470, DOI 10.1007/BF01188632. MR1545354

[67] A. Sárközy, *On difference sets of sequences of integers. III*, Acta Math. Acad. Sci. Hungar. **31** (1978), no. 3-4, 355–386, DOI 10.1007/BF01901984. MR487031 (80j:10062b)

[68] I. Schur, *Uber die Kongruenz* $x^m + y^m = z^m \pmod p$, Jahresber. Dtsch. Math.-Ver. **25** (1917), 114–117.

[69] Matthew L. Smith, *On solution-free sets for simultaneous quadratic and linear equations*, J. Lond. Math. Soc. (2) **79** (2009), no. 2, 273–293, DOI 10.1112/jlms/jdn073. MR2496514 (2010b:11134)

[70] W. Sun, *A structure theorem for multiplicative functions over the Gaussian integers and applications*, available at `arXiv:1405.0241`. To appear in *J. Anal. Math.*

[71] B. Szegedy, *On higher order Fourier analysis*, available at `arXiv:1203.2260`.

[72] Terence Tao, *A quantitative ergodic theory proof of Szemerédi's theorem*, Electron. J. Combin. **13** (2006), no. 1, Research Paper 99, 49. MR2274314 (2007i:37016)

[73] B. L. van der Waerden, *Beweis einer Baudetschen Vermutung*, Nieuw. Arch. Wisk. **15** (1927), 212–216.

[74] Mark Walters, *Combinatorial proofs of the polynomial van der Waerden theorem and the polynomial Hales-Jewett theorem*, J. Lond. Math. Soc. (2) **61** (2000), no. 1, 1–12, DOI 10.1112/S0024610799008388. MR1745405 (2001b:05219)

Department of Mathematics, University of Crete, Voutes University Campus, Heraklion 71003, Greece

E-mail address: `frantzikinakis@gmail.com`

Laboratoire d'analyse et de mathématiques appliquées, Université Paris-Est Marne-la-Vallée, UMR CNRS 8050, 5 Bd Descartes, 77454 Marne la Vallée Cedex, France

E-mail address: `bernard.host@univ-mlv.fr`