ANNALES

DE

L'INSTITUT FOURIER

Ben GREEN & Terence TAO

Quadratic uniformity of the Möbius function

Tome 58, n° 6 (2008), p. 1863-1935.

<http://aif.cedram.org/item?id=AIF_2008__58_6_1863_0>

Ann. Inst. Fourier, Grenoble
58, 6 (2008) 1863-1935

QUADRATIC UNIFORMITY OF THE MÖBIUS FUNCTION

by Ben GREEN & Terence TAO (*)

ABSTRACT. — We prove the "Möbius and Nilsequences Conjecture" for nilsystems of step 1 and 2. This paper forms a part of our program to generalise the Hardy-Littlewood method so as to handle *systems* of linear equations in primes.

RÉSUMÉ. — On établit la conjecture «Möbius et Nilsuites» pour les nilsystèmes de rang 1 et 2. Ce papier est une partie de notre programme, dont le but est une généralisation de la méthode de Hardy-Littlewood en vue d'étudier les *systèmes* d'équations linéaires dans les nombres premiers.

1. Introduction

The *Möbius function* $\mu : \mathbb{N} \to \{-1, 0, +1\}$, defined by

$$\mu(n) := \begin{cases} (-1)^k, & \text{if } n = p_1 p_2 \dots p_k \text{ for distinct primes } p_1, \dots, p_k \\ 0, & \text{if } n \text{ is not squarefree} \\ 1, & \text{if } n = 1 \end{cases}$$

plays a fundamental role in analytic number theory, especially with regard to the distribution of primes. A well-known metaprinciple holds that μ fluctuates so "randomly" that it is asymptotically orthogonal to any "low complexity" bounded sequence $f : \mathbb{N} \to \mathbb{C}$. We do not have a formal definition of "low complexity", but the examples of this section should convey the general flavour. Functions which arise from geometry and algebra, such

Keywords: Quadratic uniformity, Möbius function.
Math. classification: 11B99.
(*) The first author is a Clay Research Fellow and gratefully acknowledges the support of the Clay Institute. He also spent time, while this work was being carried out, at Trinity College, Cambridge and at the Massachusetts Institute of Technology, and is very happy to acknowledge the kind hospitality of both institutions. The second author is supported by a grant from the Packard Foundation.

as characters $n \mapsto e(n\alpha)$, are certainly of low complexity, whereas functions which depend on the prime factorization of n, such as μ itself, the von Mangoldt function Λ, and certain divisor sums arising in sieve theory, are not.

In our first example, and throughout the paper, we will use the following notation. We write $[N] := \{1, \dots, N\}$ to denote the integers from 1 to N, and $\mathbb{E}_{n\in A} f(n) := \frac{1}{|A|}\sum_{n\in A} f(n)$ to denote the average of a function $f : A \to \mathbb{C}$ on a non-empty finite set A. We also use $X \ll Y$ or $X = O(Y)$ to denote the claim that $|X| \leqslant CY$ for some absolute constant $C > 0$.

Example 1 (μ is strongly orthogonal to the constant function). — We have

$$\mathbb{E}_{n\in[N]}\mu(n) \ll e^{-c\sqrt{\log N}} \tag{1.1}$$

for all $N > 1$ and some absolute constant $c > 0$.

Remark. — This is essentially equivalent to the prime number theorem with the classical error term of Hadamard and de la Vallée Poussin.

In the next example, and throughout the paper, we use $X \ll_A Y$ or $X = O_A(Y)$ to denote the claim that $|X| \leqslant C_A Y$ for some constant $C_A > 0$ depending on A.

Example 2 (μ is strongly orthogonal to Dirichlet characters). — For any $A > 0$ we have

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{\chi(n)} \ll_A q^{1/2}\log^{-A} N \tag{1.2}$$

for all N and all Dirichlet characters χ to modulus q.

Remark. — See for instance [15, Corollary 5.29]. This may be used to prove the Siegel-Walfisz theorem concerning the distribution of primes in arithmetic progressions.

The form of the bound in (1.2) may appear strange at first sight. A key point to appreciate is that the implied constant $C = C_A$ is *ineffective*, due to the possible existence of Landau-Siegel zeros. The book [7] may be consulted for further information. It is useful to have a name for bounds of this kind.

Definition 1.1 (Strong asymptotic orthogonality). — *If $f : \mathbb{N} \to \mathbb{C}$ and $g : \mathbb{N} \to \mathbb{C}$ are two sequences on the natural numbers $\mathbb{N} = \{1, 2, 3, \dots\}$, we say that f and g are* strongly asymptotically orthogonal *if we have the estimate*

$$\mathbb{E}_{n\in[N]} f(n)\overline{g(n)} \ll_A \log^{-A} N$$

for all $N > 1$ and all $A > 0$. We allow the implied constant C_A to be ineffective*, in that we may have no explicit bounds on C_A other than that it is finite.*

Thus Example 2 shows that μ is strongly asymptotically orthogonal to all Dirichlet characters, and some Fourier analysis then shows that it is in fact strongly asymptotically orthogonal to any periodic sequence. In fact, more is true, as we shall see in the next example. Here, and throughout the paper, we use $e()$ to denote the standard character $e(x) := \exp(2\pi i x)$.

Example 3 (μ is strongly orthogonal to linear phases). — For any $\alpha \in \mathbb{R}/\mathbb{Z}$ and for any $A > 0$, we have

$$\mathbb{E}_{n\in[N]}\mu(n)e(-\alpha n) \ll_A \log^{-A} N, \tag{1.3}$$

uniformly in $\alpha \in \mathbb{R}/\mathbb{Z}$.

This bound is due to Davenport [6] and can be deduced from (1.2) by an application of Vinogradov's version of the Hardy-Littlewood major/minor arc decomposition of $\mathbb{R}/\mathbb{Z}$. See, for example, [15, Theorem 13.10]. For pedagogical reasons, and because we need this result for later sections, we give the derivation in §5. Davenport's result may be used on its own to obtain a number of self-correlation estimates on μ. For instance, by combining (1.3) with elementary Fourier analysis (the circle method) we easily obtain the estimates

$$\mathbb{E}_{x,d\in[N]}\mu(x)\mu(x+d)\mu(x+2d) \ll_A \log^{-A} N \tag{1.4}$$

and

$$\mathbb{E}_{x,h_1,h_2\in[N]}\mu(x)\mu(x+h_1)\mu(x+h_2)\mu(x+h_1+h_2) \ll_A \log^{-A} N. \tag{1.5}$$

Similar expressions in which μ is replaced by Λ, the von Mangoldt function, may be analysed using (1.3) as a key ingredient. The answers have a more complicated form involving a main term which is a product of local factors or *singular series*. See [15, §13] and [10] for different approaches to this[(1)].

A full discussion of results such as (1.4), (1.5) and the corresponding results for Λ is given in [10]. For comparison with that paper, we remark

[(1)] While the von Mangoldt function Λ is more directly related to the primes, the Möbius function μ is somewhat easier to handle analytically, being bounded by 1 and not encountering the "local" irregularities in small residue classes that Λ faces; in particular, the "major arc" terms will have a significantly simpler form. Also, the Vaughan identity for μ is slightly cleaner than that for Λ (see Lemma 4.1). Thus in this series of papers we have adopted a "Möbius first" philosophy, in which we obtain estimates on the Möbius function μ using "hard" analytic tools, and then use "softer" techniques to transfer the bounds on μ to the bounds on Λ.

that the two systems of linear forms in (1.4) and (1.5), namely $(x, x+d, x+2d)$ and $(x, x+h_1, x+h_2, x+h_1+h_2)$, both have *complexity* equal to one. This notion of complexity 1 essentially marks the limit of the classical Hardy-Littlewood circle method. The main goal of this paper is to provide some of the technical machinery needed to address the case of complexity 2.

We can reformulate (1.3) in a manner which may appear strange at first, but is well suited to generalisations, as we shall soon see. If X is any metric space, define a *Lipschitz function*[(2)] on X to be any function $f : X \to \mathbb{C}$ whose (inhomogeneous) Lipschitz norm

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

is finite.

Example 4 (μ is strongly orthogonal to 1-step nilsequences). — Suppose that G is a connected, simply-connected abelian Lie group (written multiplicatively) with a smooth metric d, and that Γ is a closed subgroup of G which is cocompact. Then G/Γ is called a 1-*step nilmanifold*; it is a torus. Let $F : G/\Gamma \to \mathbb{C}$ be a Lipschitz function, and let $T_g : G/\Gamma \to G/\Gamma$ denote the action of g on G/Γ. Then we have the estimate

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{F(T_g^n x)} \ll_{A, G/\Gamma} \|F\|_{\mathrm{Lip}} \log^{-A} N \tag{1.6}$$

for all $N > 1$, uniformly in $g \in G$ and $x \in G/\Gamma$.

The sequence $n \mapsto F(T_g^n x)$ is called a 1-*step nilsequence*. If we specialise to the *circle nilflow* case

$$G := \left(\begin{smallmatrix} 1 & \mathbb{R} \\ 0 & 1 \end{smallmatrix}\right) := \left\{\left(\begin{smallmatrix} 1 & x \\ 0 & 1 \end{smallmatrix}\right) : x \in \mathbb{R}\right\};$$
$$\Gamma := \left(\begin{smallmatrix} 1 & \mathbb{Z} \\ 0 & 1 \end{smallmatrix}\right) := \left\{\left(\begin{smallmatrix} 1 & n \\ 0 & 1 \end{smallmatrix}\right) : n \in \mathbb{Z}\right\}$$

then G/Γ is isomorphic to the unit circle $\mathbb{R}/\mathbb{Z}$, and if we identify a real number α with the group element $\left(\begin{smallmatrix} 1 & \alpha \\ 0 & 1 \end{smallmatrix}\right)$, then $T_\alpha : \mathbb{R}/\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is just the shift $x \mapsto x + \alpha \pmod 1$. Using the standard character $e : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ as the Lipschitz function F, one then sees that (1.3) is a special case of (1.6). In fact, the two examples are more-or-less equivalent, as we shall see in §6 where (1.6) will be established.

(2) The Lipschitz class is a convenient regularity class for us to use; it is smooth enough that one approximate uniformly and quantitatively by trigonometric series (see Lemma A.9), yet rough enough that one can easily extend a function in this class from a small domain to a larger domain (see Lemma A.8). Also, the Lipschitz class is meaningful in both discrete and continuous settings. Of course, the results of this paper also hold in smoother classes such as C^∞, and qualitative versions of these results (with decay factors such as $\log^{-A} N$ replaced by $o(1)$) hold for rougher classes such as the continuous class C^0, or even piecewise continuous classes, by standard limiting arguments.

The main aim of this paper is to generalise (1.6) to cover 2-step nilsequences. In the companion paper [10] to this paper, we shall show how such estimates can be used to prove various "complexity 2" estimates for the Möbius and von Mangoldt functions.

Before stating our main result, we give the definition of s-step nilsequences in general, followed by some examples.

DEFINITION 1.2 (Nilmanifolds and nilsequences). — *Let G be a connected, simply connected, Lie group. We define the* central series $G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots$ *by defining $G_0 = G_1 = G$, and $G_{i+1} = [G, G_i]$ for $i \geqslant 2$, where the commutator group $[G, G_i]$ is the group generated by $\{ghg^{-1}h^{-1} : g \in G, h \in G_i\}$. We say that G is s-*step nilpotent *if $G_{s+1} = 1$. Let $\Gamma \subseteq G$ be a discrete, cocompact subgroup. Then the quotient G/Γ is called an s-step nilmanifold. If $g \in G$ then g acts on G/Γ by left multiplication, $x \mapsto gx$. By a* (basic) s-step nilsequence, *we mean a sequence of the form $(F(T_g^n \cdot x))_{n\in\mathbb{N}}$, where $x \in G/\Gamma$ is a point, $F : G/\Gamma \to \mathbb{C}$ is a continuous function and $T_g : G/\Gamma \to G/\Gamma$ is left multiplication by g. We say that the nilsequence is* bounded *if $|F|$ takes values in $[-1, 1]$. We may (arbitrarily) endow G/Γ with a smooth Riemannian metric $d_{G/\Gamma}$. If the function F is Lipschitz with respect to this metric, we shall refer to the nilsequence $(F(T_g^n \cdot x))_{n\in\mathbb{N}}$ as Lipschitz.*

Remark. — In this paper we will usually suppress explicit mention of the metric $d_{G/\Gamma}$. Whenever an estimate is said to depend on a nilmanifold G/Γ, it should be assumed that it also depends on the choice of metric. See [10] for a more detailed discussion.

Clearly every 1-step nilsequence is a 2-step nilsequence. The next simplest example of nilsequences are quadratic phases.

Example 5 (The Heisenberg nilflow, I). — Consider the example[(3)]

$$G := \begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}; \quad \Gamma := \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}.$$

Then G/Γ is a 2-step nilmanifold. Apart from a set of zero measure, G/Γ may be identified with the fundamental domain

$$\mathcal{F} := \left\{ \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} : -1/2 < x, y, z \leqslant 1/2 \right\}$$

(3) For more detail on the Heisenberg nilflow, Appendix B may be consulted. One can also generate quadratic phase sequences such as $e(n^2\theta)$ using the slightly simpler skew shift nilflow (see *e.g.* [11, Example 12.3]), but we shall refrain from doing so here as the underlying Lie group is disconnected and thus does not quite fall within the framework of Definition 1.2.

using the easily-verified fact that

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \equiv \begin{pmatrix} 1 & \{x\} & \{y-x[z]\} \\ 0 & 1 & \{z\} \\ 0 & 0 & 1 \end{pmatrix} \pmod{\Gamma}.$$

Here, $\{x\}$ refers to the fractional part of x lying in the interval $(-1/2, 1/2]$ and $[x] := x - \{x\}$. Writing

$$g := \begin{pmatrix} 1 & -\theta & -\theta \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix},$$

where $\theta \in \mathbb{R}$, one may check that

$$g^n \equiv \begin{pmatrix} 1 & \{-n\theta\} & \{n^2\theta\} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \pmod{\Gamma}.$$

Thus we see how functions with "quadratic" behaviour arise from 2-step nilsequences. The rather natural function $e(n^2\theta)$ does not quite arise as a Lipschitz nilsequence on the 3×3 Heisenberg group, since the function

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \mapsto e(y)$$

on $\mathcal{F}$ does not extend to a continuous function on G/Γ. The situation may be remedied by splitting $e(n^2\theta)$ as the sum of (say) 10 functions $\chi(\{n\theta\})e(n^2\theta)$ where χ is a Lipschitz cutoff supported on an interval of width $1/5$. Each of the 100 functions

$$\begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \mapsto \chi(x)\chi'(z)e(y)$$

does extend to a Lipschitz function on G/Γ. By taking products one may realize $e(n^2\theta)$ as a Lipschitz nilsequence on the 2-step nilmanifold $(G/\Gamma)^{100}$.

In view of the previous example and our general intent in this paper, it is natural to ask for the estimate

$$\mathbb{E}_{n\in[N]}\mu(n)e(-\alpha n^2 - \beta n - \gamma) \ll_A \log^{-A} N, \tag{1.7}$$

with an implied constant independent of α, β and γ. We will prove such an estimate in §7. Like (1.3), this bound is a fairly standard application of Vinogradov's version of the Hardy-Littlewood method, though somewhat more complicated due to the need to estimate quadratic exponential sums rather than just linear exponential sums. The proof of it has much in common with techniques pioneered by Hua [14] and Vinogradov [23] in connection with the Goldbach-Waring problem. It should be thought of as a warm up for the main business of the paper.

As we have already mentioned, in §6 we shall see that orthogonality to linear phases is more-or-less equivalent to orthogonality to 1-step nilsequences. However, orthogonality to quadratic phases is significantly weaker

than orthogonality to 2-step nilsequences. This is because there are examples of 2-step nilsequences which do not look much like quadratic phases.

Example 6 (The Heisenberg flow, II). — We repeat the analysis of the previous example, but with a less restrictive choice of g. Take

$$g := \begin{pmatrix} 1 & \alpha & \beta \\ 0 & 1 & \gamma \\ 0 & 0 & 1 \end{pmatrix}.$$

A simple induction confirms that

$$g^n \cdot \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} 1 & x+n\alpha & y+n\beta+\frac{1}{2}n(n+1)\alpha \\ 0 & 1 & z+n\gamma \\ 0 & 0 & 1 \end{pmatrix}$$

When reduced to lie in the fundamental domain $\mathcal{F}$, one can end up with functions taking the form $[n\alpha]n\gamma$ (and related forms). These functions are known as *generalised quadratics*, and they capture the spirit of 2-step nilsequences much more completely than genuine quadratic functions do. By repeating the tricks mentioned in the previous example one may actually approximate $e(-[n\sqrt{2}]n\sqrt{3})$ (say) outside of sets of arbitrarily small density as a Lipschitz nilsequence on some product of several copies of the Heisenberg example.

The previous two examples give some idea of what a 2-step nilsequence looks like. Our main result in this paper is that the Möbius function is strongly asymptotically orthogonal to all such functions. This estimate is the case $s = 2$ of the Möbius and Nilsequences Conjecture MN(s): see [10, §6] for further discussion.

Main Theorem (MN(2) conjecture). — *Suppose that G/Γ is a 2-step nilmanifold, and that $F : G/\Gamma \to \mathbb{C}$ is a Lipschitz function. Then for every $A > 0$ we have the estimate*

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{F(T_g^n x)} \ll_{A,G/\Gamma} \|F\|_{\mathrm{Lip}} \log^{-A} N \tag{1.8}$$

uniformly in $g \in G$ and $x \in G/\Gamma$.

Remark. — We conjecture that MN(s) holds for arbitrary s, that is to say there is an analogue of the Main Theorem for s-step nilmanifolds for any $s \geqslant 1$. The fact that the bound (1.8) is uniform in x is unsurprising (since G/Γ is compact), as is the uniformity among all F with fixed Lipschitz norm (thanks to the Arzelà-Ascoli theorem). The uniformity in g is less trivial, and is quite important for applications.

We shall prove the Main Theorem as a consequence of a similar result, Theorem 2.2 below, in which the notion of a 2-step nilsequence is replaced by a more technical type of sequence (a 1-step nilsequence twisted by a

locally quadratic phase) that is more tractable for analysis. The proof of Theorem 2.2 is by far the most difficult portion of the paper and will occupy §3– §12. In comparison, the deduction of the Main Theorem from Theorem 2.2 is more standard and is performed in §2 and Appendix B.

The estimate (1.7), as well as estimates for generalised quadratic phases such as

$$\mathbb{E}_{n\in[N]}\mu(n)e(-[n\sqrt{2}]n\sqrt{3}) = o(1),$$

are consequences of our main theorem.

Remark. — The main result of this paper can then be combined with the *Gowers Inverse Theorem* from [11] to obtain a number of new correlation estimates for the Möbius function, such as

$$\mathbb{E}_{x,d\in[N]}\mu(x)\mu(x+d)\mu(x+2d)\mu(x+3d) = o_{N\to\infty}(1)$$

and

$$\mathbb{E}_{x,h_1,h_2,h_3\in[N]}\mu(x)\mu(x+h_1)\mu(x+h_2)\mu(x+h_3)\mu(x+h_1+h_2)$$
$$\mu(x+h_1+h_3)\mu(x+h_2+h_3)\mu(x+h_1+h_2+h_3) = o_{N\to\infty}(1)$$

(compare with (1.4), (1.5)). It can also be used (with some additional effort) to establish an asymptotic for expressions such as

$$\mathbb{E}_{x,d\in[N]}\Lambda(x)\Lambda(x+d)\Lambda(x+2d)\Lambda(x+3d)$$

as $N\to\infty$, thus enabling one to count the quadruples of number of primes $p_1 < p_2 < p_3 < p_4 \leqslant N$ in arithmetic progression up to a fixed level N. We defer all of these applications to the companion paper [10].

2. A technical reduction

In this section we present a technical counterpart of the Main Theorem, namely Theorem 2.2 below, in which the 2-step nilsequence is replaced by a more analytically tractable object, namely a 1-step nilsequence twisted by a locally quadratic phase. We then discuss how this result implies the Main Theorem. The proof of Theorem 2.2 will then occupy the rest of the paper (except for the Appendices). We first need some notation.

Definition 2.1 (Locally polynomial phases). — *Let $S\subset\mathbb{Z}$ be a set of integers, and let $d\geqslant 0$. A phase function $\phi : S\to\mathbb{R}/\mathbb{Z}$ is said to be* locally

degree d *on* S *if whenever* $n, h_1, \dots, h_{d+1}$ *are such that the* 2^{d+1} *quantities* $n + \epsilon_1 h_1 + \dots + \epsilon_{d+1} h_{d+1}$, $\epsilon_i \in \{0,1\}$ *lie in the set* S, *we have*

$$\sum_{\epsilon \in \{0,1\}^{d+1}} (-1)^{\epsilon_1 + \dots + \epsilon_{d+1}} \phi(n + \epsilon_1 h_1 + \dots + \epsilon_{d+1} h_{d+1}) = 0. \tag{2.1}$$

We refer to phases of local degree 1 *as* locally linear, *phases of local degree* 2 *as* locally quadratic, *and so forth.*

Examples. — Constant phases have local degree 0, while linear phases $\phi(n) := \alpha n$ for $\alpha \in \mathbb{R}$ have local degree 1. If α, β, γ are real numbers, then the phase $\phi(n) := \alpha n^2 + \beta n + \gamma \pmod 1$ is globally quadratic (*i.e.* quadratic on all of $\mathbb{Z}$). The phase $\phi(n) := \{\alpha n\}\{\beta n\}\gamma \pmod 1$ is not globally quadratic, but it is locally quadratic on the *Bohr set* $S := \{n \in \mathbb{Z} : |\{\alpha n\}|, |\{\beta n\}| \leqslant 0.1\}$, which is a set of positive density in $\mathbb{Z}$. The phase $\phi(n) := \{\alpha n\}\gamma \pmod 1$ is locally linear on the same set.

THEOREM 2.2 (μ is strongly orthogonal to local quadratics). — *Let* G/Γ *be a* 1*-step nilmanifold, let* $F : G/\Gamma \to \mathbb{C}$ *be a Lipschitz function, and let* $g \in G$ *and* $x \in G/\Gamma$ *be arbitrary. Let* $\phi : B_N \to \mathbb{R}/\mathbb{Z}$ *be a phase which is locally quadratic on the* Bohr set[(4)] $B_N := \{n \in [N] : F(T_g^n x) \neq 0\}$. *Then we have*

$$\mathbb{E}_{n \in [N]} \mu(n) \overline{F(T_g^n x)} e(-\phi(n)) \ll_{G/\Gamma, A} \|F\|_{\mathrm{Lip}} \log^{-A} N.$$

The proof of Theorem 2.2 is rather lengthy. Let us assume it for now and deduce the Main Theorem. The main proposition in achieving this deduction is

PROPOSITION 2.3 (2-step nilsequences as averages of twisted 1-step nilsequences). — *Let* G/Γ *be a* 2*-step nilmanifold and let* $0 < \varepsilon < 1/2$. *Let* $F : G/\Gamma \to \mathbb{C}$ *be a Lipschitz function with* $\|F\|_{\mathrm{Lip}} \leqslant 1$, *and let* $g \in G$ *and* $x \in G/\Gamma$ *be arbitrary. Then there exists a* 1*-step nilmanifold* $\widetilde{G}/\widetilde{\Gamma}$ *depending only on* G/Γ *and a decomposition*

$$F(T_g^n x) = \mathbb{E}_{i \in I} w_i F_i(T_{g_i}^n x_i) e(-\phi_i(n)) + O(\varepsilon) \tag{2.2}$$

where

- I *is a finite index set;*
- *For each* $i \in I$ *the* w_i *are complex numbers with* $\mathbb{E}_{i \in I} |w_i| \ll \varepsilon^{-O_{G/\Gamma}(1)}$;
- $F_i : \widetilde{G}/\widetilde{\Gamma} \to \mathbb{C}$ *is Lipschitz with norm* $O_{G/\Gamma}(1)$;

(4) This definition of a Bohr set is not quite identical to other Bohr sets in the literature, for instance in [11], but it is very closely related; see the proof of Lemma 11.4.

- $g_i \in \widetilde{G}$;
- $x_i \in \widetilde{G}/\widetilde{\Gamma}$;
- $\phi_i : B_i \to \mathbb{R}/\mathbb{Z}$ *is a phase function which is locally quadratic on the* generalised Bohr set $B_i := \{n \in [N] : F_i(T_{g_i}^n x_i) \neq 0\}$.

We have a proof of a generalisation of this proposition to k-step nilsequences (they are averages of twisted $(k-1)$-step nilsequences). This proceeds using some rather algebraic considerations involving "Kost-Kra cube groups" associated to the nilmanifold G/Γ.

In this paper we present a more computational approach involving so called Mal'cev bases [5, 16]. This approach is completely explicit when the group G is a product of Heisenberg groups $\left(\begin{smallmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{smallmatrix}\right)$. The reader will find remarks in [10] explaining that, in the theory of linear systems of complexity 2 (such as four-term APs) only examples of this type need be considered.

The use of bases may seem overly explicit to some, but it should be noted that Mal'cev bases are in fact required to prove certain foundational topological properties of nilmanifolds. Those results are needed for the approach, just alluded to, that is taken in [10, Appendix E].

The proof of Proposition 2.3 may be found in Appendix B. Assuming it and Theorem 2.2, we can now derive the Main Theorem as follows.

Proof of the Main Theorem assuming Theorem 2.2 and Proposition 2.3. Let G/Γ, F, A be as in the Main Theorem. By renormalising we may assume that $\|f\|_{\mathrm{Lip}} \leqslant 1$. We apply Proposition 2.3 with $\varepsilon := \log^{-A} N$ and obtain a decomposition (2.2). Taking inner products with μ, we obtain

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{F(T_g^n x)} \ll \mathbb{E}_{i\in I}|w_i|\mathbb{E}_{n\in[N]}\mu(n)\overline{F_i(T_{g_i}^n x_i)}e(-\phi_i(n)) + \log^{-A} N.$$

Applying Theorem 2.2, we conclude that

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{F(T_g^n x)} \ll_{A',\widetilde{G}/\widetilde{\Gamma}} \mathbb{E}_{i\in I}|w_i| \log^{-A'} N + \log^{-A} N$$

for any A'. But $\mathbb{E}_{i\in I}|w_i| \ll (\log^A N)^{O_{G/\Gamma}(1)}$, so the claim follows by taking A' suitably large. □

Remark. — Conversely it is also possible to deduce Theorem 2.2 from the Main Theorem by obtaining a suitable converse to Proposition 2.3 (*cf.* the proof of [11, Theorem 12.8]), but we will not do so here.

3. Orthogonality to periodic functions

We now begin the proof of Theorem 2.2, which is the heart of this paper. (The other major component of the paper is the proof of Proposition 2.3 in

Appendix B. This can mostly be read independently of the part of the paper concerned with Theorem 2.2, though it will utilise the harmonic analysis tools collected in Appendix A.)

Our strategy in proving Theorem 2.2 shall be to establish the strong asymptotic orthogonality of the Möbius function to increasingly large classes of sequences, starting with very simple ones and then moving on to "higher degree" sequences. Let us begin with some generalities on how one can go about proving that μ is orthogonal to some function F. There are essentially two complementary methods for doing this. The first, which will feature prominently in this section, is appropriate when F is multiplicative, for example $F = 1$ or $F = \chi$, where χ is some Dirichlet character to the modulus q. Then one may relate $\mathbb{E}_{n\in[N]}\mu(n)F(n)$ via Perron's Formula to zeros of L-functions such as $\zeta(s)$ and $L(s,\chi)$ in the critical strip, the orthogonality coming from the non-existence of zeros close to $\Re s = 1$. Siegel's theorem, concerning a possible zero near $s = 1$ when χ is real, is of particular importance. It implies the bound (1.2), which we recall now:

PROPOSITION 3.1. — *For any $A > 0$ we have*

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{\chi(n)} \ll_A q^{1/2}\log^{-A} N \tag{3.1}$$

for all Dirichlet characters χ to modulus q.

Remark. — For the proof, see [15, Prop. 5.29]. As noted in [15, p. 124] there are difficulties involved in applying the standard Perron's formula approach to $\mathbb{E}_{n\in[N]}\mu(n)\chi(n)$ directly, and it is rather easier to first obtain bounds on $\mathbb{E}_{n\in[N]}\Lambda(n)\chi(n)$. Note that the bound is only non-trivial when the period q is very small (*e.g.* $O(\log^A N)$) compared to N. If one assumed GRH then one could improve the logarithmic decay here to a polynomial decay, which would of course lead to improvements in the other bounds in this paper.

As we will see later in this section, the need to consider zeros of L-functions also appears when dealing with functions F which are not quite multiplicative. For example, they must play a role in the case $F(n) = e(an/q)$, since any Dirichlet character to modulus q is a linear combination of a few such functions F.

At the other end of the spectrum one has functions F which are far from multiplicative, such as $F(n) = e(n\sqrt{2})$. For these functions a completely different method, due originally to Vinogradov, may be brought to bear. The sum $\mathbb{E}_{n\in[N]}\mu(n)F(n)$ is decomposed into so-called *Type I* and *Type II* sums, which can be estimated without having to understand the oscillation

of μ. Provided F is not close to being multiplicative, those sums can often be shown to be small by (effective) harmonic analysis methods. We will discuss this method, in a modern and very neat incarnation due to Vaughan, in §4.

We now begin the proof of Theorem 2.2 by establishing the asymptotic orthogonality of the Möbius function to periodic sequences, which can be viewed in some sense as "0-step nilsequences", and which will be needed to handle the "major arc" case when moving on to linear phases. More precisely, we show

PROPOSITION 3.2 (Möbius is orthogonal to periodic sequences). — *Let $f : \mathbb{N} \to \mathbb{C}$ be a sequence bounded in magnitude by 1 which is periodic of some period $q \geqslant 1$. Then we have*

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{f(n)} \ll_A q\log^{-A} N$$

for all $A > 0$, where the implied constant is ineffective.

Proof. — We first establish the estimate under the additional assumption that $f(n)$ vanishes whenever $(n, q) \neq 1$. Then f can be viewed as a function on the multiplicative group $(\mathbb{Z}/q\mathbb{Z})^\times$, and thus has a Fourier expansion

$$f(n) = \sum_\chi \hat{f}(\chi)\chi(n), \text{ where } \hat{f}(\chi) := \mathbb{E}_{n\in(\mathbb{Z}/q\mathbb{Z})^\times} f(n)\overline{\chi(n)},$$

with χ ranging over all the characters on $(\mathbb{Z}/q\mathbb{Z})^\times$. Applying Proposition 3.1 and the triangle inequality, we conclude

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{f(n)} \ll_A q^{1/2}\log^{-A} N\big(\sum_\chi |\hat{f}(\chi)|\big).$$

But from Cauchy-Schwarz and Plancherel we have

$$\begin{aligned}\sum_\chi |\hat{f}(\chi)| \leqslant \phi(q)^{1/2}\Big(\sum_\chi |\hat{f}(\chi)|^2\Big)^{1/2} &= \phi(q)^{1/2}(\mathbb{E}_{n\in(\mathbb{Z}/q\mathbb{Z})^\times}|f(n)|^2)^{1/2}\\ &= O(\phi(q)^{1/2}),\end{aligned}$$

where $\phi(q) := |(\mathbb{Z}/q\mathbb{Z})^\times|$ is the Euler totient function. Since $\phi(q) \leqslant q$, the claim follows.

Now we consider the general case, in which (n, q) is not necessarily equal to 1 on the support of f. Observe that if $\mu(n)$ is non-zero, then n is square-free, and we can split $n = dm$, where $d = (n, q)$ is square-free (so $\mu^2(d) = 1$) and m is coprime to q. Furthermore we have $\mu(n) = \mu(d)\mu(m)$. We thus obtain the decomposition

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{f(n)} = \frac{1}{N}\sum_{d|q;\mu^2(d)=1}\mu(d)\sum_{1\leqslant m\leqslant N/d}\mu(m)\overline{f(dm)}1_{(m,q)=1}. \tag{3.2}$$

The sequence $m \mapsto f(dm)1_{(m,q)=1}$ is periodic of period q/d and vanishes whenever $(m, q/d) \neq 1$, hence by the preceding arguments

$$\sum_{1 \leqslant m \leqslant N/d} \mu(m)\overline{f(dm)}1_{(m,q)=1} \ll_A \frac{Nq}{d^2} \log^{-A} N.$$

Thus from (3.2) we have

$$\mathbb{E}_{n \in [N]} \mu(n)\overline{f(n)} \ll_A q \log^{-A} N \sum_{d|q} \frac{1}{d^2} \ll q \log^{-A} N,$$

concluding the proof of Proposition 3.2. □

4. Vaughan's identity

In this section we discuss Vinogradov's method for proving that the Möbius function μ is orthogonal to a function $F : \mathbb{N} \to \mathbb{C}$. As we remarked in §3, this involves a decomposition of $\mathbb{E}_{n \in [N]} \mu(n)F(n)$ into Type I and Type II sums. The first argument of this type was due to Vinogradov (who worked with the von Mangoldt function Λ instead of μ). We will use a particularly simple identity due to Vaughan [22] to effect our decomposition into Type I and II sums. See [15, Chapter 13] for a nice discussion of this and related identities.

Let us begin with a few preliminary remarks on our strategy for dealing with Vinogradov's method throughout the paper. The normal method for proving Davenport's estimate (1.3) would be to divide all $\alpha \in \mathbb{R}/\mathbb{Z}$ into two classes: the *major arcs*, where $\alpha \approx a/q$ for some reasonably small q, and the *minor arcs*, consisting of all other α. If α lies in a major arc then one would use Proposition 3.2 to estimate $\mathbb{E}_{n \in [N]} \mu(n)e(\alpha n)$. If, by contrast, α lies in a minor arc then one would establish that Type I and II sums involving $f(n) = e(\alpha n)$ are small (see below for an explanation of what this means). Vaughan [21, Chapter 3] may be consulted for details.

We will adopt what we call an "inverse" strategy. In §5 we will provide a proof of Davenport's estimate. There we will *assume* that either a Type I or a Type II sum involving $f(n) = e(\alpha n)$ is large, and then *deduce* that α lies in a major arc. The distinction between our argument and the standard one may seem rather unimportant, and indeed the two proofs are logically equivalent. However when it comes to dealing with more complicated functions f, such as locally quadratic phases which arise from the consideration of 2-step nilsequences, the inverse strategy is very helpful. There it is much less obvious what one should mean by a "major arc", and even once the

definition is made it is not obvious how to handle it in the context of Type I and II sums.

In light of Lemma A.7, it suffices to establish decay estimates for $\mathbb{E}_{N<n\leqslant 2N}\mu(n)f(n)$. The next lemma gives Vaughan's decomposition of sums of this kind.

LEMMA 4.1 (Vaughan's identity). — *Let U, V, N be positive integers with $UV \leqslant N$, and $f : \mathbb{N} \to \mathbb{C}$ be a sequence. Then we have*

$$\mathbb{E}_{N<n\leqslant 2N}\mu(n)\overline{f(n)} = -\mathrm{T}_{\mathrm{I}} + \mathrm{T}_{\mathrm{II}} \tag{4.1}$$

where T_{I} is the Type I *expression*

$$\mathrm{T}_{\mathrm{I}} := \frac{1}{N}\sum_{1\leqslant d\leqslant UV} a_d \sum_{N/d<w\leqslant 2N/d} \overline{f(dw)} \tag{4.2}$$

in which

$$a_d := \sum_{bc=d:b\leqslant U,c\leqslant V} \mu(b)\mu(c),$$

and T_{II} is the Type II *expression*

$$\mathrm{T}_{\mathrm{II}} := \frac{1}{N}\sum_{V<d\leqslant 2N/U}\ \sum_{\max(U,N/d)<w\leqslant 2N/d} \mu(w)b_d\overline{f(dw)} \tag{4.3}$$

in which

$$b_d := \sum_{c|d:c>V} \mu(c).$$

Remark. — One of the key points in the analysis of Type I sums is that the precise form of the coefficients a_d is almost completely irrelevant: we will apply the Cauchy-Schwarz inequality, and so only the mean square size of these coefficients will concern us. The same is true in the analysis of Type II sums. In this case it is the coefficients $\mu(w)$ and b_d which get removed by the Cauchy-Schwarz inequality.

There is considerable flexibility in the choice of the parameters U and V. We will take $U = V = N^{1/3}$ in our applications.

Proof. — We follow [15, §13.4–5]. Observe that for any positive integer n we have

$$\mu(n) = \sum_{b,c:bc|n} \mu(b)\mu(c).$$

We split the range of the sum over b, c into four ranges: (i) $b \leqslant U$, $c \leqslant V$; (ii) $b > U$, $c \leqslant V$; (iii) $b \leqslant U$, $c > V$ and (iv) $b > U$, $c > V$. Denoting

the associated sums $\Sigma_1, \dots, \Sigma_4$, it is easy to check that $\Sigma_2 = \Sigma_3 = -\Sigma_1$. It follows that

$$\mu(n) = -\Sigma_1 + \Sigma_4 = -\sum_{\substack{b \leqslant U;\ c \leqslant V \\ bc|n}} \mu(b)\mu(c) + \sum_{\substack{b > U;\ c > V \\ bc|n}} \mu(b)\mu(c).$$

Multiplying by $\overline{f(n)}$ and summing over $N < n \leqslant 2N$, we have Vaughan's identity:

$$\begin{aligned}\mathbb{E}_{N<n\leqslant 2N}\mu(n)\overline{f(n)} &= -\mathbb{E}_{N<n\leqslant 2N}\sum_{\substack{b\leqslant U;\ c\leqslant V\\ bc|n}}\mu(b)\mu(c)\overline{f(n)}\\ &\qquad + \mathbb{E}_{N<n\leqslant 2N}\sum_{\substack{b> U;\ c> V\\ bc|n}}\mu(b)\mu(c)\overline{f(n)}\\ &:= -\mathrm{T}_{\mathrm{I}} + \mathrm{T}_{\mathrm{II}}\,.\end{aligned}$$

It is an easy matter to confirm that T_{I} may be written in the form (4.2), after making the substitution $d = bc$ and $n = dw$. One may also check that T_{II} may be written in the form (4.3) after making the substitution $w = b$ and $n = dw$. □

Vaughan's identity tells us that if $\mathbb{E}_{N<n\leqslant 2N}\mu(n)\overline{f(n)}$ is large then either T_{I} or T_{II} is large. The next proposition shows how this information is processed, by using the Cauchy-Schwarz inequality to eliminate the parameters a_d, b_w and $\mu(w)$, leaving behind estimates which only involve the explicit function f.

PROPOSITION 4.2 (Inverse theorem for $\mathbb{E}_{N<n\leqslant 2N}\mu(n)\overline{f(n)}$). — *Let U, V, N be positive integers with $UV \leqslant N$, and let $f : \mathbb{N} \to \mathbb{C}$ be a function with $\|f\|_\infty = O(1)$ such that*

$$|\mathbb{E}_{N<n\leqslant 2N}\mu(n)\overline{f(n)}| \geqslant \delta$$

for some $\delta > 0$. Then one of the following statements holds:

- (Type I sum is large): *There exists an integer $1 \leqslant D \leqslant UV$ such that*

$$|\mathbb{E}_{N/d<w\leqslant 2N/d} f(dw)| \gg \delta \log^{-5/2} N \tag{4.4}$$

for $\gg \delta^2 D \log^{-5} N$ integers d such that $D < d \leqslant 2D$.

- (Type II sum is large): *There exist integers D, W with $V/2 \leqslant D \leqslant 4N/U$ and $N/4 \leqslant DW \leqslant 4N$, such that*

$$|\mathbb{E}_{D<d,d'\leqslant 2D}\mathbb{E}_{W<w,w'\leqslant 2W} f(dw)\overline{f(d'w)f(dw')}f(d'w')| \gg \delta^4 \log^{-14} N. \tag{4.5}$$

Remark. — The estimate (4.4) may be viewed as an assertion that f behaves periodically, while (4.5) is an assertion that f behaves multiplicatively. The numerical exponents could probably be improved slightly here, but we will not need such refinements here (especially since our bounds will eventually become ineffective anyway).

Proof. — We may of course take N to be large. Applying Lemma 4.1, we see that either $|\,\mathrm{T_I}\,| \geqslant \delta/2$ or $|\,\mathrm{T_{II}}\,| \geqslant \delta/2$.

Suppose first that the Type I expression is large, that is to say $|\,\mathrm{T_I}\,| \geqslant \delta/2$ where $\mathrm{T_I}$ is given by (4.2). Using the crude bound $|a_d| \leqslant \tau(d)$, where $\tau(d) := \sum_{b|d} 1$ is the divisor function, we have

$$\sum_{1 \leqslant d \leqslant UV} \frac{\tau(d)}{d} \big|\mathbb{E}_{N/d < w \leqslant 2N/d} f(dw)\big| \gg \delta.$$

By Cauchy-Schwarz inequality this implies that

$$\sum_{1 \leqslant d \leqslant UV} \frac{1}{d} \big|\mathbb{E}_{N/d < w \leqslant 2N/d} f(dw)\big|^2 \gg \delta^2 \Big(\sum_{1 \leqslant d \leqslant UV} \frac{\tau^2(d)}{d} \Big)^{-1}.$$

Invoking the divisor moment estimate (C.1), it follows that

$$\sum_{1 \leqslant d \leqslant UV} \frac{1}{d} \big|\mathbb{E}_{N/d < w \leqslant 2N/d} f(dw)\big|^2 \gg \delta^2 \log^{-4} N.$$

Dividing the region $1 \leqslant d \leqslant UV$ into dyadic blocks $D < d \leqslant 2D$ (allowing for some slight overlap) and applying the pigeonhole principle we obtain

$$\sum_{D < d \leqslant 2D} \big|\mathbb{E}_{N/d < w \leqslant 2N/d} f(dw)\big|^2 \gg \delta^2 D \log^{-5} N$$

for some D, $1 \leqslant D \leqslant UV$. Since the summand is bounded by $O(1)$, a simple averaging argument confirms that $\big|\mathbb{E}_{N/d < w \leqslant 2N/d} f(dw)\big| \gg \delta \log^{-5/2} N$ for at least $\gg \delta^{-2} D \log^{-5} N$ values of d, which is what we wanted to prove.

Now suppose instead that the Type II expression is large, that is $|\,\mathrm{T_{II}}\,| \geqslant \delta/2$. Using the evident bound $|b_d| \leqslant \tau(d)$, we conclude

$$\sum_{V < d \leqslant 2N/U} \tau(d) \Big| \sum_{N/d < w \leqslant 2N/d} 1_{w > U} \mu(w) f(dw) \Big| \gg N\delta.$$

Applying Cauchy-Schwarz and the divisor moment estimate (C.1) once again, we conclude that

$$\sum_{V < d \leqslant 2N/U} d \Big| \sum_{N/d < w \leqslant 2N/d} 1_{w > U} \mu(w) f(dw) \Big|^2 \gg N^2 \delta^2 \log^{-4} N.$$

By dyadic decomposition, we thus can find integers D, W with $V/2 \leqslant D \leqslant 4N/U$ and $N/4 \leqslant DW \leqslant 4N$ such that

$$\sum_{D<d\leqslant 2D} \Big| \sum_{W<w\leqslant 2W} 1_{I_d}(w)\mu(w)f(dw)\Big|^2 \gg \frac{N^2}{D}\delta^2 \log^{-5} N,$$

where I_d is the discrete interval $\{w > U : N/d < w \leqslant 2N/d\}$. Applying Lemma A.2 to remove the cutoff $1_{I_d}(w)$, we obtain

$$\sum_{D<d\leqslant 2D} \Big| \sum_{W<w\leqslant 2W} \mu(w)f(dw)e(\alpha w)\Big|^2 \gg N\delta^2 \log^{-7} N.$$

for some $\alpha \in \mathbb{R}/\mathbb{Z}$. Expanding the left-hand side as

$$\sum_{W<w,w'\leqslant 2W} \sum_{D<d\leqslant 2D} (b)(w,w')f(dw)\overline{f(dw')},$$

where we use $(b)()$ to denote a bounded function whose exact form we do not care about (see Appendix A), the required inequality (4.5) follows from the Cauchy-Schwarz inequality in the form of Lemma A.10. □

5. Orthogonality to linear phase functions

As a first application of Proposition 4.2, let us recall the standard proof of Davenport's estimate (1.3). We do this partly for expository reasons, to illustrate the "inverse" approach to dealing with Type I and II sums, and also because we will need (1.3) to treat the "major arc" case of quadratic phases in later sections. As we shall see, the linear case is particularly easy, because the exponential sums can be easily computed (using (A.1)). Here and in the rest of the paper we will be using some standard tools from harmonic analysis, together with the notations $\|x\|_{\mathbb{R}/\mathbb{Z}}$ and $\|x\|_{\mathbb{R}/\mathbb{Z},Q}$, which we summarise in Appendix A.

We begin with a partial result, which is weaker than (1.3) in that it only resolves the theorem for the "minor arc" values of α, but has the advantage of being completely effective, as it does not require any information on Siegel zeroes.

PROPOSITION 5.1 (Correlation with a linear phase implies major arc). *Let $\alpha \in \mathbb{R}$, let $A > 0$, and let N be a large integer such that*

$$\big|\mathbb{E}_{N<n\leqslant 2N}\mu(n)e(-\alpha n)\big| \geqslant \log^{-A} N. \tag{5.1}$$

Then there exists D, $1 \leqslant D \ll N^{2/3}$, such that

$$\#\Big\{1 \leqslant d \leqslant 2D : \|\alpha d\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{D}{N}\log^{4A+14} N\Big\} \gg D\log^{-4A-14} N. \tag{5.2}$$

Proof. — We apply Proposition 4.2 with $U = V = N^{1/3}$ and conclude one of the following statements:

- (Type I sum is large): There exists D, $1 \leqslant D \leqslant N^{2/3}$, such that
$$\big|\mathbb{E}_{N/d<w\leqslant 2N/d} e(\alpha dw)\big| \gg \log^{-A-5/2} N$$
for $\gg D\log^{-2A-5} N$ values of $D < d \leqslant 2D$.

- (Type II sum is large): There exist integers D, W with $N^{1/3} \ll D \ll N^{2/3}$ and $N/8 \leqslant DW \leqslant 8N$ such that
$$\big|\mathbb{E}_{D<d,d'\leqslant 2D}\mathbb{E}_{W<w,w'\leqslant 2W} e(\alpha dw - \alpha d'w - \alpha dw' + \alpha d'w')\big| \gg \log^{-4A-14} N.$$

Suppose first that the Type I sum is large. Applying (A.1) we conclude that there are $\gg D\log^{-2A-5} N$ values of d, $D < d \leqslant 2D$, for which
$$\|\alpha d\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{D}{N}\log^{-A-5/2} N.$$
This implies (5.2) with some room to spare.

Now suppose instead that the Type II sum is large. By the pigeonhole principle we can find d', w' such that
$$\big|\mathbb{E}_{D<d\leqslant 2D}\mathbb{E}_{W<w\leqslant 2W} e(\alpha dw - \alpha d'w - \alpha dw' + \alpha d'w')\big| \gg \log^{-4A-14} N$$
and hence by the triangle inequality
$$\mathbb{E}_{D<d\leqslant 2D}\big|\mathbb{E}_{W<w\leqslant 2W} e(\alpha(d-d')w)\big| \gg \log^{-4A-14} N.$$
Applying (A.1) we obtain
$$\mathbb{E}_{D<d\leqslant 2D}\min\Big(1, \frac{D}{N\|\alpha(d-d')\|_{\mathbb{R}/\mathbb{Z}}}\Big) \gg \log^{-4A-14} N,$$
and thus after a simple averaging argument we establish
$$\#\Big\{D < d \leqslant 2D : \|\alpha d - \alpha d'\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{D}{N}\log^{4A+14} N\Big\} \gg D\log^{-4A-14} N.$$
Substituting $\widetilde{d} := d - d'$, we conclude
$$\#\Big\{-2D \leqslant \widetilde{d} \leqslant 2D : \|\alpha\widetilde{d}\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{D}{N}\log^{4A+14} N\Big\} \gg D\log^{-4A-14} N.$$
Since $D \geqslant N^{1/3}$, we can easily remove the degenerate contribution when $\widetilde{d} = 0$. The claim (5.2) then follows by symmetry. □

The next task is to understand exactly what the condition (5.2) implies. It is clear that it is some sort of "major arc" condition, as it forces α to lie close to a rational number with reasonably small denominator. A naïve inspection of (5.2) would lead one to guess that this denominator is of size D or so; however it turns out that one can reduce the size of the denominator substantially, to be a power of $\log N$. Indeed, we have

COROLLARY 5.2 (Correlation with a linear phase implies major arc, II). *Let $\alpha \in \mathbb{R}$, let $A > 0$ and let N be a large integer such that* (5.1) *holds. Then*

$$\|\alpha\|_{\mathbb{R}/\mathbb{Z},16\log^{8(A+4)} N} \ll \frac{\log^{28(A+4)} N}{N}.$$

The implied constant is effective.

Proof. — We apply Proposition 5.1 to obtain D, $1 \leqslant D \leqslant N^{2/3}$, obeying (5.2). If $D \leqslant \log^{8(A+4)} N$ then the claim follows directly from (5.2). If instead $D \geqslant \log^{8(A+4)} N$, we may apply Lemma A.4(ii) with $I = \{1, \dots, 2D\}$, $\delta_1 \ll \frac{D}{N} \log^{4(A+4)} N$, and $\delta_2 \gg \log^{-4(A+4)} N$ to obtain the claim. □

When α *is* major arc, *i.e.* when $\|\alpha\|_{\mathbb{R}/\mathbb{Z},Q}$ is small, we may proceed using Proposition 3.1.

PROPOSITION 5.3 (Major arc phases are orthogonal to Möbius). — *Let N be a large integer, let α be a real number, and let $Q, K \geqslant 1$ be such that $\|\alpha\|_{\mathbb{R}/\mathbb{Z},Q} \leqslant K/N$. Then we have*

$$|\mathbb{E}_{N<n\leqslant 2N}\mu(n)e(-\alpha n)| \ll_A Q^{1/2}K^{1/2}\log^{-A} N$$

for any $A > 0$ (the implied constant is ineffective).

Proof. — Let $1 \leqslant M < N$ be a parameter to be chosen later. Then by partitioning the interval $\{N < n \leqslant 2N\}$ into intervals of length M, plus a remainder, we conclude that

$$\big|\mathbb{E}_{N<n\leqslant 2N}\mu(n)e(-\alpha n)\big| \leqslant \sup_{|I|=M;I\subset[N,2N]} \Big|\frac{1}{M}\sum_{n\in I}\mu(n)e(\alpha n)\Big| + O\Big(\frac{M}{N}\Big).$$

By hypothesis, we have integers a and $1 \leqslant q \leqslant Q$ such that $\big|\alpha - \frac{a}{q}\big| \leqslant \frac{K}{N}$. We thus have

$$e(-\alpha n) = e(-an/q)e(-(\alpha - a/q)n) = e(-an/q)e(-(\alpha - a/q)n_I) + O\Big(\frac{KM}{N}\Big)$$

for any $n, n_I \in I$. Discarding the constant phase $e(-(\alpha - a/q)n_I)$, we conclude

$$\big|\mathbb{E}_{N<n\leqslant 2N}\mu(n)e(-\alpha n)\big| \leqslant \sup_{|I|=M;I\subset[N,2N]} \Big|\frac{1}{M}\sum_{n\in I}\mu(n)e(-an/q)\Big| + O\Big(\frac{KM}{N}\Big).$$

Applying Proposition 3.2 (replacing A by $2A$) we have

$$\Big|\frac{1}{M}\sum_{n\in I}\mu(n)e(-an/q)\Big| \ll_A \frac{qN}{M}\log^{-2A} N.$$

Combining these estimates and making the optimal choice

$$M = q^{1/2}K^{-1/2}N\log^{-A} N,$$

we obtain the claim. □

Combining Corollary 5.2 with Proposition 5.3 (and selecting the parameters A appropriately) we conclude the unconditional estimate

$$\big|\mathbb{E}_{N<n\leqslant 2N}\mu(n)e(-\alpha n)\big| \ll_A \log^{-A} N,$$

uniformly in $\alpha \in \mathbb{R}/\mathbb{Z}$ and for any $A > 0$. Here the implied constant is ineffective. Davenport's estimate (1.3) then follows from Lemma A.7 (with $\varphi \equiv 1$), observing that the additional linear phase created by that lemma can be easily absorbed.

6. Orthogonality to linear objects

Our aim in this section is to prove that the Möbius function μ is orthogonal to various functions $f : \mathbb{Z} \to \mathbb{C}$ of "linear" type. We begin by proving (1.6), which asserts that μ is orthogonal to 1-step nilsequences. Then, in Proposition 6.3,we confirm that μ is orthogonal to a certain type of locally linear phase function. This proposition is needed for our later analysis of 2-step nilsequences (indeed, it essentially forms the "major arc" part of that analysis; see §12).

Proof of (1.6). — Let us begin by recalling what it is we are trying to prove. We have an abelian Lie group G and a cocompact discrete subgroup $\Gamma \leqslant G$. Let $F : G/\Gamma \to \mathbb{C}$ be any Lipschitz function. Then we wish to show that

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{F(g^n x)} \ll_{A,G/\Gamma} \|F\|_{\mathrm{Lip}}\log^{-A} N \tag{6.1}$$

uniformly in $g \in G$ and $x \in G/\Gamma$. Now G/Γ is isomorphic to the product of a torus and a finite abelian group, and hence to some subgroup of a torus $(\mathbb{R}/\mathbb{Z})^d$. By Lemma A.8, we may assume that F is defined on all of this torus. Let $0 < \varepsilon < 1$ be arbitrary. By renormalising, we may also assume that $\|F\|_{\mathrm{Lip}} = 1$. By Lemma A.9, we may write

$$F(x) = \sum_{j=1}^{J} c_j e(m_j \cdot x) + O_d(\varepsilon^{1/2})$$

(say), where $c_j = O(1)$ and $J = O_d(\varepsilon^{-d})$. Writing $g = (\alpha_1, \dots, \alpha_d)$, we have

$$F(g^n x) = \sum_{j=1}^{J} c_j e(m_j \cdot x) e\big(n(\alpha_1 m_j^{(1)} + \cdots + \alpha_d m_j^{(d)})\big) + O_d(\varepsilon^{1/2}).$$

Multiplying by μ and taking the expectation over $n \leqslant N$, the contribution of each of the J terms here is $O_A(\log^{-A} N)$ for any $A > 0$, thanks to (1.3). We therefore have

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{F(g^n x)} \ll_{A,d} \varepsilon^{-d}\log^{-A} N + \varepsilon^{1/2}.$$

Optimising this in ε and recalling that $A > 0$ was arbitrary, we obtain the claim. □

Our other goal in this section is to establish, in Proposition 6.3, orthogonality of μ to phase functions which are almost linear on Bohr sets.

DEFINITION 6.1 (Bohr sets). — *Let $N \geqslant 1$. Let G/Γ be a 1-step nilmanifold (i.e. a compact abelian Lie group). Then G/Γ can be embedded as a closed subgroup of a finite-dimensional torus $(\mathbb{R}/\mathbb{Z})^d$, and we let $d_{G/\Gamma}(x, y) := \|xy^{-1}\|_{G/\Gamma}$ be the metric on G/Γ induced from such an embedding (chosen arbitrarily), where we give the torus the metric induced by the l^∞ norm* (A.3). *For any $g \in G$ and any $n \in \mathbb{Z}$, we define the "norm" $\|n\|_g = \|n\|_{g,N}$ for all $n \in \mathbb{Z}$ by the formula*

$$\|n\|_g := \|g^n\|_{G/\Gamma} + \Big|\frac{n}{N}\Big|,$$

and then define the Bohr sets $B_g(n_0, \rho) \subset \mathbb{Z}$ *for any $n_0 \in \mathbb{Z}$ and $\rho > 0$ as*

$$B_g(n_0, \rho) := \big\{n \in \mathbb{Z} : \|n - n_0\|_g < \rho\big\}.$$

Thus we have $B_g(n_0, \rho) = n_0 + B_g(0, \rho)$.

Remarks. — These Bohr sets are closely related to the sets B_N appearing in Theorem 2.2, and also to more "traditional" Bohr sets in the literature; see the proof of Lemma 11.4 below. We observe the sub-homogeneity property $\|nm\|_g \leqslant |n|\|m\|_g$ for all $n, m \in \mathbb{Z}$, with equality $\|nm\|_g = |n|\|m\|_g$ holding whenever $|n|\|m\|_g < c$ for some constant $c_{G/\Gamma} > 0$. We shall use these facts frequently in the sequel without further comment.

Some other easy properties of Bohr sets are contained in the following lemma.

LEMMA 6.2 (Bohr set estimates). — *Let $N \geqslant 1$, let G/Γ be a 1-step nilmanifold, and let $g \in G$. Let $0 < \rho < 1/2$.*

(a) (Lower bound): *We have $|B_g(0, \rho)| \gg_{G/\Gamma} \rho^{-O_{G/\Gamma}(1)} N$.*

(b) (Doubling property): *We have* $|B_g(0,2\rho)| \ll_{G/\Gamma} |B_g(0,\rho)|$.
(c) (Divisibility): *For any integer* $d \geqslant 1$ *we have*

$$\big|\{n \in B_g(0,\rho) : d|n\}\big| \gg_{G/\Gamma} \frac{1}{d}\big|B_g(0,\rho)\big|.$$

Proof. — To obtain (a), we cover G/Γ by $O_{G/\Gamma}(\rho^{-O_{G/\Gamma}(1)})$ balls B of radius $\rho/4$, and also cover $\{1,\dots,N\}$ into intervals I of length $\rho N/4$. By the pigeonhole principle we can find an interval I and a ball B such that $S := \{n : n \in I : g^n \in B\}$ has cardinality $\gg_{G/\Gamma} \rho^{-O_{G/\Gamma}(1)}N$. The claim then follows from the triangle inequality. Indeed if $n, n_0 \in S$ then $|(n - n_0)/N| \leqslant \rho/2$ and $\|g^{n-n_0}\|_{G/\Gamma} \leqslant \rho/2$, and thus $S - n_0 \subseteq B_g(0,\rho)$. It follows that $|B_g(0,\rho)| \geqslant |S|$.

The proof of (b) is very similar. We cover the ball with centre 0 and radius 2ρ in G/Γ by $O_{G/\Gamma}(1)$ balls B of radius $\rho/4$, and the interval $\{1,\dots,\rho N\}$ by $O(1)$ intervals I of length $\rho N/4$. By the pigeonhole principle, there is an interval I and a ball B such that the set $S := \{n \in B_g(0,2\rho) : n \in I : g^n \in B\}$ has cardinality $\gg_{G/\Gamma} |B_g(0,2\rho)|$. Note, however, that if $n, n_0 \in S$ then $|(n-n_0)/N| \leqslant \rho/2$ and $\|g^{n-n_0}\|_{G/\Gamma} \leqslant \rho/2$, and so $S - n_0 \subseteq B_g(0,\rho)$. It follows that $|B_g(0,\rho)| \geqslant |S|$.

Finally, we establish (c). By the pigeonhole principle there is some residue class $X_b := \{x \in \mathbb{Z} : x \equiv b \pmod d\}$ for which $|B_g(0,\rho/2) \cap X_b| \geqslant d^{-1}|B_g(0,\rho/2)|$. Note, however, that if $n, n_0 \in B_g(0,\rho/2) \cap X_b$ then $d|(n - n_0)$ and $n - n_0 \in B_g(0,\rho)$. The result now follows from (b). □

As we have remarked, the next result will form the "major arc" part of our analysis of 2-step nilsequences. It may appear a little technical at this point, but has been designed to cover everything we need in the later application.

Proposition 6.3 (Orthogonality to almost linear phases on Bohr sets). *Let* $N \in \mathbb{N}$ *be large, let* G/Γ *be a 1-step nilmanifold, let* $g \in G$, *let* $\rho \in (0,1)$ *and let* $B_g(n_0,\rho)$ *be some Bohr set contained in* $\{N+1,\dots,2N\}$. *Let* $\psi : \mathbb{Z} \to \mathbb{R}^+$ *be a non-negative function supported on* $B_g(n_0,\rho)$ *which obeys the Lipschitz estimate*

$$|\psi(n) - \psi(m)| \ll \|n - m\|_g \tag{6.2}$$

for all $n, m \in \mathbb{Z}$. *Let* $q \in [1, N/100]$ *be an integer, let* $\varepsilon \in (0,1)$, *and let* $\phi : \mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ *be a phase obeying the approximate linearity estimate*

$$\|\phi(x + h_1 + h_2) - \phi(x + h_1) - \phi(x + h_2) + \phi(x)\|_{\mathbb{R}/\mathbb{Z}} \ll \varepsilon \tag{6.3}$$

whenever $x, x+h_1, x+h_2, x+h_1+h_2 \in B_g(n_0, 10\rho)$ *and* $q|h_1, h_2$. *Then for any* $\kappa \in (0, \rho]$ *we have*

$$\left|\mathbb{E}_{N<n\leqslant 2N}\mu(n)\psi(n)e(-\phi(n))\right| \ll_{A,G/\Gamma} \kappa^{-O_{G/\Gamma}(1)}q^3\log^{-A}N + (\varepsilon+\kappa)\mathbb{E}_{N<n\leqslant 2N}|\psi(n)|$$

for all $A > 0$ (*the constant is ineffective*).

Proof. — We can divide the interval $\{N+1, \dots, 2N\}$ into q residue classes $X_1, \dots, X_q$ modulo q. By the triangle inequality it suffices to show that

$$\left|\mathbb{E}_{N<n\leqslant 2N}\mu(n)1_{X_s}(n)\psi(n)e(-\phi(n))\right| \ll_{A,G/\Gamma} \kappa^{-C}q^2\log^{-A}N + (\varepsilon+\kappa)\mathbb{E}_{N<n\leqslant 2N}|\psi(n)|1_{X_s}(n)$$

for all s, $1 \leqslant s \leqslant q$.

Fix s. Without loss of generality we may assume that $X_s \cap B_g(n_0, \rho)$ is non-empty, thus we may choose $n_s \in X_s \cap B_g(n_0, \rho)$. We work in the group $\mathbb{Z}/p\mathbb{Z}$ where $p \in [10N, 20N]$ is some prime, abusing notation by regarding functions on $[N, 2N]$ as functions on $\mathbb{Z}/p\mathbb{Z}$ in an obvious way. Let $f : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ be the function $f(x) := \psi(x)e(-\phi(x))$, and similarly let $\widetilde{\mu} : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ be the function $\widetilde{\mu}(x) := \mu(x)1_{N<x\leqslant 2N}$. Then our task is to show

$$\mathbb{E}_{x\in\mathbb{Z}/p\mathbb{Z}}\widetilde{\mu}(x)1_{X_s}(n)f(x) \ll_{A,G/\Gamma} \kappa^{-C}q^2\log^{-A}N + (\varepsilon+\kappa)\mathbb{E}_{N<n\leqslant 2N}|\psi(n)|1_{X_s}(n). \tag{6.4}$$

Now let $F : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ be the function defined by

$$F(h) := 1_{q|h}1_{B_g(0,\kappa)}(h)e(\phi(n_s+h)).$$

Observe that if $x \in X_s \cap B_g(n_0, \rho)$ and $h_1, h_2 \in B_g(0, \kappa)$ with $q|h_1, h_2$, then from three applications of (6.3) we have (since $\kappa \leqslant \rho$)

$$\phi(x+h_1) - \phi(x) - \phi(n_s+h_1) + \phi(n_s) = O_{\mathbb{R}/\mathbb{Z}}(\varepsilon)$$
$$\phi(x+h_2) - \phi(x) - \phi(n_s+h_2) + \phi(n_s) = O_{\mathbb{R}/\mathbb{Z}}(\varepsilon)$$

and

$$\phi(x+h_1+h_2) - \phi(x+h_1) - \phi(x+h_2) + \phi(x) = O_{\mathbb{R}/\mathbb{Z}}(\varepsilon),$$

where we use $O_{\mathbb{R}/\mathbb{Z}}(\varepsilon)$ to denote a quantity whose $\|\cdot\|_{\mathbb{R}/\mathbb{Z}}$ norm is $O(\varepsilon)$. Summing these three bounds yields

$$\phi(x) = \phi(x+h_1+h_2) - \phi(n_s+h_1) - \phi(n_s+h_2) + 2\phi(n_s) + O_{\mathbb{R}/\mathbb{Z}}(\varepsilon),$$

which of course implies that

$$e(-\phi(x)) = e(-\phi(x+h_1+h_2))e(\phi(n_s+h_1))e(\phi(n_s+h_2))e(-2\phi(n_s)) + O(\varepsilon).$$

From (6.2), the Lipschitz assumption on ψ, we know that $\psi(x+h_1+h_2) = \psi(x) + O(\kappa)$ for $h_1, h_2 \in B_g(0,\kappa)$. Hence we conclude that

$$f(x) = f(x+h_1+h_2)F(h_1)F(h_2)e(-2\phi(n_s)) + O(\varepsilon+\kappa)$$

for all $x \in \mathbb{Z}/p\mathbb{Z}$ and $h_1, h_2 \in B_g(0,\kappa)$ with $q|h_1, h_2$. Since $|f(x)| \leqslant \psi(x)$ point-wise, we may sum over X_s and deduce that

$$\begin{aligned}\mathbb{E}_{x\in\mathbb{Z}/p\mathbb{Z}}\widetilde{\mu}(x)1_{X_s}(x)f(x) &\\ = \mathbb{E}_{h_1,h_2\in B_g(0,\kappa)q|h_1,h_2}\mathbb{E}_{x\in\mathbb{Z}/p\mathbb{Z}}\widetilde{\mu}(x)f(x+h_1+h_2)&\\ \times F(h_1)F(h_2)e(2\phi(n_s)) + O((\varepsilon+\kappa)\mathbb{E}_{N<n\leqslant 2N}|\psi(n)|1_{X_s}(n)).&\end{aligned}$$

To prove (6.4), then, it suffices to show that

$$\begin{aligned}\mathbb{E}_{h_1,h_2\in B_g(0,\kappa);q|h_1,h_2}\mathbb{E}_{x\in\mathbb{Z}/p\mathbb{Z}}\widetilde{\mu}(x)f(x+h_1+h_2)F(h_1)F(h_2)&\\ \ll_{A,G/\Gamma} q^2\kappa^{-C}\log^{-A}N.&\end{aligned}$$

From Lemma 6.2(a) and (c) we have

$$\#\{h \in B_g(0,\kappa) : q|h\} \gg \frac{1}{q}p\kappa^C,$$

and so it is enough to prove that

$$\mathbb{E}_{h_1,h_2,x\in\mathbb{Z}/p\mathbb{Z}}\widetilde{\mu}(x)f(x+h_1+h_2)F(h_1)F(h_2) \ll_A \log^{-A}N.$$

To demonstrate this we use the Fourier transform[(5)] on $\mathbb{Z}/p\mathbb{Z}$, noting in particular the identity

$$\mathbb{E}_{x,h_1,h_2\in\mathbb{Z}/p\mathbb{Z}}\widetilde{\mu}(x)f(x+h_1+h_2)F(h_1)F(h_2) = \sum_{\xi\in\mathbb{Z}/p\mathbb{Z}}\widehat{\widetilde{\mu}}(\xi)\hat{f}(-\xi)\widehat{F}(\xi)^2.$$

Since f and F are bounded, we see from Plancherel's formula that $|\hat{f}(-\xi)| = O(1)$ and $\sum_{\xi\in\mathbb{Z}/p\mathbb{Z}}|\widehat{F}(\xi)|^2 = O(1)$. Also, from (1.3) we have $\widehat{\widetilde{\mu}}(\xi) \ll_A \log^{-A}N$ for any ξ. The claim follows. □

Remark. — What we have in effect done here is approximate $\psi(n)e(-\phi(n))$ by something akin to a *dual function* coming from the Gowers U^2-norm. By the general theory of this norm we know that any bounded function which is orthogonal to all linear exponentials (*cf.* (1.3)) is orthogonal to all such dual functions. The Fourier argument at the end of the proof of Proposition 6.3 is basically the standard proof of this fact. See [11] for further discussion.

(5) If $g : \mathbb{Z}/p\mathbb{Z} \to \mathbb{C}$ is a function, and if $\xi \in \mathbb{Z}/p\mathbb{Z}$, we write $\hat{g}(\xi) := \mathbb{E}_{x\in\mathbb{Z}/p\mathbb{Z}}g(x)e(-x\xi/p)$.

Remark. — The results of this section may be used to show that μ is orthogonal to various other types of function, which need not be Lipschitz or even continuous, but which are still somehow "approximately linear" in n. Examples of such functions include the bracket-linear phases $e(\beta_1\lfloor\alpha_1 n\rfloor + \cdots + \beta_d\lfloor\alpha_d n\rfloor)$. We omit the details.

7. Orthogonality to quadratic phases

In this section our aim is to prove the estimate (1.7). Strictly speaking, this section is unnecessary, since (1.7) does not represent the heart of the Main Theorem in the same way that (1.3) forms the substance of (1.6). See the introduction for some remarks on this point.

This section is included for two pedagogical reasons. First of all the argument does have some features in common with the (far more complicated) analysis of later sections, and thus introduces the main ideas of those sections in a simplified setting. Secondly, it represents a good opportunity to introduce some notation for inequalities which will be very helpful for the rest of the paper.

The definition of asymptotic orthogonality involves establishing that $X \ll_A \log^{-A} N$, for various quantities X and for all $A > 0$, and it is convenient to have a notation specific to this kind of situation. In each argument that follows, the value of A will be arbitrary, but fixed throughout the argument. When we write $X \lessapprox Y$ or $Y \gtrapprox X$, we mean that

$$|X| \leqslant C_A Y \log^{C(A+1)} N \tag{7.1}$$

for some constant C which does not depend on A, and some constant C_A which can depend (possibly in an ineffective manner) on A. The constants C and C_A can be different in different instances of this notation. In all our arguments the exponent C can be chosen effectively, but it may not be possible to give an explicit value of C_A due to the possibility of Siegel zeros.

In some cases, statements of the form $X \lessapprox Y$ will appear as both hypotheses and conclusions of a proposition. In such cases it is understood that the implied constants in the conclusions are dependent on the implied constants in the hypotheses. Somewhat more subtly, in the course of an argument we may divide into several cases using this notation (*e.g.* we may divide into two cases $X \lessapprox Y$ and $X \not\lessapprox Y$). Once again, the implied constants in the conclusion of this argument will depend on the implied

constants used to create the division of cases. When necessary we shall draw attention to these dependence-of-constants issues[(6)].

Our argument here shall broadly follow that used to prove (1.3) in §5. We begin with the analogue of Proposition 5.1.

PROPOSITION 7.1 (Correlation with quadratic phase implies major arc). *Let α, β, γ be real numbers, $A > 0$, and let N be a large integer such that*

$$\left|\mathbb{E}_{N<n\leqslant 2N}\mu(n)e(-\alpha n^2 - \beta n - \gamma)\right| \geqslant \log^{-A} N. \tag{7.2}$$

Then there exists D, $1 \leqslant D \ll N^{2/3}$, an integer $q \lessapprox 1$ and a $\theta \in \mathbb{R}$ such that

$$\#\{d \in (D, 2D] : \|q\alpha d^2 - \theta\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{D^2}{N^2}\} \gtrapprox D. \tag{7.3}$$

Furthermore if $D < N^{1/3}$ we can take $\theta = 0$.

Proof. — We can discard the constant phase $e(-\gamma)$. As before, we apply Proposition 4.2 with $U = V = N^{1/3}$ and conclude one of the following statements:

- (Type I sum is large): There exists D, $1 \leqslant D \ll N^{2/3}$, such that
$$\left|\mathbb{E}_{N/d<w\leqslant 2N/d}e(\alpha d^2 w^2 + \beta dw)\right| \gtrapprox 1$$
for $\gtrapprox D$ values of $d \in (D, 2D]$.

- (Type II sum is large): There exist integers D, W with $N^{1/3} \ll D \ll N^{2/3}$ and $N/4 \leqslant DW \leqslant 4N$, such that
$$\left|\mathbb{E}_{D<d,d'\leqslant 2D}\mathbb{E}_{W<w,w'\leqslant 2W}e(\phi(dw) - \phi(d'w) - \phi(dw') + \phi(d'w'))\right| \gtrapprox 1$$
where $\phi(n) := \alpha n^2 + \beta n$.

Suppose first that the Type I sum is large. Applying Lemma A.11, we can find an integer $q \lessapprox 1$ such that $\|qd^2\alpha\|_{\mathbb{R}/\mathbb{Z}} \lessapprox D^2/N^2$ for $\gtrapprox D$ values of $D < d \leqslant 2D$, which implies (7.3) (with $\theta = 0$).

Now suppose instead that the Type II sum is large. By the pigeonhole principle, we can find d', w' such that

$$\left|\mathbb{E}_{D<d\leqslant 2D}\mathbb{E}_{W<w\leqslant 2W}e(\phi(dw) - \phi(d'w) - \phi(dw') + \phi(d'w'))\right| \gtrapprox 1$$

(6) One can of course rewrite all the arguments in this paper replacing every appearance of $X \lessapprox Y$ or $Y \gtrapprox X$ by suitably explicit long-hand forms (7.1), although some of the constants may be ineffective. However we have found that this tended to clutter the estimates with distracting numerical constants, and so we have chosen instead to suppress all of these constants.

and hence

$$\left|\mathbb{E}_{W<w\leqslant 2W} e(\phi(dw)-\phi(d'w)-\phi(dw')+\phi(d'w'))\right| \gtrapprox 1$$

for $\gtrapprox D$ values of d. Now the phase $\phi(dw)-\phi(d'w)-\phi(dw')+\phi(d'w')$ is quadratic in w with a leading coefficient of $\alpha(d^2-(d')^2)$. We may thus apply Lemma A.11 and conclude that there exists $q \lessapprox 1$ such that

$$\|q\alpha(d^2-(d')^2)\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{D^2}{N^2}. \tag{7.4}$$

Pigeonholing in q, we conclude there exists a single value of q such that (7.4) follows for $\gtrapprox D$ values of $d \in (D, 2D]$. Setting $\theta := q\alpha(d')^2$, the claim follows. □

By using Lemma A.4, we can now conclude the analogue of Corollary 5.2.

Proposition 7.2 (Correlation with quadratic phase implies major arc, II). — *Let α, β, γ be real numbers, $A > 0$, and let N be a large integer such that* (7.2) *holds. Then we have*

$$\|\alpha\|_{\mathbb{R}/\mathbb{Z},Q} \lessapprox N^{-2}$$

for some $Q \lessapprox 1$.

Proof. — We apply Proposition 7.1 to obtain D, $1 \leqslant D \leqslant N^{2/3}$, and $q \lessapprox 1$ obeying (7.3). If[7] $D \lessapprox 1$ then certainly $D \ll N^{1/3}$, and so we may take $\theta = 0$. There then exists $d \in (D, 2D]$ such that

$$\|q\alpha d^2\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{D^2}{N^2} \lessapprox N^{-2},$$

and the claim follows on replacing q by qd^2.

Now let us suppose that $D \not\lessapprox 1$. We will not be able to apply Lemma A.12 as it is not sufficiently "amplified" for our use here. Instead, we use the triangle inequality and (7.3) to obtain

$$\#\Big\{d, d' \in (D, 2D] : \|q\alpha(d^2-(d')^2)\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{D^2}{N^2}\Big\} \gtrapprox D^2.$$

The diagonal case $d = d'$ is negligible since $D \not\lessapprox 1$, *i.e.*

$$\#\Big\{d, d' \in (D, 2D] : d \neq d', \|q\alpha(d^2-(d')^2)\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{D^2}{N^2}\Big\} \gtrapprox D^2. \tag{7.5}$$

Writing $d^2-(d')^2 = d_1d_2$, where $d_1 := d-d'$ and $d_2 := d+d'$, we conclude

$$\#\Big\{d_1, d_2 : 1 \leqslant |d_1|, |d_2| \leqslant 4D : \|q\alpha d_1 d_2\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{D^2}{N^2}\Big\} \gtrapprox D^2.$$

[7] This is an instance of the subtlety of the $\lessapprox$ notation. By this we mean that $D \leqslant C_A \log^{C(A+1)} N$, where C is chosen so that if $D > C_A \log^{C(A+1)} N$ then the later estimate (7.5) holds true.

By reflection symmetry we may take d_1, d_2 to be positive. In particular, for $\gtrapprox D$ values of d_1 in $[1, 4D]$, we have

$$\#\Big\{d_2 \in [1,4D] : \|q\alpha d_1 d_2\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{D^2}{N^2}\Big\} \gtrapprox D.$$

Applying Lemma A.4 (ii) we thus conclude that for each such d_1, there exists $q_{d_1} \lessapprox 1$ such that

$$\|q\alpha d_1 q_{d_1}\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{D}{N^2}.$$

Applying the pigeonhole principle, we can thus find $q' \lessapprox 1$ such that

$$\#\Big\{1 \leqslant d_1 \leqslant 4D : \|q\alpha d_1 q'\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{D}{N^2}\Big\} \gtrapprox D.$$

Applying Lemma A.4 (ii) again, we conclude that there exists $q'' \lessapprox 1$ such that

$$\|q\alpha q'' q'\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{1}{N^2}.$$

Since $qq'q'' \lessapprox 1$, the claim follows. □

On the other hand, we have the quadratic analogue of Proposition 5.3:

PROPOSITION 7.3 (Major arc quadratic phases are orthogonal to Möbius). — *Let N be a large integer, let $\alpha, \beta, \gamma \in \mathbb{R}/\mathbb{Z}$, and let $Q, K \geqslant 1$ be such that $\|\alpha\|_{\mathbb{R}/\mathbb{Z},Q} \leqslant K/N^2$. Then we have*

$$\mathbb{E}_{N<n\leqslant 2N}\mu(n)e(-\alpha n^2 - \beta n - \gamma) \ll_{A'} Q^{1/3}K^{1/3}\log^{-A'} N$$

for any $A' > 0$ (the implied constant is ineffective).

Proof. — Let $1 \leqslant M < N$ be a parameter to be chosen later. We can set $\gamma = 0$. Arguing as in the proof of Proposition 5.3, we have

$$\begin{aligned}\big|\mathbb{E}_{N<n\leqslant 2N}\mu(n)e(-\alpha n^2 - \beta n)\big| \\ \ll \sup_{|I|=M; I\subset[N,2N]}\Big|\frac{1}{M}\sum_{n\in I}\mu(n)e(-\alpha n^2 - \beta n)\Big| + \frac{M}{N}.\end{aligned}$$

By hypothesis, we have an integer a and $1 \leqslant q \leqslant Q$ such that $|\alpha - \frac{a}{q}| \leqslant \frac{K}{N^2}$. We thus have

$$\begin{aligned}e(\alpha n^2) &= e(an^2/q)e((\alpha - a/q)n^2) \\ &= e(an^2/q)e(2(\alpha - a/q)(n - n_I))e((\alpha - a/q)n_I^2) + O\Big(\frac{KM^2}{N^2}\Big) \\ &= e(an^2/q)e(2(\alpha - a/q)n)(b)(\alpha, a/q, n_I) + O\Big(\frac{KM^2}{N^2}\Big)\end{aligned}$$

for any $n, n_I \in I$, where we use the $(b)()$ notation from Appendix A. Discarding the constant phase $(b)(\alpha, a/q, n_I)$, and absorbing the linear phase $e(2(\alpha - a/q)n)$ into the $e(\beta n)$ factor we conclude

$$\mathbb{E}_{N<n\leqslant 2N}\mu(n)e(-\alpha n^2 - \beta n)$$
$$\ll \sup_{|I|=M; I\subset[N,2N]; \beta'\in\mathbb{R}} \Big|\frac{1}{M}\sum_{n\in I}\mu(n)e(-an^2/q - \beta' n)\Big| + \frac{KM^2}{N^2} + \frac{M}{N}.$$

The function $e(an^2/q)$ is periodic of period q, and can thus be decomposed as a Fourier series $e(an^2/q) = \sum_{b=0}^{q-1} c_b e(bn/q)$ where the coefficients c_b are Gauss sums and can be computed explicitly. From Plancherel's theorem and the Cauchy-Schwarz inequality we have $\sum_{b=0}^{q-1}|c_b| = O(q^{1/2})$ (*cf.* the proof of Proposition 3.2). Applying (1.3) (with A replaced by $2A'$) we conclude that

$$\sum_{n\in I}\mu(n)e(-an^2/q - \beta' n) \ll_{A'} Nq^{1/2}\log^{-2A'} N,$$

and hence

$$\mathbb{E}_{N<n\leqslant 2N}\mu(n)e(-\alpha n^2 - \beta n - \gamma) \ll_{A'} \frac{N}{M}q^{1/2}\log^{-2A'} N + \frac{KM^2}{N^2} + \frac{M}{N}.$$

If we set $M := K^{-1/3}q^{1/6}N\log^{-A'} N$ we obtain the claim. □

Propositions 7.2 and 7.3 together imply (1.7), though the $\lessapprox$ notation does take some unravelling. Suppose for a contradiction that (7.2) holds. Then Proposition 7.2 implies that $\|\alpha\|_{\mathbb{R}/\mathbb{Z},Q} \leqslant K/N^2$, where we may take $K = Q = C_A\log^{C(A+1)} N$ for some absolute C. Proposition 7.3 now implies, taking $A' = C(A+1)$, that

$$\mathbb{E}_{N<n\leqslant 2N}\mu(n)e(-\alpha n^2 - \beta n - \gamma) \ll_{A'} \log^{-A'/3} N.$$

We may clearly assume that $C > 3$, and so this does contradict our assumption that (7.2) holds, at least if $N > N_0(A)$ is sufficiently large. To conclude the proof of (1.7), one simply applies Lemma A.7 with $\varphi \equiv 1$.

Remark. — It is straightforward to iterate the above argument, as is done in the standard theory of Weyl exponential sums, to obtain a generalisation of (1.7) in which $\alpha n^2+\beta n+\gamma$ is replaced by an arbitrary polynomial. We will, however, not pursue this generalisation here.

8. Locally quadratic phase functions, I: a technical reduction

We now begin the (onerous) task of proving Theorem 2.2. Let us begin by recalling the statement:

THEOREM 2.2 (μ is strongly orthogonal to local quadratics). — *Let G/Γ be a 1-step nilmanifold, let $F : G/\Gamma \to \mathbb{C}$ be a Lipschitz function, and let $g \in G$ and $x \in G/\Gamma$ be arbitrary. Let $\phi : B_N \to \mathbb{R}/\mathbb{Z}$ be a phase which is locally quadratic on the* Bohr set $B_N := \{n \in [N] : F(T_g^n x) \neq 0\}$. *Then we have*

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{F(T_g^n x)}e(-\phi(n)) \ll_{G/\Gamma,A} \|F\|_{\mathrm{Lip}} \log^{-A} N.$$

Our objective in this (rather technical) section is to reduce this to a similar result which has certain important technical advantages. The most critical of these is that ϕ can be extended to a function which is quadratic somewhat beyond the domain $B_N = \mathrm{Supp}_n F(g^n x)$. This refined formulation reads as follows.

PROPOSITION 8.1 (μ is strongly orthogonal to extendible local quadratics). — *Let $g \in G$, $x \in G/\Gamma$, $n_0 \in \mathbb{Z}$, and let $\rho_0 \in (0, 10^{-5})$ be a small radius. Suppose that $B_g(n_0, 100\rho_0)$ is contained in $\{n \in \mathbb{Z} : N < n \leqslant 2N\}$, and suppose that $\phi : \mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ is a function which is locally quadratic when restricted to $B_g(n_0, 100\rho_0)$. Let $\psi : \mathbb{Z} \to \mathbb{R}^+$ be a function supported on $B_g(n_0, \rho_0)$ which obeys the Lipschitz property*

$$|\psi(n) - \psi(m)| \leqslant \|n - m\|_g \text{ for all } n, m \in \mathbb{Z}. \tag{8.1}$$

Then we have

$$\left|\mathbb{E}_{N<n\leqslant 2N}\mu(n)\psi(n)e(-\phi(n))\right| \ll_{A,G/\Gamma} \log^{-A} N. \tag{8.2}$$

Proof that Proposition 8.1 implies Theorem 2.2. — By renormalising, we may assume that $\|F\|_{\mathrm{Lip}} \leqslant 1$.

The essential idea is that a "ball" (say B_N) can be covered by balls $B_g(n_0, \epsilon)$ of a much smaller radius. Most of these will have the property that $B_g(n_0, 100\epsilon)$ is still contained in B_N, and hence that ϕ is still quadratic on $B_g(n_0, 100\epsilon)$.

We turn to the details. First of all, an application of Lemma A.7 implies that it suffices to establish the estimate

$$\mathbb{E}_{N\leqslant n\leqslant 2N}\mu(n)\varphi(n/N)F(T_g^n x)e(-\phi(n)) \ll_{A,G/\Gamma} \log^{-A} N, \tag{8.3}$$

where $\varphi : \mathbb{R} \to \mathbb{R}$ is the function

$$\varphi(t) := \begin{cases} 6(t - \frac{7}{6}) & \text{if } \frac{7}{6} \leqslant t \leqslant \frac{3}{2}; \\ 6(\frac{11}{6} - t) & \text{if } \frac{3}{2} \leqslant t \leqslant \frac{11}{6}. \end{cases}$$

(any similar function would work). The phase $e(\alpha n)$ which featured in that lemma has been absorbed into the quadratic phase $e(-\phi(n))$.

We now replace F by a "smooth-thresholded" function $\widetilde{F}$, as constructed in Lemma A.13. Let $\rho_0 \in (0, 10^{-5})$ be a parameter to be chosen later, and set $\delta := 10^4\rho_0$ in Lemma A.13. This provides a Lipschitz function $\widetilde{F} : G/\Gamma \to \mathbb{R}$ satisfying properties (i), (ii) and (iii) of that lemma. In particular from Lemma A.13 (iii) we see that

$$\begin{aligned}(8.4)\quad \mathbb{E}_{N\leqslant n\leqslant 2N}\mu(n)\varphi(n/N)F(T_g^n x) \\ = \mathbb{E}_{N\leqslant n\leqslant 2N}\mu(n)\varphi(n/N)\widetilde{F}(T_g^n x) + O(\rho_0).\end{aligned}$$

Now take a partition of unity $1 = \sum_\alpha \chi_\alpha$ on G/Γ, where

(1) Each χ_α is supported on a ball of diameter at most $\rho_0/2$;
(2) Each χ_α is bounded in magnitude by 1 and satisfies $\|\chi_\alpha\|_{\mathrm{Lip}} \ll_{G/\Gamma} \rho_0^{-1}$;
(3) The number of χ_α is $O(\rho_0^{-O_{G/\Gamma}(1)})$.

We leave the construction of such a partition to the reader: modelling G/Γ by a torus, one may be quite explicit. This partition of unity induces a decomposition

$$\widetilde{F} = \sum_\alpha F_\alpha,$$

where $F_\alpha := F\chi_\alpha$. Note that since both F and χ_α are bounded we have, using Lemma A.13 (i), that

$$(8.5)\qquad \|F_\alpha\|_{\mathrm{Lip}} \ll \|\widetilde{F}\|_{\mathrm{Lip}} + \|\chi_\alpha\|_{\mathrm{Lip}} \ll_{G/\Gamma} \rho_0^{-1}.$$

We may also effect a Lipschitz decomposition

$$\varphi = \sum_\beta \varphi_\beta$$

of φ into $O(\rho_0^{-1})$ Lipschitz functions φ_β with Lipschitz constant $O(\rho_0^{-1})$, each supported on an interval of diameter $\rho_0/2$. Write

$$\psi_{\alpha,\beta}(n) := F_\alpha(T_g^n x)\varphi_\beta(n/N).$$

Noting that

$$\varphi(n/N)\widetilde{F}(T_g^n x) = \sum_{\alpha,\beta}\psi_{\alpha,\beta},$$

it follows from (8.4) and the triangle inequality that

$$\begin{aligned}(8.6)\quad \mathbb{E}_{N\leqslant n\leqslant 2N}\mu(n)\varphi(n/N)F(T_g^n x) \\ \ll_{G/\Gamma} \rho_0^{-O_{G/\Gamma}(1)} \sup_{\alpha,\beta}\bigl|\mathbb{E}_{N<n\leqslant 2N}\mu(n)\psi_{\alpha,\beta}(n)e(-\phi(n))\bigr| + \rho_0.\end{aligned}$$

Suppose that $n, n' \in \mathrm{Supp}(\psi_{\alpha,\beta})$. Then $\varphi_\beta(n/N), \varphi_\beta(n'/N) \neq 0$, which means that $|n-n'|/N \leqslant \rho_0/2$. Furthermore $F_\alpha(T_g^n x), F_\alpha(T_g^{n'} x) \neq 0$, meaning that $\|g^{n-n'}\|_{G/\Gamma} \leqslant \rho_0/2$. It follows that $\|n-n'\|_g \leqslant \rho_0$, and so the support of $\psi_{\alpha,\beta}$ is contained in some ball $B_g(n_0, \rho_0)$.

We are, of course, going to apply Proposition 8.1. It is therefore necessary to confirm that ϕ is defined on $B_g(n_0, 100\rho_0)$, and also to say something concerning the Lipschitz constant of $\psi_{\alpha,\beta}$.

Starting with the first task, suppose that $\mathrm{Supp}(\psi_{\alpha,\beta}) \subseteq B_g(n_0, \rho_0)$ and that $\psi_{\alpha,\beta}(n_1) \neq 0$ for some $n_1 \in B_g(n_0, \rho_0)$ (we may clearly ignore those α, β for which $\psi_{\alpha,\beta} \equiv 0$). Then $\varphi_\beta(n/N) \neq 0$ and so, due to the choice of φ, we have $7/6 \leqslant n_1/N \leqslant 11/6$. It follows that if $n \in B_g(n_0, 100\rho_0)$ then $|n-n_1|/N \leqslant 101\rho_0$ and thus, since ρ_0 is so small, that $N < n \leqslant 2N$. We also have that $F_\alpha(g^{n_1}x) \neq 0$, which implies that $\widetilde{F}(g^{n_1}x) \neq 0$. Now if $n \in B_g(n_0, 100\rho_0)$ then $d_{G/\Gamma}(g^n x, g^{n_1}x) \leqslant 101\rho_0$. It follows from Lemma A.13 and our choice of δ that $F(g^n x) \neq 0$. We have shown that $B_g(n_0, 100\rho_0) \subseteq B_N$, and hence ϕ is indeed defined on the desired set.

We now examine the Lipschitz constant of $\psi_{\alpha,\beta}$, with the $\|\cdot\|_g$ metric on $\mathbb{Z}$. We have, recalling (8.5), that

$$|\varphi_\beta(n) - \varphi_\beta(n')| \ll \rho_0^{-1} \frac{|n-n'|}{N} \leqslant \rho_0^{-1} \|n-n'\|_g$$

and

$$\left|F_\alpha(g^n x) - F_\alpha(g^{n'} x)\right| \ll \rho_0^{-1} \|g^{n-n'}\|_{G/\Gamma} \leqslant \rho_0^{-1} \|n-n'\|_g.$$

Since both F_α and φ_β are bounded, the Lipschitz constant of $\psi_{\alpha,\beta}$ is $O_{G/\Gamma}(\rho_0^{-1})$.

We are now in a position to apply (a renormalised version of) Proposition 8.1. We deduce that

$$\mathbb{E}_{N<n\leqslant 2N} \mu(n)\psi_{\alpha,\beta}(n) e(-\phi(n)) \ll_A \rho_0^{-1} \log^{-A} N$$

uniformly in α, β. Thus, from (8.6), we see that

$$\mathbb{E}_{N\leqslant n\leqslant 2N} \mu(n)\varphi(n/N) F(g^n x) \ll_{G/\Gamma} \rho_0^{-O_{G/\Gamma}(1)} \log^{-A} N + \rho_0.$$

Setting $\rho_0 := \log^{-A/2C} N$, and recalling that A can be arbitrary, we do indeed conclude Theorem 2.2. □

It will be convenient later on (in the proof of Lemma 11.4) to add some further technical assumptions to the hypotheses of Proposition 8.1. We may assume that ψ is real. Next, recall that G/Γ was embedded isometrically in a torus $(\mathbb{R}/\mathbb{Z})^d$; we may in fact simply replace G/Γ by that torus $(\mathbb{R}/\mathbb{Z})^d$ (using Lemma A.8) since this does not affect anything. It will be convenient to work in $\mathbb{Z}/p\mathbb{Z}$ where p is some prime between $10MN$ and $20MN$. We

can approximate the group element g by the nearest p^{th} root of unity $\widetilde{g}$ in G/Γ, thus $\|g^{-1}\widetilde{g}\|_{G/\Gamma} \ll 1/N$ and $\widetilde{g}^p \in \Gamma$. Observe that the $\|\cdot\|_g$ and $\|\cdot\|_{\widetilde{g}}$ norms are comparable, thanks to the factor of $|\frac{n}{N}|$ in the definition of these norms. Thus we may, after making some trivial adjustments to the constants such as 100 in the proof of Proposition 8.1, replace g by $\widetilde{g}$, that is we may assume that g is a p^{th} root of unity.

9. Locally quadratic phase functions, II: Explicit quadratic and quartic behaviour

We now begin the proof of Proposition 8.1. We are going to show that if (8.2) is false then the phase ϕ is somehow "major arc". Ultimately we will relate it to the type of phases in Proposition 6.3 which, in view of the main result of that proposition, will lead to a contradiction. We have already seen several instances where a hypothesis that the Möbius function μ correlates with some phase implies that the phase is "major arc": Propositions 5.1, 5.2, 7.1 and 7.2 are examples of this. In those cases the phase involved, being either linear or quadratic, was of a simple algebraic kind, but the phases that interest us now are not so explicitly given. The two technical lemmas in this section show that these phases do, nevertheless, enjoy some algebraic structure.

Suppose, for the remainder of the section, that $\phi : B_g(n_0, 100\rho_0) \to \mathbb{R}/\mathbb{Z}$ is a locally quadratic phase. If $\|h_1\|_g, \|h_2\|_g \leqslant 30\rho_0$ then we define

$$\phi''(h_1, h_2) := \phi(n_0 + h_1 + h_2) - \phi(n_0 + h_1) - \phi(n_0 + h_2) + \phi(n_0).$$

This expression is clearly symmetric in h_1, h_2. Since ϕ is locally quadratic on the Bohr set $B_g(n_0, 100\rho_0)$, we conclude the "Taylor expansion"

$$\phi''(h_1, h_2) = \phi(n + h_1 + h_2) - \phi(n + h_1) - \phi(n + h_2) + \phi(n) \tag{9.1}$$

whenever $n \in B_g(n_0, 40\rho_0)$. By telescoping the right-hand side, we conclude the local bilinearity properties

$$\begin{aligned} \phi''(h_1 + h_1', h_2) &= \phi''(h_1, h_2) + \phi''(h_1', h_2); \\ \phi''(h_1, h_2 + h_2') &= \phi''(h_1, h_2) + \phi''(h_1, h_2') \end{aligned} \tag{9.2}$$

whenever $\|h_1\|_g, \|h_2\|_g, \|h_1'\|_g, \|h_2'\|_g \leqslant 15\rho_0$.

As another corollary of Lemma 9.1, we see that ϕ behaves like a genuine quadratic function on certain short arithmetic progressions:

COROLLARY 9.1 (Explicit quadratic structure). — *If $n \in B_g(n_0, 20\rho_0)$, $L \in \mathbb{Z}$ and $h \in B_g(0, 20\rho_0/L)$, then there exist $\alpha, \beta \in \mathbb{R}/\mathbb{Z}$ (depending on n and h) such that*

$$\phi(n + hl) = \tfrac{1}{2}l(l-1)\phi''(h,h) + \alpha l + \beta$$

for all l, $1 \leqslant l \leqslant L$.

Proof. — From (9.1) we obtain the recurrence

$$\phi(n + h(l+2)) - 2\phi(n + h(l+1)) + \phi(n + hl) = \phi''(h,h)$$

for all l, $1 \leqslant l \leqslant L-2$. The claim follows. □

This corollary is strong enough for us to understand the behaviour of the Type I sums which will appear when, in subsequent sections, we analyse

$$\mathbb{E}_{n \in [N]} \mu(n)\psi(n)e(-\phi(n))$$

using Proposition 4.2. The corresponding Type II sums are more difficult. The basic issue here is to understand the algebraic structure of the expression $\psi(dw)e(\phi(dw))$, as a function of d and w. Since ϕ is already quadratic, the phase $\phi(dw)$ here is *quartic* (think of it as being like d^2w^2). We would like some analogue of Corollary 9.1 that makes this quartic structure manifest, for instance we would like $\phi((d+sl)(w+tm))$ to exhibit some explicitly quartic behaviour in l and m, under suitable hypotheses on d, s, l, w, t, m of course. This turns out to be a little tricky, because of the cross terms tdm and slw present in the expression $(d+sl)(w+tm)$. By introducing suitably many constraints (which will be available to us after later arguments) and taking enough differences of the phase, we can eliminate these cross terms and obtain the sought-after quartic structure.

LEMMA 9.2 (Explicit quartic structure). — *Let $d, w; s, t$ and L, M be integers such that*

$$LM\|st\|_g \leqslant \rho_0 \tag{9.3}$$

and let $P : \mathbb{Z} \times \mathbb{Z} \to \mathbb{Z}$ be the quadratic polynomial

$$P(l,m) := (d + sl)(w + tm).$$

Suppose that the integers $l_0, l_1, l_2, m_0, m_1, m_2$ are such that $|l_i|, |m_i| \leqslant L$, and furthermore that all sixteen of the values

$$P(l_0 + i_1l_1 + i_2l_2, j_1m_1 + j_2m_2), \quad i_1, i_2, j_1, j_2 \in \{0,1\}, \tag{9.4}$$

lie in $B_g(n_0, \rho_0)$. *Then we have*

$$\sum_{i_1,i_2,j_1,j_2 \in \{0,1\}} (-1)^{i_1+i_2+j_1+j_2} \phi(P(l_0 + i_1 l_1 + i_2 l_2, m_0 + j_1 m_1 + j_2 m_2)) = 2 l_1 l_2 m_1 m_2 \phi''(st, st). \tag{9.5}$$

Remark. — This lemma is a generalisation of the observation that if $\phi(n) = an^2 + bn + c$ is a quadratic, and one differentiates $\phi(P(l, m))$ twice in the l variable and twice in the m variable, one gets $2 \times \phi'' \times st \times st$, where $\phi'' = 2a$ is the double derivative of ϕ. It is key here that we have the sixteen constraints (9.4): this gives us sufficient instances where (9.1) and (9.2) may be applied. Later arguments (involving many applications of the Cauchy-Schwarz inequality) will put us in a situation where we have such a multiplicity of constraints at our disposal.

Proof. — By replacing d, w by $d + l_0 s$ and $w + m_0 t$ we may assume that $l_0 = m_0 = 0$. Let l_1, l_2, m_1, m_2 be as in the hypothesis of the lemma, that is to say $|l_i|, |m_i| \leqslant L$ and the sixteen constraints (9.4) are satisfied. From the identities

$$dw = P(0,0)$$
$$swl_1 = P(l_1, 0) - P(0,0), \quad swl_2 = P(l_2, 0) - P(0,0)$$
$$tdm_1 = P(0, m_1) - P(0,0), \quad tdm_2 = P(0, m_2) - P(0,0),$$

we see that

$$dw \in B_g(n_0, \rho_0) \quad \text{and} \quad swl_1, swl_2, tdm_1, tdm_2 \in B_g(0, 2\rho_0). \tag{9.6}$$

Now fix $i_1, i_2 \in \{0, 1\}$ and consider the sum

$$\phi(P(i_1 l_1 + i_2 l_2, m_1 + m_2)) - \phi(P(i_1 l_1 + i_2 l_2, m_1)) - \phi(P(i_1 l_1 + i_2 l_2, m_2)) + \phi(P(i_1 l_1 + i_2 l_2, 0)). \tag{9.7}$$

We can rewrite this as

$$\phi(n + h_1 + h_2) - \phi(n + h_1) - \phi(n + h_2) + \phi(n) \tag{9.8}$$

where $n := w(d + i_1 sl_1 + i_2 sl_2)$, $h_1 := (d + i_1 sl_1 + i_2 sl_2) tm_1$ and $h_2 := (d + i_1 sl_1 + i_2 sl_2) tm_2$. From (9.3) and (9.6) we see that $n \in B_g(n_0, 5\rho_0)$, and that $h_1, h_2 \in B_g(0, 4\rho_0)$. Thus all four of $n, n + h_1, n + h_2, n + h_1 + h_2$ lie in $B_g(n_0, 13\rho_0)$ and (9.1) is applicable, which means we can rewrite (9.8) as

$$\phi''((d + i_1 sl_1 + i_2 sl_2) tm_1, (d + i_1 sl_1 + i_2 sl_2) tm_2).$$

Applying (9.2) and (9.6), (9.3), we can expand this as

$$(9.7) = X + i_1 Y + i_2 Z + 2 i_1 i_2 l_1 m_1 l_2 m_2 \phi''(st, st)$$

where X, Y, Z are quantities which depend on $\phi, d, s, t, l_1, m_1, l_2, m_2$ but are independent of i_1, i_2. If one then takes an alternating sum of this identity over the four possible choices of $i_1, i_2 \in \{0,1\}$ to eliminate the X, Y, Z terms, one obtains (9.5). □

10. Quadratic bias implies major arc

With the above preliminaries out of the way, we now begin the proof of Proposition 8.1 in earnest. In this section we shall establish the main step of this proof, namely that a quadratic bias necessarily implies a "major arc" condition on ϕ. We persist in our use of the notations $X \lessapprox Y$ and $X \gtrapprox Y$, which were introduced in §7. Recall (*cf.* (7.1)) that $X \lessapprox Y$ means that

$$X \leqslant C_A Y \log^{C(A+1)} N$$

for some constant C which does not depend on A. That constant is, from now on, allowed to depend on the underlying 2-step nilmanifold G/Γ (in actuality, it will depend on the *dimension* of that nilmanifold). The constant C_A is of course also allowed to depend on G/Γ. Recall also from Appendix A the notation

$$\|\alpha\|_{\mathbb{R}/\mathbb{Z},Q} := \sup_{q \leqslant Q} \|q\alpha\|_{\mathbb{R}/\mathbb{Z}}.$$

The main result of this section is as follows.

Proposition 10.1. — *Let the notation and assumptions be as in the previous section. Suppose that*

$$\left|\mathbb{E}_{N<n\leqslant 2N} \mu(n)\psi(n)e(-\phi(n))\right| \geqslant \log^{-A} N. \tag{10.1}$$

Then there exist $X_0 \lessapprox 1$, $D \leqslant 4N^{2/3}$ and $Q \lessapprox 1$ with the following property: for any X with $X_0 < X < N^{1/10}$, there exists a set $\mathcal{D} \subseteq [1, D]$, $|\mathcal{D}| \gtrapprox D/X^{1/2}$, such that if $d \in \mathcal{D}$ and $w \in \mathbb{Z}$ satisfies $\|dw\|_g \leqslant 1/X$ then

$$\|\phi''(dw, dw)\|_{\mathbb{R}/\mathbb{Z},Q} \lessapprox X^{-2}.$$

Remark. — The conclusion here is an assertion that $\phi''(h,h)$ is major arc for many values of h. We shall recast this conclusion into a more tractable form in the next section (in particular it is necessary to show that as d, w range over the values allowed in the conclusion of the proposition, $h = dw$ takes on many different values).

Proof. — Since ψ is Lipschitz and supported on $B(n_0, \rho_0)$, we have $\|\psi\|_\infty \ll \rho_0$, and so we conclude from (10.1) that

$$\rho_0 \gtrapprox 1. \tag{10.2}$$

In practice, this will make it fairly easy to verify hypotheses such as $\|h\|_g \leqslant \rho_0$ which occur in the lemmas of the previous section.

We now apply Proposition 4.2 with $f(n) := \psi(n)e(\phi(n))$ and $U = V = N^{1/3}$ to conclude one of the following statements must be true:

- (Type I sum is large): There exists an integer $1 \leqslant D \leqslant N^{2/3}$ such that

$$\left|\mathbb{E}_{N/d<w\leqslant 2N/d}\psi(dw)e(\phi(dw))\right| \gtrapprox 1 \tag{10.3}$$

  for $\gtrapprox D$ integers d such that $D < d \leqslant 2D$.

- (Type II sum is large): There exists integers D, W with $\frac{1}{2}N^{1/3} \leqslant D \leqslant 4N^{2/3}$ and $N/4 \leqslant DW \leqslant 4N$, such that

$$\begin{aligned}\Big|\mathbb{E}_{D<d,d'\leqslant 2D}\mathbb{E}_{W<w,w'\leqslant 2W}\psi(dw)\psi(d'w)\psi(dw')\psi(d'w') \\ \times e(\phi(dw) - \phi(d'w) - \phi(dw') + \phi(d'w'))\Big| \gtrapprox 1.\end{aligned} \tag{10.4}$$

We can thus assume that either (10.3) or (10.4) holds, and see what this implies about ϕ. We handle the two cases separately.

Large Type I sums. — Let us consider the (substantially simpler) Type I case when (10.3) holds for many values of D. The bulk of the argument is contained inside the following lemma.

LEMMA 10.2 (Large Type I sum implies major arc). — *Let d, $D \leqslant d < 2D$, be such that* (10.3) *holds, that is to say*

$$\left|\mathbb{E}_{N/d<w\leqslant 2N/d}\psi(dw)e(\phi(dw))\right| \gtrapprox 1.$$

Assume that N is large depending on A. Then there exist $Q \lessapprox 1$ and $\varepsilon \gtrapprox 1$ such that

$$\|\phi''(dt, dt)\|_{\mathbb{R}/\mathbb{Z},Q} \lessapprox L^{-2}$$

whenever $L \geqslant 1$ and $t \in \mathbb{Z}$ is such that $\|dt\|_g \leqslant \varepsilon/L$.

Proof. — The idea is to analyse the quantity in (10.3) locally on short progressions of common difference t and length L. Since ψ is supported on $(N, 2N]$, we have

$$\left|\sum_w \psi(dw)e(\phi(dw))\right| \gtrapprox \frac{N}{D}.$$

From the averaging identity

$$\sum_w f(w) = \sum_w \mathbb{E}_{1 \leqslant l \leqslant L} f(w + tl),$$

valid for any compactly supported function $f : \mathbb{Z} \to \mathbb{C}$, we conclude

$$\Big| \sum_w \mathbb{E}_{1 \leqslant l \leqslant L} \psi(dw + dtl) e(\phi(dw + dtl)) \Big| \gtrapprox \frac{N}{D}.$$

Since ψ is supported on $(N, 2N]$ and

$$|dtl| \leqslant LN |dtl/N| \leqslant LN \|dt\|_g \leqslant \varepsilon N,$$

we see that in this sum we still have the constraint $|dw| = O(N)$, and whence $w = O(N/D)$. Thus by the pigeonhole principle we can find w such that

$$\big| \mathbb{E}_{1 \leqslant l \leqslant L} \psi(dw + dtl) e(\phi(dw + dtl)) \big| \gtrapprox 1.$$

By (8.1) we have

$$\psi(dw + dtl) = \psi(dw) + O(l \|dt\|_g) = \psi(dw) + O(\varepsilon)$$

and hence (if $\varepsilon \gtrapprox 1$ is chosen suitably small)

$$\big| \mathbb{E}_{1 \leqslant l \leqslant L} \psi(dw) e(\phi(dw + dtl)) \big| \gtrapprox 1.$$

Since $\psi(dw)$ is bounded and independent of l, it can be discarded and this becomes

$$\big| \mathbb{E}_{1 \leqslant l \leqslant L} e(\phi(dw + dtl) \big| \gtrapprox 1.$$

We apply Corollary 9.1 with $n := dw$ and $h := dt$. We may assume, in view of (10.2), that $\varepsilon \leqslant 20\rho_0$ which means that $h \in B_g(0, 20\rho_0/L)$. Since $n \in \mathrm{Supp}(\psi) \subseteq B_g(n_0, \rho_0)$, Corollary 9.1 does indeed apply and we may infer the existence of $\alpha, \beta \in \mathbb{R}/\mathbb{Z}$ such that

$$\big| \mathbb{E}_{1 \leqslant l \leqslant L} e(\tfrac{1}{2} l(l-1) \phi''(dt, dt) + \alpha l + \beta) \big| \gtrapprox 1.$$

Now if $L \geqslant \log^{C(A+1)} N$, for sufficiently large C, then Lemma A.11 applies and we may indeed conclude that $\|\phi''(dt, dt)\|_{\mathbb{R}/\mathbb{Z}, Q} \lessapprox L^{-2}$. If L is not this large then (because so much may be hidden inside the $\lessapprox$ symbol) the conclusion is trivial anyway. □

The deduction of Proposition 10.1 in the Type I case is almost immediate. Indeed from the preceding lemma we see that for $\gtrapprox D$ values of $d \in [D, 2D)$ we have

$$\|\phi''(dt, dt)\|_{\mathbb{R}/\mathbb{Z}, Q} \lessapprox L^{-2}$$

whenever $t \in \mathbb{Z}$ is such that $\|dt\|_g \leqslant \varepsilon/L$. Now simply let $\mathcal{D}$ be the set of such d, set $L := X/\varepsilon$, and require that $X_0 \lessapprox 1$ be large enough that $L \geqslant 1$ whenever $X > X_0$.

Large Type II sums. — We move on now to the much more complicated Type II case where (10.4) holds. That is to say, we work under the assumption that

$$\big|\mathbb{E}_{D<d,d'\leqslant 2D}\mathbb{E}_{W<w,w'\leqslant 2W}\psi(dw)\psi(d'w)\psi(dw')\psi(d'w') \times e(\phi(dw)-\phi(d'w)-\phi(dw')+\phi(d'w'))\big| \gtrapprox 1$$

where $\frac{1}{2}N^{1/3}\leqslant D\leqslant 4N^{2/3}$ and $\frac{1}{4}N\leqslant DW\leqslant 4N$.

LEMMA 10.3 (Type II sum implies major arc). — *Let $\frac{1}{2}N^{1/3}\leqslant D\leqslant 4N^{2/3}$ be such that $\frac{1}{4}N\leqslant DW\leqslant 4N$ and (10.4) holds. Assume that N is large depending on A. Then there exist $Q\lessapprox 1$ and $\varepsilon\gtrapprox 1$ with the property that*

$$\|\phi''(st,st)\|_{\mathbb{R}/\mathbb{Z},Q}\lessapprox 1/L^2M^2$$

whenever $s,t\in\mathbb{Z}$ and $L,M\in\mathbb{Z}^+$ are such that $L|s|\leqslant\varepsilon D$, $M|t|\leqslant\varepsilon W$, $L,M\geqslant 1/\varepsilon$ and $\|st\|_g\leqslant\varepsilon^2/LM$.

Proof. — It will be convenient to use the $\mathbf{b}(x_1,\dots,x_k)$ notation introduced in Appendix A. Thus for instance we can write (10.4) as

$$\big|\mathbb{E}_{D<d,d'\leqslant 2D}\mathbb{E}_{W<w,w'\leqslant 2W}\psi(dw)e(\phi(dw))\mathbf{b}(d,w')\mathbf{b}(d',w)\mathbf{b}(d',w')\big|\gtrapprox 1.$$

By the pigeonhole principle, we can thus find d',w' such that

$$\big|\mathbb{E}_{D<d\leqslant 2D}\mathbb{E}_{W<w\leqslant 2W}\psi(dw)e(\phi(dw))\mathbf{b}(d,w')\mathbf{b}(d',w)\mathbf{b}(d',w')\big|\gtrapprox 1$$

which upon relabeling the bounded functions $\mathbf{b}$ becomes simply

$$|X|\gtrapprox 1$$

where X is the quantity

$$X:=\mathbb{E}_{D<d\leqslant 2D}\mathbb{E}_{W<w\leqslant 2W}\psi(dw)e(\phi(dw))\mathbf{b}(d)\mathbf{b}(w).$$

Now we argue somewhat as in the proof of Lemma 10.2, averaging d and w over arithmetic progressions. For any $1\leqslant l\leqslant L$ and $1\leqslant m\leqslant M$ we can make the change of variables $d\to d+sl$, $w\to w+tm$ to obtain

$$X=\mathbb{E}_{D-sl<d\leqslant 2D-sl}\mathbb{E}_{W-tm<w\leqslant 2W-tm}\psi((d+sl)(w+tm)) \times e(\phi((d+sl)(w+tm)))\mathbf{b}(d+sl)\mathbf{b}(w+tm).$$

From our assumption that $L|s|\leqslant\varepsilon D$ and $M|t|\leqslant\varepsilon W$ we infer that

$$X=\mathbb{E}_{D<d\leqslant 2D}\mathbb{E}_{W<w\leqslant 2W}\psi((d+sl)(w+tm))e(\phi((d+sl)(w+tm))) \times\mathbf{b}(d+sl)\mathbf{b}(w+tm)+O(\varepsilon).$$

Averaging over l and m gives

$$X = \mathbb{E}_{D<d\leqslant 2D}\mathbb{E}_{W<w\leqslant 2W}\mathbb{E}_{1\leqslant l\leqslant L}\mathbb{E}_{1\leqslant m\leqslant M}\psi((d+sl)(w+tm)) \\ \times e(\phi((d+sl)(w+tm)))\mathbf{b}(d+sl)\mathbf{b}(w+tm)+O(\varepsilon).$$

If $\varepsilon \gtrapprox 1$ is sufficiently small, the assumption that $X \gtrapprox 1$ implies that

$$\big|\mathbb{E}_{D<d\leqslant 2D}\mathbb{E}_{W<w\leqslant 2W}\mathbb{E}_{1\leqslant l\leqslant L}\mathbb{E}_{1\leqslant m\leqslant M}\psi((d+sl)(w+tm)) \\ \times e(\phi((d+sl)(w+tm)))\mathbf{b}(d+sl)\mathbf{b}(w+tm)\big| \gtrapprox 1.$$

Hence by the pigeonhole principle there exist d, w such that

$$\big|\mathbb{E}_{1\leqslant l\leqslant L}\mathbb{E}_{1\leqslant m\leqslant M}\psi((d+sl)(w+tm))e(\phi((d+sl)(w+tm))) \\ \times \mathbf{b}(d+sl)\mathbf{b}(w+tm)\big| \gtrapprox 1.$$

Fix such d, w. By relabeling the $\mathbf{b}$'s, we can write $\mathbf{b}(d+sl)\mathbf{b}(w+tm)$ simply as $\mathbf{b}(l)\mathbf{b}(m)$. We also set

$$P(l,m) := (d+sl)(w+tm).$$

We have, then, that

$$\Big|\sum_{l,m} f(l,m)\mathbf{b}(l)\mathbf{b}(m)\Big| \gtrapprox LM$$

where

$$f(l,m) := \psi(P(l,m)))e(\phi(P(l,m)))1_{1\leqslant l\leqslant L}1_{1\leqslant m\leqslant M}. \tag{10.5}$$

Using Lemma A.10 to eliminate the $\mathbf{b}(l)\mathbf{b}(m)$ factors, we conclude

$$\Big|\sum_{l,l',m,m'} f(l,m)\overline{f(l,m')f(l',m)}f(l',m')\Big| \gtrapprox L^2M^2.$$

We write $l = l_0$, $l' = l_0 + l_1$, $m = m_0$, $m' = m_0 + m_1$ to obtain

$$\Big|\sum_{l_1,m_1}\sum_{l_0,m_0} F(l_0,m_0;l_1,m_1)\Big| \gtrapprox L^2M^2$$

where

$$F(l_0,m_0;l_1,m_1) := f(l_0,m_0)\overline{f(l_0,m_0+m_1)f(l_0+l_1,m_0)} \\ \times f(l_0+l_1,m_0+m_1).$$

Applying Lemma A.10 again, this time in the l_0 and m_0 variables, we see that

$$\Big|\sum_{l_1,m_1}\sum_{l_0,m_0,l_0',m_0'} F(l_0,m_0;l_1,m_1)\overline{F(l_0,m_0';l_1,m_1)F(l_0',m_0;l_1,m_1)} \times F(l_0',m_0';l_1,m_1)\Big| \gtrapprox L^3M^3. \tag{10.6}$$

Writing $l_0' = l_0 + l_2$, $m_0' = m_0 + m_2$, this becomes

$$\Big|\sum_{l_0,l_1,l_2,m_0,m_1,m_2} G(l_0,l_1,l_2,m_0,m_1,m_2)\Big| \gtrapprox L^3M^3$$

where

$$\begin{aligned} G(l_0,l_1,l_2,m_0,m_1,m_2) &:= F(l_0,m_0;l_1,m_1)\overline{F(l_0,m_0+m_2;l_1,m_1)} \\ &\quad \times \overline{F(l_0,m_0+m_2;l_1,m_1)} \\ &\quad \times F(l_0+m_2,m_0+m_2;l_1,m_1) \\ &= \prod_{(i_1,i_2,j_1,j_2)\in\{0,1\}^4} \mathcal{C}^{i_1+i_2+j_1+j_2} f(l_0+i_1l_1+i_2l_2, m_0 \\ &\quad + j_1m_1 + j_2m_2) \end{aligned}$$

and $\mathcal{C} : z \mapsto \overline{z}$ is the conjugation operator. Observe that the support of the sum in (10.6) is still contained in the region $|l_i| \leqslant L$, $|m_i| \leqslant M$. By the pigeonhole principle, we can find l_0 and m_0 such that

$$\Big|\sum_{l_1,l_2,m_1,m_2}\prod_{(i_1,i_2,j_1,j_2)\in\{0,1\}^4} \mathcal{C}^{i_1+i_2+j_1+j_2} f \times \big(l_0+i_1l_1+i_2l_2, m_0+j_1m_1+j_2m_2\big)\Big| \gtrapprox L^2M^2. \tag{10.7}$$

Let us now expand the product using (10.5); this creates a very long product involving sixteen phases (coming from the terms $e(\phi(P(l,m)))$ in the definition of f) and fourty-eight cutoffs (coming from the terms $\psi(P(l,m))1_{1\leqslant l\leqslant L}1_{1\leqslant m\leqslant M}$). The sixteen phases $e(\phi(P(l,m)))$ combine to form a single phase

$$e\Big(\sum_{i_1,i_2,j_1,j_2\in\{0,1\}} (-1)^{i_1+i_2+j_1+j_2}\phi\big(P(l_0+i_1l_1+i_2l_2, m_0+j_1m_1+j_2m_2)\big)\Big).$$

The presence of the fourty-eight cutoffs is just what we need to apply Lemma 9.2, which allows us write the phase in (10.7) as

$$e\big(2l_1l_2m_1m_2\phi''(st,st)\big).$$

Note that the condition (9.3) required by that lemma is a consequence of the condition $\|st\|_g \leqslant \epsilon^2/LM$ we are working under here, provided that ε is chosen sufficiently small; indeed recall from (10.2) that $\rho_0 \gtrapprox 1$.

The fourty-eight cutoffs have now served their purpose of explicitly quartilinearising the phase, and we shall now set about obliterating them with further applications of the Cauchy-Schwarz inequality. To do this, we observe by inspection that fourty-seven of these cutoffs depend on at most three of the variables l_1, l_2, m_1, m_2, with the lone exception being $\psi(P(l_0 + l_1 + l_2, m_0 + m_1 + m_2))$. Also, let us recall once more that the cutoffs restrict l_1, l_2 to have magnitude at most L, and m_1, m_2 to have magnitude at most M. We thus have

$$\begin{aligned}(10.8)\quad \Big| \sum_{|l_1|,|l_2|\leqslant L|m_1|,|m_2|\leqslant M} & e(2l_1l_2m_1m_2\phi''(st,st)) \\ & \times \psi\big(P(l_0+l_1+l_2, m_0+m_1+m_2)\big) \\ & \times \mathbf{b}(l_2,m_1,m_2)\mathbf{b}(l_1,m_1,m_2)\mathbf{b}(l_1,l_2,m_2)\mathbf{b}(l_1,l_2,m_1)\Big| \gtrapprox L^2M^2.\end{aligned}$$

We would like to eliminate all the $\mathbf{b}()$ factors using Lemma A.10, but we need to deal with the exceptional cutoff $\psi(P(l_0+l_1+l_2, m_0+m_1+m_2))$ first. First observe that if ψ were a multiplicative function then the quadratic nature of P would ensure that $\psi(P(l_0 + l_1 + l_2, m_0 + m_1 + m_2))$ would factor into the product of expressions, each of which only depends on at most three (in fact, at most two) of the l_1, l_2, m_1, m_2. Of course, ψ is not multiplicative, but thanks to (8.1) we can write $\psi(n) = \Psi(g^n, n/N)$ for $N < n \leqslant 2N$, where $\Psi : G/\Gamma \times (\mathbb{R}/\mathbb{Z}) \to \mathbb{R}$ is Lipschitz on the orbit $\{(g^n, n/N) : N < n \leqslant 2N\}$ and hence, by Lemma A.8, is the restriction of a Lipschitz function on all of $G/\Gamma \times (\mathbb{R}/\mathbb{Z})$. Let $\delta \gtrapprox 1$ be a parameter to be chosen later. Using Lemma A.9, we can approximate Ψ uniformly to accuracy $O(\delta)$ on $(N, 2N]$ by a linear combination of at most $O(\delta^{-C})$ characters on $G/\Gamma \times (\mathbb{R}/\mathbb{Z})$, each of which has the form $(x, \theta) \mapsto \chi(x)e(k\theta)$ where $\chi \in (G/\Gamma)^*$ and $k \in \mathbb{Z}$. The coefficients in this linear combination are all $O(1)$. Thus we can estimate the left-hand side of (10.8) by

$$\begin{aligned} O\Bigg(\delta^{-C} \sup_{\substack{\chi\in(G/\Gamma)^*\\ k\in\mathbb{Z}}} \Big| \sum_{\substack{|l_1|,|l_2|\leqslant L\\ |m_1|,|m_2|\leqslant M}} & e(2l_1l_2m_1m_2\phi''(st,st)) \\ & \times \chi(g^{P(l_0+l_1+l_2,m_0+m_1+m_2)})e(kP(l_0+l_1+l_2,m_0+m_1+m_2)) \\ & \times \mathbf{b}(l_2,m_1,m_2)\mathbf{b}(l_1,m_1,m_2)\mathbf{b}(l_1,l_2,m_2)\mathbf{b}(l_1,l_2,m_1)\Big|\Bigg) + O(\delta L^2M^2).\end{aligned}$$

Choosing $\delta \gtrapprox 1$ suitably small, we thus conclude that there exist χ and k such that the inner sum is $\gtrapprox \delta^C L^2M^2 \gtrapprox L^2M^2$. By the quadratic nature of P we may absorb the terms $\chi(g^{P(l_0+l_1+l_2,m_0+m_1+m_2)})$ and $e(kP(l_0+l_1+l_2,m_0+m_1+m_2))$ into the four unspecified bounded functions $\mathbf{b}()$, thereby obtaining

$$\Big|\sum_{|l_1|,|l_2|\leqslant L|m_1|,|m_2|\leqslant M} e(2l_1l_2m_1m_2\phi''(st,st)) \times \mathbf{b}(l_2,m_1,m_2)\mathbf{b}(l_1,m_1,m_2)\mathbf{b}(l_1,l_2,m_2)\mathbf{b}(l_1,l_2,m_1)\Big| \gtrapprox L^2M^2.$$

Applying Lemma A.10 to eliminate the $\mathbf{b}()$ factors, we deduce

$$\Big|\sum_{\substack{|l_1|,|l_1'|,|l_2|,|l_2'|\leqslant L\\ |m_1|,|m_1'|,|m_2|,|m_2'|\leqslant M}} e(2(l_1-l_1')(l_2-l_2')(m_1-m_1')(m_2-m_2') \times \phi''(st,st))\Big| \gtrapprox L^4M^4.$$

By the pigeonhole principle, we can thus find $l_1', l_2' = O(L)$ and $m_1', m_2' = O(M)$ such that

$$\Big|\sum_{\substack{|l_1|,|l_2|\leqslant L\\ |m_1|,|m_2|\leqslant M}} e(2(l_1-l_1')(l_2-l_2')(m_1-m_1')(m_2-m_2')\phi''(st,st))\Big| \gtrapprox L^2M^2.$$

Summing in m_2 using (A.1), we obtain

$$\sum_{|l_1|,|l_2|\leqslant L}\sum_{|m_1|\leqslant M} \min\Big(1, \frac{1}{M}\|2(l_1-l_1')(l_2-l_2')(m_1-m_1')\phi''(st,st)\|_{\mathbb{R}/\mathbb{Z}}^{-1}\Big) \gtrapprox L^2M.$$

Shifting l_1, l_2, m_1 by l_1', l_2', m_1' respectively, and doubling m_1 to absorb the factor of two this creates, we thus have

$$\sum_{|l_1|,|l_2|\leqslant 2L}\sum_{|m_1|\leqslant 4M} \min\Big(1, \frac{1}{M\|l_1l_2m_1\phi''(st,st)\|_{\mathbb{R}/\mathbb{Z}}}\Big) \gtrapprox L^2M.$$

It follows that

$$\min\Big(1, \frac{1}{M\|l_1l_2m_1\phi''(st,st)\|_{\mathbb{R}/\mathbb{Z}}}\Big) \gtrapprox 1$$

for $\gtrapprox L^2M$ triples (l_1,l_2,m_1), which means that

$$\|l_1l_2m_1\phi''(st,st)\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{1}{M}$$

for those triples. In particular, we have $\gtrapprox L^2$ pairs (l_1,l_2) for which this inequality holds for $\gtrapprox M$ values of $m_1 = O(M)$. If $M \geqslant \log^{C_1(A+1)} N$ for some sufficiently large C_1 then we may apply Lemma A.4 (ii) with

parameters $\delta_1 \lessapprox 1/M$, $\delta_2 \gtrapprox 1$ and $|I| \sim M$ to conclude that for each such pair (l_1, l_2), there exists $q \lessapprox 1$ such that

$$\|l_1 l_2 q\phi''(st, st)\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{1}{M^2}. \tag{10.9}$$

This condition on M may be met by choosing $\varepsilon \gtrapprox 1$ sufficiently small, since one of the hypotheses of the lemma was that $M \geqslant 1/\varepsilon$. Applying the pigeonhole principle to (10.9), we can now locate a single $q \lessapprox 1$ such that the above bound holds for $\gtrapprox L^2$ pairs (l_1, l_2).

Taking ε sufficiently small we may assume that $L, M \geqslant \log^{C_2(A+1)} N$ for suitable C_2 and apply Lemma A.4 to l_2 instead of m_1. The parameters in that lemma are now $\delta_1 \approx 1/M^2$, $\delta_2 \approx 1$ and $|I| \sim L$, and we conclude the existence of $q' \lessapprox 1$ such that

$$\|l_1 q' q\phi''(st, st)\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{1}{LM^2}$$

for $\gtrapprox L$ values of l_1. Applying Lemma A.4 one last time, now with $\delta_1 \approx 1/LM^2$, $\delta_2 \approx 1$ and $|I| \sim L$, we find a $q'' \lessapprox 1$ such that

$$\|q'' q' q\phi''(st, st)\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{1}{L^2 M^2}.$$

Since $q'' q' q' \lessapprox 1$, the proof of Lemma 10.3 is complete. □

It remains to use this lemma to complete the proof of Proposition 10.1 in the Type II case. We take $\mathcal{D}$ to be simply the whole interval $[D/2X^{1/2}, D/X^{1/2}]$. There is a very important subtlety here: this set of integers can only be guaranteed to have size $\gtrapprox D/X^{1/2}$ if we assume that $D/X^{1/2} \gg 1$. Note, however, that in the Type II case this is so since we are working under that assumption that $D \gg N^{1/3}$ and $X \leqslant N^{1/10}$. This is not just a technical artefact of our approach — it is simply not possible to bound a general bilinear form, such as the Type II sum

$$T_{\mathrm{II}} = \sum_{d \sim D} \sum_{w \sim W} a_d b_w f(dw)$$

when one of the ranges $d \sim D$ or $w \sim W$ is too short, as the weights a_d, b_w could conspire to give no cancellation.

Suppose, then, that $d \in \mathcal{D}$ and that $w \in \mathbb{Z}$ satisfies the condition of Proposition 10.1, namely that $\|dw\|_g \leqslant 1/X$. In Lemma 10.3 take $L = M := \varepsilon X^{1/2}/10$ and $s := d$, $t := w$. If $X_0 \lessapprox 1$ is sufficiently large and $X > X_0$ then certainly the two conditions $L, M \geqslant 1/\varepsilon$ are satisfied. Furthermore we have

$$L|s| \leqslant \frac{\varepsilon X^{1/2}}{10} \cdot \frac{D}{X^{1/2}} \leqslant \varepsilon D$$

and

$$M|t| = \frac{\varepsilon X^{1/2}}{10} \cdot \frac{|dw|}{|d|} \leqslant \frac{\varepsilon X^{1/2}}{10} \cdot \frac{N\|dw\|_g}{|d|} \leqslant \frac{\varepsilon N}{5D} \leqslant \varepsilon W,$$

and finally $\|st\|_g \leqslant \varepsilon^2/LM$ by the definition of L and M. All the conditions of Lemma 10.3 are thus satisfied, and we may infer that

$$\|\phi''(dw, dw)\|_{\mathbb{R}/\mathbb{Z},Q} \lessapprox X^{-2}$$

for some $Q \lessapprox 1$, as required.

□

We may now forget about Type I and II sums, and work with the conclusion of Proposition 10.1 instead. In the next section we will use divisor moment estimates to cast this conclusion in a more tractible form.

11. Massaging the major arc condition

In the last two sections we showed that if $\psi(n)e(-\phi(n))$ correlates with Möbius (specifically if (10.1) holds true) then ϕ must exhibit some kind of "major arc" behaviour. Indeed we proved Proposition 10.1, which we urge the reader to recall now. Our first task in this section is to cast the conclusion of that proposition in a more useable form. Through this section, we assume that $\phi : B_g(n_0, 100\rho_0) \to \mathbb{R}/\mathbb{Z}$ is a phase for which (10.1), and hence the conclusion of Proposition 10.1, holds true.

PROPOSITION 11.1. — *Let ϕ be as above, and suppose that the parameter ρ_1 satisfies*

$$N^{-c} < \rho_1 < \rho_0 \log^{-C_1(A+1)} N \tag{11.1}$$

for some $c > 0$ and some C_1 which is sufficiently large depending on G/Γ (in reality ρ_1 will be much larger than N^{-c}, so the lower bound here is hardly relevant). Then

$$\|\phi''(n, n)\|_{\mathbb{R}/\mathbb{Z},Q} \lessapprox \rho_1^2 \tag{11.2}$$

for $\gtrsim \rho_1^{3/2}|B_g(0, \rho_1)|$ values of $n \in B_g(0, \rho_1)$, where $Q \lessapprox 1$.

Remarks. — Note that since $\rho_1^{3/2}$ is so much bigger than ρ_1^2, the conclusion is in the spirit of the hypotheses of Lemmas such as A.12, where a quadratic whose fractional part was "close to zero unexpectedly often" was shown to be major arc. We will, in fact, apply exactly that lemma later in this section. The fact that we can arrange the exponents 3/2 and 2 in this way is ultimately due to the lower bound $|\mathcal{D}| \gtrapprox D/X^{1/2}$ in Proposition 10.1; $|\mathcal{D}| \gtrapprox D/X$ would not suffice.

Proof. — Set $X := 1/\rho_1$ in Proposition 10.1; we may certainly suppose that C_1 is so large that this is permissible. We find $D \ll N^{2/3}$ and a set $\mathcal{D} \subseteq \{1, \dots, D\}$ of cardinality $|\mathcal{D}| \gtrapprox D/X^{1/2}$ such that

$$\|\phi''(dw, dw)\|_{\mathbb{R}/\mathbb{Z}, Q} \lessapprox \rho_1^2$$

whenever $d \in \mathcal{D}$ and $w \in \mathbb{Z}$ are such that $dw \in B_g(0, \rho_1)$. Thus, if we define the sets

$$\Omega := B_g(0, \rho_1) \cap \mathbb{Z}^+; \quad \Omega_d := \{n \in \Omega : d|n\}$$

for each integer $d > 1$, it will suffice (noting that $B_g(0, \rho_0)$ is symmetric about the origin) to prove the estimate

$$\Big|\bigcup_{d \in \mathcal{D}} \Omega_d\Big| \gtrapprox \rho_1^{3/2}|\Omega|. \tag{11.3}$$

Observing from Lemma 6.2 that

$$|\Omega| \gg \rho_1^C N \text{ and } |\Omega_d| \gg \frac{1}{D}|\Omega| \text{ for all } d \in \mathcal{D},$$

where C depends only on G/Γ, it follows by taking $\kappa := 1/2C$ in Lemma C.2 of Appendix C that

$$\Big|\bigcup_{d \in \mathcal{D}} \Omega_d\Big| \gg \frac{|\mathcal{D}|^2}{|D|^2}|\Omega|\rho_1^{1/2} \log^{-C_2} N.$$

Since $|\mathcal{D}| \gtrapprox D/X^{1/2}$, the result follows immediately. □

As we remarked, the conclusion of Proposition 11.1 has the form "$\phi''(n, n)$ is surprisingly close to an integer very often" on a small Bohr set $B_g(0, \rho_1)$. The next step is to amplify this to obtain $\phi''(h, h)$ major arc for a significantly larger set of h (working on $B_g(0, \rho_0)$ rather than $B_g(0, \rho_1)$). More precisely, we now establish a more pleasant characterisation of major arc:

Lemma 11.2 (Major arcs have small second derivative). — *Let ϕ be as above. Then there exists $Q_1 \lessapprox 1$ such that*

$$\|\phi''(h, h)\|_{\mathbb{R}/\mathbb{Z}, Q_1} \lessapprox \|h\|_g^2$$

for all $h \in B_g(0, \rho_0)$.

Proof. — The idea is to make the quadratic structure of ϕ'' so explicit that we can apply Lemma A.12.

We will choose $\rho_1 := \log^{-C_2(A+1)} N$, where $C_2 \gg C_1$ is some constant to be specified later. In particular if C_2 is large enough then the conditions of Proposition 11.1 are satisfied, and we can find as a result some set $\mathcal{S} \subseteq B_g(0, \rho_1)$ such that

$$|\mathcal{S}| \gtrapprox \rho_1^{3/2}|B_g(0, \rho_1)| \tag{11.4}$$

and

$$\|\phi''(n,n)\|_{\mathbb{R}/\mathbb{Z},Q} \lessapprox \rho_1^2 \tag{11.5}$$

for all $n \in \mathcal{S}$. Note that the implied constants in the $\gtrapprox$ and $\lessapprox$ notations here *do not depend on* C_2. Note also that for reasons like this one must exercise extreme caution with these notations.

Select some $C_3 \gg C_2$. If $\|h\|_g \geqslant \log^{-C_3(A+1)} N$ then the lemma holds vacuously, and so we assume henceforth that $\|h\|_g \leqslant \log^{-C_3(A+1)} N$. Now from (11.4) and Lemma 6.2(b) we have

$$\mathbb{E}_{n \in B_g(0,2\rho_1)} 1_{\mathcal{S}}(n+m) \gtrapprox \rho_1^{3/2}$$

for all $m \in B_g(0,\rho_1)$. Applying this to $m = hl$ for all $l \in \{1,\dots,L\}$, where $L := \lfloor \frac{\rho_1}{\|h\|_g} \rfloor$, and then averaging in L, we conclude

$$\mathbb{E}_{n \in B_g(0,2\rho_1)} \mathbb{E}_{1 \leqslant l \leqslant L} 1_{\mathcal{S}}(n+hl) \gtrapprox \rho_1^{3/2},$$

and thus by the pigeonhole principle we can find $n \in B_g(0,2\rho_1)$ such that

$$\mathbb{E}_{1 \leqslant l \leqslant L} 1_{\mathcal{S}}(n+hl) \gtrapprox \rho_1^{3/2}.$$

In particular, we have

$$\|\phi''(n+hl,n+hl)\|_{\mathbb{R}/\mathbb{Z},Q} \lessapprox \rho_1^2$$

for $\gtrapprox \rho_1^{3/2} L$ values of $l \in \{1,\dots,L\}$. Applying the pigeonhole principle again, we can thus find a single $q \lessapprox 1$ such that

$$\|q\phi''(n+hl,n+hl)\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \rho_1^2$$

for $\gtrapprox \rho_1^{3/2} L$ values of $l \in \{1,\dots,L\}$. Now from Corollary 9.1 (and (10.2)) we can write

$$q\phi''(n+hl,n+hl) = ql^2\phi''(h,h) + \alpha l + \beta$$

for some quantities $\alpha, \beta \in \mathbb{R}/\mathbb{Z}$ which depend on q, ϕ, n, h but are independent of l. Thus

$$\|ql^2\phi''(h,h) + \alpha l + \beta\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \rho_1^2$$

for $\gtrapprox \rho_1^{3/2} L$ values of $l \in \{1,\dots,L\}$. Now Lemma A.12 applies to exactly this kind of situation. In that lemma we take $\delta_1 \approx \rho_1^2$ and $\delta_2 \approx \rho_1^{3/2}$, and note that the requisite conditions $\delta_1 \leqslant \frac{1}{4}\delta_2$ and $L \geqslant 2^{58}\delta_2^{-12}$ are handsomely satisfied if C_2, C_3 are chosen judiciously. The conclusion is that

$$\|q\phi''(h,h)\|_{\mathbb{R}/\mathbb{Z},2^{-43}\delta_2^{-9}} \leqslant 2^{141}\delta_2^{-28}L^{-2}.$$

Setting $Q_1 := 2^{-43}\delta_2^{-9} \lessapprox 1$ and noting that $L \gtrapprox \|h\|_g^{-1}$, the conclusion follows. □

In the next lemma, we bootstrap Lemma 11.2 to a depolarized version of itself.

LEMMA 11.3 (Major arcs have small second derivative, II). — *Let ϕ be as above. Then there exist $Q_2 \lessapprox 1$ and $\rho_2 \gtrapprox 1$ such that $\rho_2 \leqslant \rho_0$ and*

$$\|\phi''(h,h')\|_{\mathbb{R}/\mathbb{Z},Q_2} \lessapprox \|h\|_g \|h'\|_g$$

for all $h, h' \in B_g(0,\rho_2)$.

Proof. — Let $\rho_2 = \log^{-C_4(A+1)} N$, for some large C_4 to be chosen later. By symmetry we may assume $\|h'\|_g \leqslant \|h\|_g$. Let $L > 1$ be the least integer such that $L\|h'\|_g > \|h\|_g$. For any $l \in \{1,\dots,L\}$, we use (9.2) and the hypotheses $h, h' \in B_g(0,\rho_2)$ to conclude

$$4l\phi''(h,h') = \phi''(h+lh', h+lh') - \phi''(h-lh', h-lh').$$

Applying Lemma 11.2 and the triangle inequality, we infer

$$\|4l\phi''(h,h')\|_{\mathbb{R}/\mathbb{Z},Q_1} \lessapprox \|h\|_g^2$$

and hence

$$\|l\phi''(h,h')\|_{\mathbb{R}/\mathbb{Z},4Q_1} \lessapprox \|h\|_g^2$$

for all $l \in \{1,\dots,L\}$. Let C_5 be a further constant to be specified later. If $L \leqslant \log^{-C_5(A+1)} N$ then we can set $l = 1$ and the argument is finished. Suppose, then, that $L \geqslant \log^{-C_5(A+1)} N$. By the pigeonhole principle, we can find $q \leqslant Q_1 \lessapprox 1$ such that

$$\|ql\phi''(h,h')\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \|h\|_g^2$$

for $\gtrapprox L$ values of $l \in \{1,\dots,L\}$. We are now in a position to apply Lemma A.4(ii) with $\delta_1 \approx \|h\|_g^2$ and $\delta_2 \approx 1$. If C_4 is large enough then we certainly have $\delta_1 \leqslant \frac{1}{4}\delta_2$, whilst C_5 may be chosen so that $L > 2/\delta_2^2$. In those circumstances the lemma is applicable and we deduce that

$$\|q\phi''(h,h')\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \|h\|_g^2/L \leqslant \|h\|_g\|h'\|_g.$$

This concludes the proof. □

The above lemma says that for any pair h, h' each having small $\|\cdot\|_g$ norms, the second derivative $\phi''(h,h')$ is close to a rational number a/q for some small q. However, this q can currently depend on h, h'. Fortunately, it is possible to "clear denominators" and make q independent of h, h', by taking advantage of a certain "finite dimensionality" of the Bohr set $B_g(0,\rho_2)$. More precisely, we have

LEMMA 11.4 (Major arcs have small second derivative, III). — *Let ϕ be as above. Then there exists $\rho_3 \gtrapprox 1$ and an integer $q \lessapprox 1$ such that*

$$\|q\phi''(h,h')\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \|h\|_g \|h'\|_g$$

for all $h, h' \in B_g(0,\rho_3)$.

Proof. — We shall use some standard results from the geometry of numbers to obtain a "basis" for the Bohr set $B_g(0,\rho_2)$. These result are discussed in several places: see, for example, [2, 11] and [20, Ch. 3]. Recall at this point the discussion at the end of §8, where we remarked that $g \in (\mathbb{R}/\mathbb{Z})^d$ can be taken to be an p^{th} root of unity, where $p \in [10N, 20N]$ is the prime we have associated to N for those arguments where it is convenient to work in a cyclic group. This is such an argument. We identify $B_g(0,\rho_2)$, which is certainly contained in $\{1,\dots,N\}$, with a subset of $\mathbb{Z}/p\mathbb{Z}$. Write

$$g = \Big(\frac{\xi_1}{p},\dots,\frac{\xi_d}{p}\Big)$$

in $(\mathbb{R}/\mathbb{Z})^d$, where $\xi_1,\dots,\xi_d \in \mathbb{Z}/p\mathbb{Z}$. Let $S \subseteq \mathbb{Z}/p\mathbb{Z}$ be the set of frequencies

$$S := \{1, \xi_1, \dots, \xi_d\}.$$

In the notation of [11], the Bohr set $B_g(0,\rho_2)$ is then comparable to a "traditional" Bohr set

$$B(S,\rho) := \big\{x \in \mathbb{Z}/p\mathbb{Z} : \|\xi x/p\|_{\mathbb{R}/\mathbb{Z}} < \rho\big\}$$

in the sense that

$$B(S, \tfrac{1}{20}\rho_2) \subseteq B_g(0,\rho_2) \subseteq B(S, 2\rho_2). \tag{11.6}$$

Applying [11, Corollary 10.5], and redefining $d := d+1$, we can then find a proper[8] generalised arithmetic progression

$$P = \big\{l_1 v_1 + \ldots + l_d v_d : |l_j| \leqslant L_j \text{ for all } 1 \leqslant j \leqslant d\big\}$$

for some $L_1,\dots,L_d \geqslant 1$ and $v_1,\dots,v_d \in \mathbb{Z}/p\mathbb{Z}$, such that

$$B_g(0, c\rho_2) \subseteq P \subseteq B_g(0,\rho_2)$$

for some $c = c(d) > 0$. In fact by applying that result to $B_g(0, \frac{1}{4}\rho_2)$ (and redefining P and the L_j slightly) we may insist on the slightly stronger inclusions

$$B_g(0, \tfrac{1}{4}c\rho_2) \subseteq P_{1/4} \subseteq P \subseteq B_g(0,\rho_2) \tag{11.7}$$

[8] By *proper* we mean that all the sums $l_1 v_1 + \ldots + l_d v_d$ are distinct.

where P_θ is defined for any $\theta \in (0,1]$ by

$$P_\theta := \{l_1 v_1 + \ldots + l_d v_d : |l_j| \leqslant \theta L_j \text{ for all } 1 \leqslant j \leqslant d\}.$$

We will prove the lemma with $\rho_3 := \frac{1}{4} c\rho_2$. Let us note from (11.7) that

$$\|v_j\|_g \leqslant \frac{\rho_2}{L_j} \leqslant \frac{1}{L_j}$$

for each j, $1 \leqslant j \leqslant d$. Thus by Lemma 11.3 we may find for each j, j', $1 \leqslant j, j' \leqslant d$, a $q_{j,j'} \lessapprox 1$ such that

$$\|q_{j,j'}\phi''(v_j, v_{j'})\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{1}{L_j L_{j'}}.$$

If we let q be the least common multiple of all the $q_{j,j'}$, then we still have $q \lessapprox 1$ and

$$\|q\phi''(v_j, v_{j'})\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \frac{1}{L_j L_{j'}}$$

for all j, j', $1 \leqslant j, j' \leqslant d$. Note that at this point the implied constants in the $\lessapprox$ notation have become heavily dependent on d. By bilinearity (9.2) it follows that

$$\|q\phi''(h, h')\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \|h\|_P \|h'\|_P \tag{11.8}$$

for all $h, h' \in P$, where the norm $\|\cdot\|_P$ on P is defined by

$$\|l_1 v_1 + \ldots + l_d v_d\|_P := \sup_{1 \leqslant j \leqslant d} \frac{|l_j|}{L_j}.$$

We claim that $\|h\|_P \lessapprox \|h\|_g$ for all $h \in B_g(0, \rho_3)$. In view of (11.8), this will suffice to prove the lemma.

We may assume that $h \neq 0$ since the claim is trivial otherwise. Observe that $h \in P_{1/2}$. Let $M > 1$ be the smallest positive integer such that $Mh \notin P_{1/4}$; since $Mh = (M-1)h + h$, we see that $Mh \in P_{1/2}$. Thus $\|Mh\|_P \leqslant 1/2$, which implies that $\|h\|_P \leqslant 1/2M$ (here we use the hypothesis that P is proper, which implies that the co-ordinates $l_1, \ldots, l_d$ of Mh are M times the co-ordinates of h). On the other hand, since $Mh \notin P_{1/4}$, we have $Mh \notin B_g(0, \rho_3)$, which implies $M\|h\|_g \geqslant \rho_3$ and hence that $\|h\|_g \geqslant \rho_3/M \gtrapprox 1/M$. Combining these estimates we obtain the claim, and hence the lemma. □

12. Handling the major arcs

Let us summarise the current state of affairs. In our effort to prove Proposition 8.1, we assumed that its conclusion (8.2) was false. After a long and

complicated analysis, we deduced from this assumption that the phase ϕ is *major arc*, in the sense that we have an estimate

$$\|q\phi''(h,h')\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \|h\|_g\|h'\|_g$$

whenever $h, h' \in B_g(0,\rho_3)$, for some $q \lessapprox 1$ and some $\rho_3 \gtrapprox 1$. This was, of course, the content of Lemma 11.4. To close the argument, we relate major arc phases of this type to those appearing in Proposition 6.3. This is not hard (though a little technical), and leads quickly to a contradiction (of the assumption that (8.2) was false).

Let q be as above. By bilinearity (9.2) again, we see that

$$\|\phi''(h,h')\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \|h\|_g\|h'\|_g$$

for all $h, h' \in B_g(0,\rho_3)$ such that $q|h,h'$. Let $\varepsilon \ll \rho_3$, $\varepsilon \gtrapprox 1$, be a small number to be chosen later. Applying (9.1), we conclude the approximate linearity relationship

$$\|\phi(n+h_1+h_2)-\phi(n+h_1)-\phi(n+h_2)+\phi(n)\|_{\mathbb{R}/\mathbb{Z}} \lessapprox \varepsilon^2 \ll \epsilon \tag{12.1}$$

whenever $n \in B_g(n_0,2\rho_0)$, whenever $h_1,h_2 \in B_g(0,20\varepsilon)$ are such that $q|h_1,h_2$, and provided that ε is small enough.

Now due to the finite dimensionality of the space $(\mathbb{R}/\mathbb{Z})^d \times \mathbb{R}$ from which the metric $\|n-m\|_g$ is naturally descended (*cf.* the remarks following Definition 6.1) we may cover $B_g(n_0,\rho_0)$ with $O(\varepsilon^{-C})$ Bohr sets $B_g(n_\alpha,\varepsilon)$ such that each point is contained in $O(1)$ of these Bohr sets. This induces a corresponding partition of ψ into $O(\varepsilon^{-C})$ functions ψ_α, each of which is supported on a Bohr set $B_g(n_\alpha,\varepsilon)$ and still obeys the Lipschitz bound (8.1).

Now observe that if $n, n+h_1, n+h_2, n+h_1+h_2 \in B_g(n_\alpha,10\varepsilon)$, and if $q|h_1,h_2$ then (12.1) holds. Thus we may apply Proposition 6.3 (with $\rho=\varepsilon$) to conclude that for any $\kappa \leqslant \epsilon$, and for some A' to be chosen later, we have

$$\left|\mathbb{E}_{N<n\leqslant 2N}\mu(n)\psi_\alpha(n)e(-\phi(n))\right| \ll_{A'} \kappa^{-C}q^3\log^{-A'}N + (\varepsilon+\kappa)\mathbb{E}_{N<n\leqslant 2N}|\psi_\alpha|.$$

Summing in α, using the bounded overlap of the Bohr sets and the fact that $\|\psi\|_\infty \ll \rho_0 \ll 1$, we conclude

$$\left|\mathbb{E}_{N<n\leqslant 2N}\mu(n)\psi(n)e(-\phi(n))\right| \ll_{A'} (\varepsilon\kappa)^{-C}q^3\log^{-A'}N + \varepsilon + \kappa. \tag{12.2}$$

At this point we set[(9)] $\kappa = \varepsilon = \log^{-C(A+1)}N$ for some $C>1$ which is so large that (12.1) holds. Recalling that $q \lessapprox 1$, we see that A' may be chosen so that the right-hand side of (12.2) is $\ll \log^{-A}N$.

[(9)] We kept the parameters ε and κ separate in Proposition 6.3 for pedagogical reasons, to make the dependencies clear.

We have, at long last, contradicted the supposition that (8.2) is false. This implies Proposition 8.1. By the analysis of §8, Theorem 2.2 is also true, and thus, by the deduction immediately after the statement of Theorem 2.2, so is the Main Theorem.

Appendix A. Some harmonic analysis tools

In this appendix we collect some simple harmonic analysis tools which are used frequently in the paper. We begin by introducing some norms on the unit circle $\mathbb{R}/\mathbb{Z}$, which can be lifted up to the real line $\mathbb{R}$.

Definition A.1 (Circle norms). — *If α is an element of the real line $\mathbb{R}$ or the circle $\mathbb{R}/\mathbb{Z}$, we use $\|\alpha\|_{\mathbb{R}/\mathbb{Z}}$ to denote the distance from α to the nearest integer (if α is real) or to zero (if α is on the circle $\mathbb{R}/\mathbb{Z}$). If $Q \geqslant 1$ is an integer, we use $\|\alpha\|_{\mathbb{R}/\mathbb{Z},Q}$ to denote the quantity*

$$\|\alpha\|_{\mathbb{R}/\mathbb{Z},Q} := \inf_{1 \leqslant q \leqslant Q} \|q\alpha\|_{\mathbb{R}/\mathbb{Z}}.$$

The quantity $\|\alpha\|_{\mathbb{R}/\mathbb{Z}}$ is subadditive, thus $\|\alpha+\beta\|_{\mathbb{R}/\mathbb{Z}} \leqslant \|\alpha\|_{\mathbb{R}/\mathbb{Z}} + \|\beta\|_{\mathbb{R}/\mathbb{Z}}$. We caution however that the quantity $\|\alpha\|_{\mathbb{R}/\mathbb{Z},Q}$ (which is large when α lies in a "minor arc", and small when α lies in a "major arc") is *not* subadditive.

Define a *discrete interval* to be any set of the form $\{n \in \mathbb{Z} : a \leqslant n \leqslant b\}$ for some a, b. By summing the geometric series, we observe the elementary exponential sum estimate

$$\Big|\sum_{n \in I} e(\alpha n)\Big| \leqslant 4 \min\Big(|I|, \frac{1}{\|\alpha\|_{\mathbb{R}/\mathbb{Z}}}\Big) \tag{A.1}$$

for any discrete interval $I \subset \mathbb{Z}$ and any $\alpha \in \mathbb{R}/\mathbb{Z}$ (or any $\alpha \in \mathbb{R}$). One consequence of this is the following Pólya-Vinogradov type completion of sums lemma, which allows one to estimate a partial sum by a completed sum at the cost of a logarithm and an exponential phase.

Lemma A.2 (Completion of sums). — *Let $I \subset \mathbb{Z}$ be a discrete interval, and $f : \mathbb{Z} \to \mathbb{C}$ be a function. Then we have*

$$\sup_{J \subseteq I} \Big|\sum_{n \in J} f(n)\Big| \ll \log(1+|I|) \sup_{\alpha \in \mathbb{R}/\mathbb{Z}} \Big|\sum_{n \in I} f(n) e(\alpha n)\Big|$$

where the supremum on the left ranges over discrete sub-intervals of I. More generally, if $I' \subset \mathbb{Z}$ is another discrete interval, and $K : \mathbb{Z} \times \mathbb{Z} \to \mathbb{C}$ is a function, then we have

$$\sum_{m \in I'} \Big|\sum_{n \in I} 1_{J_m}(n) K(n,m)\Big|^2 \ll \log^2(1+|I|) \sup_{\alpha \in \mathbb{R}/\mathbb{Z}} \sum_{m \in I'} \Big|\sum_{n \in I} K(n,m) e(\alpha n)\Big|^2$$

where for each $m \in I'$, $J_m \subset \mathbb{Z}$ is an arbitrary discrete interval.

Proof. — We may assume I is non-empty. By translation we may take $I = \{1, \dots, L\}$ for some $L \geqslant 1$, which we then identify with $\mathbb{Z}/L\mathbb{Z}$. If J is any interval in $\mathbb{Z}/L\mathbb{Z}$, we can use Fourier expansion in $\mathbb{Z}/L\mathbb{Z}$ to write

$$\begin{aligned}\sum_{n\in J} f(n) &= \sum_{n\in\mathbb{Z}/L\mathbb{Z}} 1_J(n) f(n) \\ &= \sum_{\xi\in\mathbb{Z}/L\mathbb{Z}} \widehat{1_J}(\xi) \sum_{n\in\mathbb{Z}/L\mathbb{Z}} e(n\xi/L) f(n)\end{aligned}$$

where

$$\widehat{1_J}(\xi) := \mathbb{E}_{n\in\mathbb{Z}/L\mathbb{Z}} 1_J(n) e(-n\xi/L).$$

Applying (A.1), we have

$$|\widehat{1_J}(\xi)| \leqslant 4 \min(1, \frac{1}{L\|\xi/L\|_{\mathbb{R}/\mathbb{Z}}}).$$

Thus by the triangle inequality, we have

$$\begin{aligned}\Big|\sum_{n\in J} f(n)| &\leqslant 4 \sum_{\xi\in\mathbb{Z}/L\mathbb{Z}} \min\Big(1, \frac{1}{L\|\xi/L\|_{\mathbb{R}/\mathbb{Z}}}\Big) \Big| \sum_{n\in\mathbb{Z}/L\mathbb{Z}} e(n\xi/L) f(n)\Big| \\ &\ll \sum_{\xi\in\mathbb{Z}/L\mathbb{Z}} \min\Big(1, \frac{1}{L\|\xi/L\|_{\mathbb{R}/\mathbb{Z}}}\Big) \sup_{\alpha\in\mathbb{R}/\mathbb{Z}} \Big|\sum_{n\in I} e(n\alpha) f(n)\Big| \\ &\ll \log(1+L) \sup_{\alpha\in\mathbb{R}/\mathbb{Z}} \Big|\sum_{n\in I} e(n\alpha) f(n)\Big|,\end{aligned}$$

which gives the first inequality. Using similar arguments, as well as the triangle inequality in l^2, we have

$$\begin{aligned}&\Big(\sum_{m\in I'} \Big|\sum_{n\in I} 1_{J_m}(n) K(n,m)\Big|^2\Big)^{1/2} \\ &\quad\ll \Big(\sum_{m\in I'} \Big(\sum_{\xi\in\mathbb{Z}/N\mathbb{Z}} \min(1, \frac{1}{L\|\xi/L\|_{\mathbb{R}/\mathbb{Z}}}) \Big|\sum_{n\in I} e(n\xi/L) K(n,m)\Big|\Big)^2\Big)^{1/2} \\ &\quad\ll \sum_{\xi\in\mathbb{Z}/N\mathbb{Z}} \min\Big(1, \frac{1}{L\|\xi/L\|_{\mathbb{R}/\mathbb{Z}}}\Big) \Big(\sum_{m\in I'} \Big|\sum_{n\in I} e(n\xi/L) K(n,m)\Big|^2\Big)^{1/2}\end{aligned}$$

$$\ll \log(1+L) \sup_{\alpha \in \mathbb{R}/\mathbb{Z}} \Big(\sum_{m \in I'} \Big| \sum_{n \in I} e(\alpha \xi) K(n,m) \Big|^2 \Big)^{1/2}.$$

□

In a similar spirit, we now recall the well-known Erdős-Turán inequality:

PROPOSITION A.3 (Erdős-Turán inequality). — *Let $(u_l)_{l=1}^L$ be a sequence in $\mathbb{R}/\mathbb{Z}$, and define the* discrepancy $\Delta(\alpha,\beta)$ *for any $-\frac{1}{2} \leqslant \alpha < \beta < \frac{1}{2}$ by the formula*

$$\Delta(\alpha,\beta) := \#\{l \in \{1,\dots,L\} : u_l \in [\alpha,\beta]\} - (\beta-\alpha)L.$$

Then for any positive integer Q we have

$$|\Delta(\alpha,\beta)| \leqslant \frac{L}{Q} + 3\sum_{q=1}^{Q} \frac{1}{q} \Big| \sum_{l=1}^{L} e(q u_l) \Big|.$$

Proof. — See for instance [17]. The constant 3 is unimportant for us, and could be improved slightly. □

An important application of this inequality for us (which we will use extremely frequently) will be the following observation, which says that if a linear sequence αl stays close to an integer for many l in an interval I, then α must be "major arc", in the sense that $\|\alpha\|_{\mathbb{R}/\mathbb{Z},Q}$ is small for some small Q.

LEMMA A.4 (Recurrent linear functions are major arc). — *Let $I \subseteq \mathbb{Z}$ be a discrete interval, let $\alpha \in \mathbb{R}/\mathbb{Z}$, and suppose that the set*

$$\mathfrak{L} := \big\{ l \in I : \|\alpha l\|_{\mathbb{R}/\mathbb{Z}} \leqslant \delta_1 \big\}$$

has cardinality at least $\delta_2|I|$ for some $0 < \delta_1, \delta_2 < 1$ with $\delta_1 \leqslant \frac{1}{4}\delta_2$.

(i) *If $|I| > 1/\delta_2$, then $\|\alpha\|_{\mathbb{R}/\mathbb{Z},8/\delta_2} \leqslant 2^8/\delta_2^2|I|$.*
(ii) *If $|I| > 2/\delta_2^2$, then $\|\alpha\|_{\mathbb{R}/\mathbb{Z},16/\delta_2^2} \leqslant 2^{15}\delta_1/\delta_2^6|I|$.*

Proof. — Write $I = \{M+1,\dots,M+L\}$, and let $(u_l)_{l=1}^L$ be the sequence $u_l := \alpha(M+l) \pmod 1$. Then the lower bound on $\mathfrak{L}$ implies the discrepancy estimate

$$\Delta(-\delta_1,\delta_1) \geqslant (\delta_2 - 2\delta_1)L \geqslant \tfrac{1}{2}\delta_2 L.$$

Let us now prove (i). Applying Proposition A.3 we conclude

$$\tfrac{1}{2}\delta_2 L \leqslant \frac{L}{Q} + 3\sum_{q=1}^{Q} \frac{1}{q} \Big| \sum_{l=1}^{L} e(q u_l) \Big|$$

for any Q. Taking $Q =: \lceil 4/\delta_2 \rceil$, this implies that there is $q \leqslant 8/\delta_2$ such that

$$\Big|\sum_{l=1}^{L} e(qu_l)\Big| \geqslant 2^{-6}\delta_2^2 L.$$

Applying (A.1), the result follows.

We now use a standard "amplification" argument, exploiting the smallness of δ_1 compared to δ_2, to bootstrap (i) to the stronger estimate (ii). We may assume that $\delta_1 < \delta_2^2/16$ since the result follows immediately from (i) otherwise. Let $1 \leqslant m \leqslant L$ be an integer to be chosen later; then by the pigeonhole principle and the lower bound on $|\mathfrak{L}|$, there exists some b such that the set

$$\mathfrak{L}_b := \{b+1, \dots, b+m\} \cap \mathfrak{L}$$

has cardinality at least $\delta_2 m/2$. We fix b, and note that if $x \in m\mathfrak{L} + \mathfrak{L}_b$, that is to say if $x = ml + l'$ with $l \in \mathfrak{L}$ and $l' \in \mathfrak{L}_b$, then $\|\alpha x\|_{\mathbb{R}/\mathbb{Z}} \leqslant 2m\delta_1$. Furthermore we have $|m\mathfrak{L} + \mathfrak{L}_b| \geqslant \delta_2^2 mL/2$, and also $m\mathfrak{L} + \mathfrak{L}_b$ is a subset of the interval

$$I' := \big\{m(M+1) + b + 1, \dots, m(M+L) + b + m\big\},$$

which has cardinality at most mL. We can apply (i) with I, δ_1, δ_2 replaced by I', $2m\delta_1$, and $\delta_2^2/2$, provided that $m \leqslant \delta_2^2/16\delta_1$ and $mL > 2/\delta_2^2$. It being sensible to take m essentially as large as possible, set $m := \lfloor \delta_2^2/16\delta_1 \rfloor$. The result follows quickly. □

Next, we record a version of summation by parts. Define the *total variation* $\|\psi\|_{\mathrm{TV}}$ of a sequence $\psi : \mathbb{Z} \to \mathbb{C}$ to be the quantity

$$\|\psi\|_{\mathrm{TV}} := \sup_{n\in\mathbb{Z}} |\psi(n)| + \sum_{n\in\mathbb{Z}} |\psi(n+1) - \psi(n)|,$$

and more generally define the *total variation modulo* q for any $q \geqslant 1$ to be the quantity

$$\|\psi\|_{\mathrm{TV},q} := \sup_{n\in\mathbb{Z}} |\psi(n)| + \sum_{n\in\mathbb{Z}} |\psi(n+q) - \psi(n)|.$$

LEMMA A.5 (Summation by parts). — *If f, $\psi : \mathbb{Z} \to \mathbb{C}$ and I is an interval, then*

$$\Big|\sum_{n\in I} f(n)\psi(n)\Big| \leqslant \|\psi\|_{\mathrm{TV}} \sup_{J\subseteq I} \Big|\sum_{n\in J} f(n)\Big|.$$

More generally, for any $q \geqslant 1$ we have

$$\Big|\sum_{n\in I} f(n)\psi(n)\Big| \leqslant q\|\psi\|_{\mathrm{TV},q} \sup_{J\subseteq I, a\in\mathbb{Z}/q\mathbb{Z}} \Big|\sum_{n\in J} f(n)1_{n=a \pmod q}\Big|.$$

Proof. — Write $I = \{u, \dots, v\}$, and denote by $S_n := \sum_{j=u}^{n} f(j)$ the partial sums of f. Recalling the summation by parts formula

$$\sum_{n\in I} f(n)\psi(n) = S_v\psi(v) + \sum_{n=u}^{v-1} S_n(\psi(n) - \psi(n+1)),$$

the first inequality follows immediately. The second bound follows by splitting I into q residue classes modulo q and applying a rescaled version of the first identity to each component. □

Corollary A.6 (Completion of sums, II). — *Let $I \subset \mathbb{Z}$ be a discrete interval, and $f : \mathbb{Z} \to \mathbb{C}$ and $\psi : \mathbb{Z} \to \mathbb{C}$ be functions. Then we have*

$$\sum_{n\in I} \psi(n)f(n) \ll \log(1+|I|)\|\psi\|_{\mathrm{TV}} \sup_{\alpha\in\mathbb{R}/\mathbb{Z}} \Big|\sum_{n\in I} f(n)e(\alpha n)\Big|$$

and more generally for any $q \geqslant 1$

$$\sum_{n\in I} \psi(n)f(n) \ll q\log(1+|I|)\|\psi\|_{\mathrm{TV},q} \sup_{\alpha\in\mathbb{R}/\mathbb{Z}} \Big|\sum_{n\in I} f(n)e(\alpha n)\Big|.$$

Proof. — The first part is immediate from Lemmas A.2 and A.5. To obtain the second bound, we begin with an invocation of the second bound in Lemma A.5. It is now sufficient to prove that

$$\sup_{J\subseteq I, a\in\mathbb{Z}/q\mathbb{Z}} \Big|\sum_{n\in J} f(n)1_{n\equiv a \pmod q}\Big| \leqslant \sup_{\alpha\in\mathbb{R}/\mathbb{Z}} \Big|\sum_{n\in I} f(n)e(\alpha n)\Big|.$$

To see this, expand $1_{n\equiv a \pmod q}$ as a Fourier series

$$1_{n\equiv a \pmod q} = \frac{1}{q}\sum_{\xi\in\mathbb{Z}/q\mathbb{Z}} e\Big(\frac{(a-n)\xi}{q}\Big),$$

and apply Lemma A.2 and the triangle inequality. □

As a consequence of this Corollary, we can obtain the following convenient lemma, which allows us to replace the range $1 \leqslant n \leqslant N$ by a smooth cutoff to the interval $N < n \leqslant 2N$, at the expense of adding an arbitrary linear phase to the function (which in our applications will be totally harmless).

Lemma A.7. — *Let $f : \mathbb{N} \to \mathbb{C}$ be a sequence bounded by $O(1)$. Let $\varphi : \mathbb{R} \to \mathbb{R}$ be a Lipschitz non-negative function of Lipschitz norm $O(1)$ which is at least 1 on $[4/3, 5/3]$. Suppose that we know that*

$$\mathbb{E}_{N<n\leqslant 2N}\varphi(\frac{n}{N})f(n)e(\alpha n) \ll_A \log^{-A} N$$

for all $A > 0$, $N \geqslant 1$, and $\alpha \in \mathbb{R}/\mathbb{Z}$. Then we have

$$\mathbb{E}_{n\in[N]} f(n) \ll_{A,\varphi} \log^{-A} N$$

for all $A > 0$ and $N \geqslant 1$.

Proof. — For large N we can write

$$\mathbb{E}_{4N/3<n\leqslant 5N/3} f(n) \ll_\varphi \Big|\mathbb{E}_{N<n\leqslant 2N} \varphi\Big(\frac{n}{N}\Big) f(n)g(n)\Big|$$

where

$$g(n) = 1_{4N/3<n\leqslant 5N/3}\varphi^{-1}(\frac{n}{N}).$$

Since φ^{-1} is Lipschitz on $[4/3, 5/3]$, we have $\|g\|_{\mathrm{TV}} \ll_\varphi 1$, and hence by Corollary A.6 and hypothesis

(A.2)
$$\mathbb{E}_{4N/3<n\leqslant 5N/3} f(n) \ll_\varphi \sup_{\alpha\in\mathbb{R}/\mathbb{Z}} \big|\mathbb{E}_{N<n\leqslant 2N} \varphi(\frac{n}{N}) f(n) e(\alpha n)\big| \ll_A \log^{-A} N.$$

Now we may decompose the interval $\{1, \dots, N\}$ into $O(\log N)$ intervals of type $4M/3 < n \leqslant 5M/3$ together with $O(\log N)$ extra points. Combining (A.2) with the bound $f = O(1)$, we obtain the lemma. □

Another harmonic analysis tool we will need often is to approximate Lipschitz functions by exponentials. We first recall a well-known extension lemma:

Lemma A.8 (Lipschitz extension). — *If Y is a non-empty subset of a metric space $X = (X, d)$, and $f : Y \to \mathbb{R}$ is a Lipschitz function then there exists a Lipschitz extension $f_{\mathrm{ext}} : X \to \mathbb{R}$ of f from Y to X with $\|f_{\mathrm{ext}}\|_{\mathrm{Lip}} = \|f\|_{\mathrm{Lip}}$. Similarly, if $f : Y \to \mathbb{C}$ is Lipschitz then there exists an extension $f_{\mathrm{ext}} : X \to \mathbb{C}$ with $\|f_{\mathrm{ext}}\|_{\mathrm{Lip}} \leqslant 2\|f\|_{\mathrm{Lip}}$.*

Proof. — If f is real-valued one can for instance define

$$f_{\mathrm{ext}}(x) := \min\Big(\inf\big\{f(y) + Md(x, y) : y \in Y\big\}, \sup_{y\in Y} f(y)\Big),$$

where $M := \|f\|_{\mathrm{Lip}}$. The complex case then follows by splitting f into real and imaginary parts. □

Lemma A.9 (Fourier approximation of Lipschitz functions). — *Let $(\mathbb{R}/\mathbb{Z})^d$ be the standard d-dimensional torus, with metric induced by the l^∞ norm*

$$\|(x_1, \dots, x_d)\|_{(\mathbb{R}/\mathbb{Z})^d} := \sup_{1\leqslant j\leqslant d} \|x_j\|_{\mathbb{R}/\mathbb{Z}}. \tag{A.3}$$

Let Y be a subset of $(\mathbb{R}/\mathbb{Z})^d$, and let $f : Y \to \mathbb{C}$ be a Lipschitz function bounded in magnitude by 1. Then for any $N \geqslant 1$ there exist $J = O_d(N^d)$,

$c_1, \dots, c_J = O(1)$, and $m_1, \dots, m_J \in \mathbb{Z}^d$ such that

$$f(x) = \sum_{j=1}^{J} c_j e(m_j \cdot x) + O_d\Big(\frac{\|f\|_{\mathrm{Lip}} \log N}{N}\Big)$$

for all $x \in Y$. Furthermore, the values of $m_1, \dots, m_J$ depend on L, d, N but are otherwise independent of f or Y.

Proof. — By Lemma A.8 we may take $Y = (\mathbb{R}/\mathbb{Z})^d$. Let $\sigma_N : (\mathbb{R}/\mathbb{Z})^d \to \mathbb{R}^+$ be the Fejér kernel

$$\sigma_N(x_1, \dots, x_d) := \prod_{j=1}^{d} \frac{1}{N} \frac{\sin^2(\pi N x_j)}{\sin^2(\pi x_j)}.$$

Note that

$$\hat{\sigma}_N(m) = \prod_{j=1}^{d} \Big(1 - \frac{|m_j|}{N}\Big) 1_{|m_j| \leqslant N}$$

for all $m \in \mathbb{Z}^d$. We have

$$f * \sigma_N(x) = \sum_m \widehat{f * \sigma_N}(m) e(m \cdot x) = \sum_m \hat{f}(m) \hat{\sigma}_N(m) e(m \cdot x)$$

which, since $\|f\|_\infty = O(1)$, has the form $\sum_{j=1}^{J} c_j e(m_j \cdot x)$ where $J = O_d(N^d)$ and $c_j = O(1)$. To conclude the proof of the lemma, then, it suffices to show that $\|f - f * \sigma_N\|_\infty = O_d(\|f\|_{\mathrm{Lip}} \log N/N)$. To this end, note that

$$|f(x) - f * \sigma_N(x)| = \Big| \int_{(\mathbb{R}/\mathbb{Z})^d} (f(x) - f(y)) \sigma_N(x - y)\, dy \Big|,$$

and hence by the change of variables $z := x - y$ it will suffice to show that

$$\int_{(\mathbb{R}/\mathbb{Z})^d} \|z\|_{(\mathbb{R}/\mathbb{Z})^d} \sigma_N(z)\, dz = O_d(\log N/N).$$

Since σ_N has total mass one, the portion of the integral on the region $\|z\|_{(\mathbb{R}/\mathbb{Z})^d} \leqslant N^{-1}$ is acceptable. Now, for each integer $n \geqslant 0$, consider the portion of the integral on the annular region $2^n N^{-1} \leqslant \|z\|_{(\mathbb{R}/\mathbb{Z})^d} \leqslant 2^{n+1} N^{-1}$. We have

$$\begin{aligned}
&\Big| \int_{\|z\|_{(\mathbb{R}/\mathbb{Z})^d} \sim 2^n N^{-1}} \|z\|_{(\mathbb{R}/\mathbb{Z})^d} \sigma_N(z)\, dz \Big| \\
&\qquad \ll 2^n N^{-1} \int_{\|t\|_{(\mathbb{R}/\mathbb{Z})^d} \gg 2^n N^{-1}} |\sigma_N(t)|\, dt \\
&\qquad \ll_d 2^n N^{-1} \int_{\|t_1\|_{\mathbb{R}/\mathbb{Z}} \gg 2^n N^{-1}} \frac{1}{N} \frac{\sin^2(\pi N t_1)}{\sin^2(\pi t_1)}\, dt_1
\end{aligned}$$

$$\ll_d 2^n N^{-1} \int_{\|t_1\|_{\mathbb{R}/\mathbb{Z}} \gg 2^n N^{-1}} \frac{1}{N\|t_1\|_{\mathbb{R}/\mathbb{Z}}^2}\, dt_1$$
$$\ll_d \frac{1}{N}.$$

Summing this over $n = 0, 1, \ldots, N$ we obtain the claim. □

We shall adopt the following convenient notation from [11]: we use $\mathbf{b}(x_1, \ldots, x_k)$ to denote any function of the variables $x_1, \ldots, x_k$ which is bounded by $O(1)$; the exact value of $\mathbf{b}()$ may vary from line to line, just as with the $O()$ notation. We use this notation to denote functions whose exact value is not of interest to us, invariably because they are destined to be annihilated in the course of a Cauchy-Schwarz argument such as the following.

LEMMA A.10 (Cauchy-Schwarz inequality). — *Let X, Y be finite non-empty sets, and let $f : X \times Y \to \mathbb{C}$ be a function. Then*

$$\left|\mathbb{E}_{x\in X}\mathbb{E}_{y\in Y}\mathbf{b}(x)f(x,y)\right| \ll \left|\mathbb{E}_{x\in X}\mathbb{E}_{y,y'\in Y} f(x,y)\overline{f(x,y')}\right|^{1/2}$$

and

$$\left|\mathbb{E}_{x\in X}\mathbb{E}_{y\in Y}\mathbf{b}(x)\mathbf{b}(y)f(x,y)\right| \ll \left|\mathbb{E}_{x,x'\in X}\mathbb{E}_{y,y'\in Y} f(x,y)\overline{f(x,y')f(x',y)}f(x',y')\right|^{1/4}.$$

Similarly, if $K : X^4 \to \mathbb{C}$ is a function, then

$$\left|\mathbb{E}_{x_1,x_2,x_3,x_4\in X}\mathbf{b}(x_2,x_3,x_4)\mathbf{b}(x_1,x_3,x_4)\mathbf{b}(x_1,x_2,x_4)\mathbf{b}(x_1,x_2,x_3) \times K(x_1,x_2,x_3,x_4)\right|$$
$$\ll \left|\mathbb{E}_{x_{1,0},x_{1,1},\ldots,x_{4,0},x_{4,1}\in X} \prod_{i_1,i_2,i_3,i_4\in\{0,1\}} \mathcal{C}^{i_1+\ldots+i_4} K(x_{1,i_1},\ldots,x_{4,i_4})\right|^{1/16}$$

where $\mathcal{C} : z \mapsto \overline{z}$ is the conjugation operator.

Remark. — These estimates are part of the theory of the Gowers uniformity norms $\|f\|_{U^d}$ and $\|K\|_{\square^d}$; see for instance [10, 9, 13, 12, 11, 19].

Proof. — From the triangle inequality and Cauchy-Schwarz we have

$$\left|\mathbb{E}_{x\in X}\mathbb{E}_{y\in Y}\mathbf{b}(x)f(x,y)\right| \ll \mathbb{E}_{x\in X}\left|\mathbb{E}_{y\in Y}f(x,y)\right| \leqslant \left(\mathbb{E}_{x\in X}\left|\mathbb{E}_{y\in Y}f(x,y)\right|^2\right)^{1/2}$$

and the first claim follows. The second claim follows by two iterations of the first, and the third follows from four iterations of the first. □

Now, we develop some quadratic analogues to the linear phase estimates given above. We begin with a quadratic counterpart to (A.1). We do not

pretend that the exponents here are even remotely optimal; we have opted for a statement which is conveniently derived from our earlier lemmas.

LEMMA A.11 (Weyl's inequality). — *Let $\alpha, \beta, \gamma \in \mathbb{R}$ and let $\delta \in (0,1)$. Let $I \subset \mathbb{Z}$ be a discrete interval such that $|I| \geqslant 2^{16}/\delta^6$ and*

$$\left|\mathbb{E}_{l \in I} e(\alpha l^2 + \beta l + \gamma)\right| \geqslant \delta.$$

Then we have

$$\|\alpha\|_{\mathbb{R}/\mathbb{Z}, 2^{12}\delta^{-4}} \leqslant \frac{2^{43}}{\delta^{14}|I|^2}.$$

Proof. — By translating I we may take $I = \{1, \dots, L\}$ for some L. Squaring the expression gives a double sum over variables l', l; setting $l' = l + h$, we find that

$$\Big|\sum_{h=-L}^{L} \sum_{l=\max(1-h,1)}^{\min(L-h,L)} e(2\alpha hl + \alpha h^2 + \beta h)\Big| \geqslant \delta^2 L^2.$$

Summing the inner geometric series using (A.1) we see that

$$\sum_{h=-L}^{L} \min\big(L, \frac{1}{\|2\alpha h\|_{\mathbb{R}/\mathbb{Z}}}\big) \geqslant \delta^2 L^2/2$$

and therefore that

$$\sum_{h=1}^{L} \min\big(L, \frac{1}{\|2\alpha h\|_{\mathbb{R}/\mathbb{Z}}}\big) \geqslant \delta^2 L^2/8.$$

It follows that there are at least $\delta^2 L/16$ values of $h \in \{1, \dots, L\}$ such that $\|2\alpha h\|_{\mathbb{R}/\mathbb{Z}} \leqslant 16/\delta^2 L$. The claim then follows from Lemma A.4(ii). □

One can now repeat the proof of Lemma A.4(i), using Lemma A.11 in place of (A.1), to conclude

LEMMA A.12 (Recurrent quadratics are non-diophantine). — *Let $I \subseteq \mathbb{Z}$ be a discrete interval, let α, β, γ be real numbers, and suppose that the set*

$$\big\{l \in I : \|\alpha l^2 + \beta l + \gamma\|_{\mathbb{R}/\mathbb{Z}} \leqslant \delta_1\big\}$$

has cardinality at least $\delta_2|I|$ for some $0 < \delta_1, \delta_2 < 1$ with $\delta_1 \leqslant \frac{1}{4}\delta_2$. If $|I| \geqslant 2^{58}\delta_2^{-12}$, then we have

$$\|\alpha\|_{\mathbb{R}/\mathbb{Z}, 2^{43}\delta_2^{-9}} \leqslant 2^{141}\delta_2^{-28}|I|^{-2}.$$

The final tool we assemble in this appendix is a technical lemma used in §8. This allows us to approximate a Lipschitz function F by a "soft-thresholded" function $\widetilde{F}$.

LEMMA A.13 (Soft-thresholding a Lipschitz function). — *Let $F : X \to [-1,1]$ be any Lipschitz function on a metric space (X,d), and let $\delta > 0$ be a parameter. Then there is a Lipschitz function $\widetilde{F} : X \to [-1,1]$ satisfying the following properties:*

(1) $\|\widetilde{F}\|_{\mathrm{Lip}} \leqslant \|F\|_{\mathrm{Lip}}$;
(2) *If $x \in \mathrm{Supp}(\widetilde{F})$ and $d(x,x') \leqslant \delta$ then $x' \in \mathrm{Supp}(F)$;*
(3) $\|F - \widetilde{F}\|_\infty \leqslant \delta \|F\|_{\mathrm{Lip}}$.

Proof. — We will set

$$\widetilde{F}(x) := \max(|F(x)| - \lambda, 0)\,\mathrm{sgn}(F(x))$$

for an appropriate value of $\lambda \geqslant 0$ which we shall shortly specify. Let us first prove that *any* such function satisfies (i). Since $|\widetilde{F}|$ is pointwise bounded by $|F|$, it suffices to show that if $x, x' \in X$ then

$$|\widetilde{F}(x) - \widetilde{F}(x')| \leqslant |F(x) - F(x')|.$$

But this follows because the function $x \mapsto \max(|x| - \lambda, 0)\,\mathrm{sgn}(x)$ is easily seen to be a contraction. This proves (i).

Now set $\lambda := \delta \|F\|_{\mathrm{Lip}}$. Statement (iii) is then obvious. To prove (ii), note that if $x \in \mathrm{Supp}(\widetilde{F})$ then $|F(x)| > \lambda$. Thus if $d(x,x') \leqslant \delta$ then

$$|F(x')| \geqslant |F(x)| - |F(x) - F(x')| \geqslant |F(x)| - \delta\|F\|_{\mathrm{Lip}} > 0. \tag{A.4}$$

□

Appendix B. Nilsequences and locally polynomial phases

The purpose of this appendix if to give the proof of Proposition 2.3, the statement of which we recall now.

PROPOSITION 2.3 (2-step nilsequences are averages of twisted 1-step nilsequences). — *Let G/Γ be a 2-step nilmanifold and let $0 < \varepsilon < 1/2$. Let $F : G/\Gamma \to \mathbb{C}$ be a bounded Lipschitz function with $\|F\|_{\mathrm{Lip}} \leqslant 1$, and let $g \in G$ and $x \in G/\Gamma$ be arbitrary. Then there exists a 1-step nilmanifold $\widetilde{G}/\widetilde{\Gamma}$ depending only on G/Γ and a decomposition*

$$F(T_g^n x) = \mathbb{E}_{i \in I} w_i F_i(T_{g_i}^n x_i) e(-\phi_i(n)) + O(\varepsilon)$$

where

- *I is a finite index set;*
- *For each $i \in I$ the w_i are complex numbers with $\mathbb{E}_{i\in I}|w_i| \ll_{\varepsilon} O_{G/\Gamma}(1)$;*

- $F_i : \widetilde{G}/\widetilde{\Gamma} \to \mathbb{C}$ *is bounded* $O_{G/\Gamma}(1)$*-Lipschitz;*
- $g_i \in \widetilde{G}$*;*
- $x_i \in \widetilde{G}/\widetilde{\Gamma}$*;*
- $\phi_i : B_i \to \mathbb{R}/\mathbb{Z}$ *is a phase function which is locally quadratic on the* generalised Bohr set $B_i := \{n \in [N] : F_i(T_{g_i}^n x_i) \neq 0\}$.

As we remarked in §2, we are going to give a rather hands-on calculational approach to this theorem, using Mal'cev bases and the Heisenberg nilmanifold as an illustrative example. The reader interested in a comprehensive discussion of Mal'cev bases may consult the book [5].

Let G be a connected, simply connected, 2-step nilpotent Lie group. Thus G is a Lie group, and the central series $G_0 = G_1 = G$, $G_2 := [G, G_1]$, $G_3 := [G, G_2]$ terminates at the third step, so that $G_3 = \{e\}$. Let Γ be a discrete, cocompact subgroup of G.

The Heisenberg example. — To motivate our arguments, let us first prove the above Proposition in the model case of the Heisenberg nilmanifold G/Γ, with

$$G := \left\{ \begin{pmatrix} 1 & x_1 & x_3 \\ 0 & 1 & x_2 \\ 0 & 0 & 1 \end{pmatrix} : x_1, x_2, x_3 \in \mathbb{R} \right\}$$

and

$$\Gamma := \left\{ \begin{pmatrix} 1 & m_1 & m_3 \\ 0 & 1 & m_2 \\ 0 & 0 & 1 \end{pmatrix} : m_1, m_2, m_3 \in \mathbb{Z} \right\}.$$

Clearly $G_1 = G$ and

$$G_2 := [G, G_1] = \left\{ \begin{pmatrix} 1 & 0 & t \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} : t \in \mathbb{R} \right\}$$

and $G_3 := [G, G_2] = \{I\}$.

Let us distinguish elements

$$e_1 = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, e_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}, e_3 = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

To these are associated the *one-parameter subgroups* $(e_i^t)_{t \in \mathbb{R}}$:

$$e_1^{t_1} = \begin{pmatrix} 1 & t_1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, e_2^{t_2} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & t_2 \\ 0 & 0 & 1 \end{pmatrix}, e_3^{t_3} = \begin{pmatrix} 1 & 0 & t_3 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Note that

$$e_1^{t_1} e_2^{t_2} e_3^{t_3} = \begin{pmatrix} 1 & t_1 & t_3 + t_1 t_2 \\ 0 & 1 & t_2 \\ 0 & 0 & 1 \end{pmatrix}.$$

The collection $\{e_1, e_2, e_3\}$ is an example of a *Mal'cev basis* for G which respects Γ, the key feature to note being that Γ is precisely the set $\{e_1^{m_1} e_2^{m_2} e_3^{m_3} : m_1, m_2, m_3 \in \mathbb{Z}\}$.

For Mal'cev coordinates to be of any use, we need to know how the group operation in G interacts with them. It is easy to explore this for the Heisenberg nilmanifold. Every element $x = e_1^{t_1} e_2^{t_2} e_3^{t_3} \in G$ may be written

in Mal'cev coordinates as $\langle t_1, t_2, t_3\rangle_{\mathrm{T_{II}}}$. It is a simple matter to check that multiplication in G is given by the rule

$$\langle t_1, t_2, t_3\rangle_{\mathrm{T_{II}}} * \langle u_1, u_2, u_3\rangle_{\mathrm{T_{II}}} = \langle t_1 + u_1, t_2 + u_2, t_3 + u_3 - t_2u_1\rangle_{\mathrm{T_{II}}}. \tag{B.1}$$

A trivial induction confirms that if $g = \langle \alpha_1, \alpha_2, \alpha_3\rangle_{\mathrm{T_{II}}}$ then

$$g^n = \langle n\alpha_1, n\alpha_2, n\alpha_3 - \tfrac{1}{2}n(n-1)\alpha_1\alpha_2\rangle_{\mathrm{T_{II}}}, \tag{B.2}$$

an expression which provides the first indication that 2-step nilmanifolds are somehow associated with "quadratic" types of behaviour.

To coordinatize the nilmanifold G/Γ, we pick a fundamental domain for the action of Γ on G. A very natural one is

$$\mathcal{F} := \{\langle x_1, x_2, x_3\rangle_{\mathrm{T_{II}}} : -\tfrac{1}{2} < x_1, x_2, x_3 \leqslant \tfrac{1}{2}\}.$$

If $x = \langle x_1, x_2, x_3\rangle_{\mathrm{T_{II}}} \in G$, then we write γ_x for the unique element of Γ such that $x\gamma_x \in \mathcal{F}$. We have

$$\gamma_x = \langle -[x_1], -[x_2], -[x_3 - [x_1]x_2]\rangle_{\mathrm{T_{II}}},$$

where $[u] = u - \{u\}$ denotes the nearest integer function (fractional parts are taken to have values in $(-\frac{1}{2}, \frac{1}{2})$). Defining

$$\tau(x) = x\gamma_x,$$

we therefore have

$$\tau(x) = \langle \{x_1\}, \{x_2\}, \{x_3 - [x_1]x_2\}\rangle_{\mathrm{T_{II}}}.$$

For any element x we have that x and $\tau(x)$ are equivalent under the action of Γ on G.

We may now analyse the map $T_g : G/\Gamma \to G/\Gamma$. Recall that if $\psi : G \to G/\Gamma$ is the canonical projection then the transformation $T_g : G/\Gamma \to G/\Gamma$ is defined via the rule $T_g(\psi(x)) = \psi(gx)$. Persisting with the notation $g = \langle \alpha_1, \alpha_2, \alpha_3\rangle_{\mathrm{T_{II}}}$ and using coordinates on the fundamental domain $\mathcal{F}$ to represent G/Γ, we have

$$\begin{aligned} T_g^n(0) &= \tau(g^n 0) \\ &= \big\langle \{n\alpha_1\}, \{n\alpha_2\}, \{n\alpha_3 - \tfrac{1}{2}n(n-1)\alpha_1\alpha_2 - [n\alpha_1]n\alpha_2\}\big\rangle_{\mathrm{T_{II}}} \\ &\equiv \big\langle n\alpha_1, n\alpha_2, n\alpha_3 - \tfrac{1}{2}n(n-1)\alpha_1\alpha_2 - [n\alpha_1]n\alpha_2\big\rangle_{\mathrm{T_{II}}} \pmod 1. \end{aligned} \tag{B.3}$$

This provides the first indication that nilmanifolds encode behaviour somewhat more general than simply quadratic; here we have "generalised" quadratic behaviour typified by the appearance of the "bracket quadratic" $[n\alpha_1]n\alpha_2$. We have now assembled everything we need to prove Proposition 2.3 for the Heisenberg nilmanifold.

Proof of Proposition 2.3 for the Heisenberg nilmanifold. — Let $F(T_g^n x)$ be a nilsequence on G/Γ. For the sake of exposition we take $x = 0$ so that (B.3) applies. Let $\pi : G \to G/G_2$ be the canonical projection and, by abuse of notation, write $\pi : G/\Gamma \to G/\Gamma G_2$ for the induced projection. Now $G/\Gamma G_2$ is a 1-step nilmanifold, being the quotient of G/G_2 by $\Gamma/\Gamma \cap G_2$, and we may identify it with $(\mathbb{R}/\mathbb{Z})^2$ via the coordinatization

$$\pi(\langle t_1, t_2, t_3\rangle_{\mathrm{T_{II}}}) = (t_1, t_2).$$

Observe that $(\pi(T_g^n 0))_{n\in\mathbb{N}} = (T_{\pi(g)}^n 0)_{n\in\mathbb{N}}$ is an orbit on $G/\Gamma G_2$, generated by the rotation $T_{\pi(g)} : (t_1, t_2) \to (t_1 + \alpha_1, t_2 + \alpha_2)$ on the torus. Let

$$1 = \sum_{l=1}^{d} \psi_l,$$

$d = O(1)$, be a Lipschitz partition of unity on $(\mathbb{R}/\mathbb{Z})^2$ with the property that for each l there are x_1, x_2 such that

$$\mathrm{Supp}(\psi_l) = [x_1, x_1 + \tfrac{1}{10}] \times [x_2, x_2 + \frac{1}{10}].$$

Then we have

$$F(T_g^n 0) = \sum_{l=1}^{d} \psi_l(T_{\pi(g)}^n 0))F(T_g^n 0).$$

We will look at each constituent nilsequence $\psi_l(T_{\pi(g)}^n 0)F(T_g^n 0)$, and write it in terms of local quadratics on 1-step Bohr sets defined on $G/\Gamma G_2$.

Fix l, $1 \leqslant l \leqslant d$ together with the associated x_1 and x_2. Now the set $U := \{x \in G/\Gamma : \pi(x) \in [x_1, x_1 + \frac{1}{10}] \times [x_2, x_2 + \frac{1}{10}]\}$ is diffeomorphic to the direct product

$$[x_1, x_1 + \tfrac{1}{10}] \times [x_2, x_2 + \frac{1}{10}] \times \mathbb{R}/\mathbb{Z},$$

which itself is diffeomorphic to a subset of $(\mathbb{R}/\mathbb{Z})^3$. Write $\pi_3 : U \to \mathbb{R}/\mathbb{Z}$ for projection onto the third coordinate. Write S for the set of all $n \in \mathbb{N}$ such that $T_g^n 0 \in U$. Note that S is a 1-step Bohr set, since

$$S = \big\{n : \psi_l(T_{\pi(g)}^n 0) \neq 0\big\}.$$

Lemma B.1 (Local quadratic behaviour). — *Suppose that n, h_1, h_2 and h_3 are such that all eight of the points $n + \epsilon_1 h_1 + \epsilon_2 h_2 + \epsilon_3 h_3$, $\epsilon_1, \epsilon_2, \epsilon_3 \in \{0, 1\}$, lie in S. Then the π_3-coordinates are subject to the quadratic constraint*

$$\sum_{\epsilon_1,\epsilon_2,\epsilon_3\in\{0,1\}} (-1)^{\epsilon_1+\epsilon_2+\epsilon_3} \pi_3(T_g^{n+\epsilon_1 h_1+\epsilon_2 h_2+\epsilon_3 h_3} 0) = 0.$$

Proof. — Recall (B.3). Writing

$$f_1(n) := n\alpha_3 - \tfrac{1}{2}n(n-1)\alpha_1\alpha_2 - [n\alpha_1]n\alpha_2,$$

we are to show that

$$\sum_{\epsilon_1,\epsilon_2,\epsilon_3\in\{0,1\}} (-1)^{\epsilon_1+\epsilon_2+\epsilon_3} f_1(n+\epsilon_1h_1+\epsilon_2h_2+\epsilon_3h_3) = 0$$

whenever the $n+\epsilon_1h_1+\epsilon_2h_2+\epsilon_3h_3$ are all in S. We may write f_1 as the sum of a quadratic polynomial and $f_2(n) := \{n\alpha_1\}n\alpha_2$. It suffices, then, to verify the result for this function f_2 instead. To do this, we note that the obvious relations

$$\{(\epsilon_1h_1+\epsilon_2h_2+\epsilon_3h_3)\alpha_1\} \equiv \{(n+\epsilon_1h_1+\epsilon_2h_2+\epsilon_3h_3)\alpha_1\} - \{n\alpha_1\} \pmod 1$$

are actually equalities in $\mathbb{R}$, and not just in $\mathbb{R}/\mathbb{Z}$, by virtue of the constraint that all quantities $\{(n+\epsilon_1h_1+\epsilon_2h_2+\epsilon_3h_3)\alpha_1\}$ lie in the interval $[x_1, x_1+\frac{1}{10}]$. Furthermore we have such relations as

$$\{h_1\alpha_1\} + \{h_2\alpha_1\} = \{(h_1+h_2)\alpha_1\}.$$

By employing these together with a few simple manipulations, the lemma follows. □

To introduce locally quadratic exponentials, we use Lemma A.9 to approximate $F = F(u_1,u_2,u_3)$, considered as a function on $U \subseteq (\mathbb{R}/\mathbb{Z})^3$, by a sum of exponentials. For any ϵ we may pick $J = O(\epsilon^{-3}\log^3(1/\epsilon))$ together with complex numbers $c_1,\dots,c_J = O(1)$ and frequencies $m_1,\dots,m_J \in \mathbb{Z}^3$ so that

$$F(u_1,u_2,u_3) = \sum_{j=1}^{J} c_j e(m_j\cdot u) + O(\epsilon)$$

for all $u = (u_1,u_2,u_3) \in U$. Using (B.3) we obtain the formula

$$F(T_g^n 0) = \sum_{j=1}^{J} c_j e(m_j^{(1)}\{n\alpha_1\} + m_j^{(2)}\{n\alpha_2\} + m_j^{(3)}\pi_3(T_g^n 0)) + O(\epsilon).$$

Each function $e(m_j^{(1)}\{n\alpha_1\} + m_j^{(2)}\{n\alpha_2\})$ is a Lipschitz nilsequence on $G/\Gamma G_2$, that is to say it can be written in the form $f_k(T^n_{\pi(g)}0)$. Thus we can write

$$\psi_l(\pi(T_g^n 0))F(T_g^n 0) = \sum_{j=1}^{J} \widetilde{f}_j(T^n_{\pi(g)}0)e(m_j^{(3)}\pi_3(T_n^g 0)) + O(\epsilon).$$

By Lemma B.1, each of the constituents here is a local quadratic on a 1-step Bohr set. This concludes the proof of Proposition 2.3 in the special case of the Heisenberg nilmanifold. □

The general case. — The above arguments can be can be extended to more general nilpotent groups. To do so, we need to involve the Lie algebra $\mathfrak{g}$ associated to G together with the exponential map

$$\exp : \mathfrak{g} \to G.$$

For the Heisenberg nilmanifold $\mathfrak{g}$ may be identified with the Lie algebra of strictly upper triangular 3×3 matrices over $\mathbb{R}$ with 0's on the diagonal, that is to say

$$\mathfrak{g} = \left\{ \begin{pmatrix} 0 & u_1 & u_3 \\ 0 & 0 & u_2 \\ 0 & 0 & 0 \end{pmatrix} : u_1, u_2, u_3 \in \mathbb{R} \right\}.$$

The exponential map is given by matrix exponentiation, so $\exp(X) = e^X$, which in practice means that if

$$X = \begin{pmatrix} 0 & u_1 & u_3 \\ 0 & 0 & u_2 \\ 0 & 0 & 0 \end{pmatrix}$$

then

$$\exp(X) = \begin{pmatrix} 1 & u_1 & u_3 + \frac{1}{2} u_1 u_2 \\ 0 & 1 & u_2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Wth the notation of Lie algebras and the exponential map it is possible to define, for a connected, simply-connected, nilpotent Lie group G, the 1-parameter subgroup $(g^t)_{t \in \mathbb{R}}$ associated to an element $g \in G$. Thus we set

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

for all $X \in \mathfrak{g}$ and $t \in \mathbb{R}$.

We can now obtain Mal'cev coordinates for any nilmanifold arising from a connected and simply connected Lie group:

Proposition B.2 (Mal'cev coordinates of the second kind). — *Let G be a connected and simply connected s-step nilpotent Lie group with central series*

$$G = G_0 = G_1 \supseteq G_2 \supseteq G_3 \supseteq \cdots \supseteq G_{s+1} = \{e\}.$$

Let Γ be a discrete, cocompact subgroup of G. Then there is a collection

$$\{e_1, \ldots, e_{i_1}, e_{i_1+1}, \ldots, e_{i_2}, e_{i_2+1}, \ldots, e_{i_k}\}$$

such that

(i) *Suppose that $j \in \{1, \ldots, s+1\}$, and define $i_0 := 1$. Then every element of G_j can be written uniquely as $e_{i_j+1}^{t_{i_j+1}} \ldots e_{i_{s+1}}^{t_{s+1}}$, for real numbers $t_{i_j+1}, \ldots, t_{s+1}$.*

(ii) *We have*

$$\Gamma = \{e_1^{m_1} \ldots e_{s+1}^{m_{s+1}} : m_1, \ldots, m_{s+1} \in \mathbb{Z}\}.$$

It turns out to be more natural to deal with *coordinates of the first kind*, which are defined on the Lie algebra $\mathfrak{g}$. Before defining these, we assemble some slightly disparate facts about how the exponential map provides a link between $\mathfrak{g}$ and G in the nilpotent case. It is not particularly easy to find proofs of all of these statements in one place: our main resources were [3] and [5].

PROPOSITION B.3 (Nilpotent Lie algebras and groups). — *Let G be a connected, simply connected, s-step nilpotent Lie group. Let $\mathfrak{g}$ be the corresponding Lie algebra, and let $\exp : \mathfrak{g} \to G$ be the exponential map. We have the following statements.*

(i) $\exp$ *is a diffeomorphism between $\mathfrak{g}$ and G, both of which are diffeomorphic to some Euclidean space $\mathbb{R}^d$.*

(ii) *Define the central series of $\mathfrak{g}$ by $\mathfrak{g}_0 = \mathfrak{g}_1 := \mathfrak{g}$ and $\mathfrak{g}_{i+1} = [\mathfrak{g}, \mathfrak{g}_i]$ for $i \geqslant 1$. Then $\exp(\mathfrak{g}_i) = G_i$. In particular, the Lie algebra $\mathfrak{g}$ is s-step nilpotent. We have the relations $[\mathfrak{g}_i, \mathfrak{g}_j] \subseteq \mathfrak{g}_{i+j}$ and $[G_i, G_j] \subseteq G_{i+j}$.*

(iii) (Baker-Campbell-Hausdorff Formula): *We have*

$$\exp(X)\exp(Y) = \exp(Z),$$

where

$$Z = X + Y + \tfrac{1}{2}[X,Y] + \frac{1}{12}[X,[X,Y]] + \frac{1}{12}[Y,[Y,X]] + \dots$$

Remarks. — The dots in (iii) are supposed to indicate that the Baker-Campbell-Hausdorff formula has terms involving commutators of fourth and higher order. Note, however, that since $\mathfrak{g}$ is nilpotent, the series does terminate. It is possible to give a description of the whole series, though it does not have a particularly simple closed form. See [3].

We describe now the Mal'cev coordinates of the first kind:

THEOREM B.4 (Mal'cev coordinates of the first kind). — *Let G be a connected, simply-connected, nilpotent Lie group with Mal'cev basis $\{e_1, \dots, e_k\}$. Thus any element $g \in G$ may be written uniquely as $e_1^{t_1} \dots e_k^{t_k}$, giving rise to the Mal'cev coordinates of the second kind $\langle t_1, \dots, t_k \rangle_{\mathrm{II}}$. Write $e_i = \exp(X_i)$, where $X_i \in \mathfrak{g}$. Then for any $g \in G$ there are unique $\xi_1, \dots, \xi_k \in \mathbb{R}$ such that $g = \exp(\xi_1 X_1 + \dots + \xi_k X_k)$. We refer to the elements of the k-tuple $\langle \xi_1, \dots, \xi_k \rangle_{\mathrm{I}}$ as the* Mal'cev coordinates of the first kind.

Remark. — In view of Proposition B.2 (i) and Proposition B.3 (ii), we have

$$\mathfrak{g}_j = \mathrm{Span}_{\mathbb{R}}(X_{i_j+1}, \dots, X_{i_k}).$$

For the Heisenberg nilmanifold, note that

$$\langle t_1, t_2, t_3\rangle_{\mathrm{T_{II}}} = \big\langle t_1, t_2, t_3 + \tfrac{1}{2}t_1t_2\big\rangle_{\mathrm{T_I}}.$$

Writing $\tau : \mathbb{R}^3 \to G$ for the map which identifies coordinates of the first kind with the element in G they represent, we see that $\tau^{-1}(\Gamma)$ is not a lattice. Fortunately, something nearly as good is true.

Proposition B.5 (Fundamental domain description of a nilmanifold). [1, Ch IV.6]. *Let G/Γ be a nilmanifold, and suppose that $X_1, \dots, X_k$ is a Mal'cev basis of the first kind in $\mathfrak{g}$. Let $\tau : \langle \xi_1, \dots, \xi_k\rangle_{\mathrm{I}} \mapsto \exp(\xi_1 X_1 + \cdots + \xi_k X_k)$ be the coordinate map, and let $\mathcal{F}$ be any region of the form*

$$\big\{\langle \xi_1, \dots, \xi_k\rangle_{\mathrm{I}} : a_i \leqslant \xi_i < a_i + 1 \ \text{ for all } i\big\}.$$

Then each point of G is equivalent, under the right action of Γ, to precisely one point in $\exp(\mathcal{F})$. Furthermore the natural projection map $\pi : G \to G/\Gamma$ is continuous on $\exp(\mathcal{F})$ and is a homeomorphism when restricted to the interior $\exp(\mathcal{F})^\circ$.

Our aim now is to describe the action of some $g = \langle \beta_1, \dots, \beta_k\rangle_{\mathrm{T_I}}$ on G/Γ by finding formulæ analogous to (B.1), (B.2) and (B.3). The key tool is the Baker-Campbell-Hausdorff formula. For notational simplicity we restrict to the 2-step case from now on, and write $m := i_2$ and $n := i_3$. Thus the Mal'cev basis of the first kind for G is $\{X_1, \dots, X_m, X_{m+1}, \dots, X_n\}$, where

$$\mathrm{Span}_{\mathbb{R}}(X_{m+1}, \dots, X_n) = \mathfrak{g}_2 = [\mathfrak{g}, \mathfrak{g}].$$

The Lie algebra $\mathfrak{g}$ is completely specified by its *structure constants*, a collection of real numbers $(a_{ijk})_{1\leqslant i,j\leqslant m, m+1\leqslant k\leqslant n}$ such that

$$[X_i, X_j] = \sum_{k=m+1}^{n} a_{ijk} X_k. \tag{B.4}$$

These constants can be arbitrary so long as $(a_{ijk})_{i,j\leqslant m}$ is antisymmetric for each k, though if we want G to possess a cocompact subgroup Γ then certain rationality conditions must hold [16].

Lemma B.6 (Multiplication in coordinates of the first kind). — *Suppose that G is a connected and simply-connected 2-step nilpotent Lie group with group operation $*$, and abuse notation by identifying elements of G with their coordinates of the first kind. Then we have*

$$\begin{aligned}\langle \xi_1, \dots, \xi_n\rangle_{\mathrm{T_I}} * \langle \nu_1, \dots, \nu_n\rangle_{\mathrm{T_I}} = \big\langle &\xi_1 + \nu_1, \dots, \xi_m + \nu_m, \xi_{m+1} \\ &+ \nu_{m+1} + \phi_{m+1}(\xi_{\leqslant m}, \nu_{\leqslant m}), \dots, \xi_n + \nu_n + \phi_n(\xi_{\leqslant m}, \nu_{\leqslant m})\big\rangle_{\mathrm{T_I}},\end{aligned}$$

where the $\xi_{\leqslant m} := (\xi_1, \dots, \xi_m)$, $\nu_{\leqslant m} := (\nu_1, \dots, \nu_m)$ and the ϕ_j are antisymmetric bilinear forms.

Proof. — This is a simple matter of combining the Baker-Campbell-Hausdorff formula with the existence of structure constants (B.4). We remark that the presentation of a 2-step nilmanifold in this form is essentially the same as an example discussed by Furstenberg in [8].

Observe in particular that

$$g^n = \langle n\beta_1, \dots, n\beta_n\rangle_{\mathrm{T_I}}, \tag{B.5}$$

and thus

$$T_g^n x = \langle n\beta_1 + x_1, \dots, n\beta_m + x_m, n\beta_1' + x_{m+1}, \dots, n\beta_n' + x_n\rangle_{\mathrm{T_I}} \tag{B.6}$$

for certain constants β_j' depending on g, x and the bilinear forms ϕ_j.

To coordinatize G/Γ we pick, in view of Proposition B.5, the very natural fundamental domain

$$\mathcal{F} := \big\{\langle x_1, \dots, x_n\rangle_{\mathrm{T_I}} : -\tfrac{1}{2} < x_1, \dots, x_n \leqslant \tfrac{1}{2}\big\}.$$

If $x = \langle x_1, \dots, x_n\rangle_{\mathrm{T_I}} \in G$, then we write γ_x for the unique element of Γ such that $x\gamma_x \in \mathcal{F}$. Write $\tau(x) = x\gamma_x$. We need a formula for γ_x in terms of coordinates of the first kind, and to obtain such a result we need a description of the lattice Γ in terms of these coordinates. Since Γ may be identified with $\mathbb{Z}^n$ in coordinates of the second kind, such a description can be obtained by finding the relation between the two types of coordinate. Such a relation is easy to obtain. Indeed by definition we have

$$\langle t_1, \dots, t_n\rangle_{\mathrm{T_{II}}} = \langle t_1, 0, \dots, 0\rangle_{\mathrm{T_I}} * \cdots * \langle 0, \dots, 0, t_n\rangle_{\mathrm{T_I}}.$$

By inductive use of Lemma B.6 this quickly implies that

$$\langle t_1, \dots, t_n\rangle_{\mathrm{T_{II}}} = \langle t_1, \dots, t_m, q_{m+1}(t_{\leqslant m}), \dots, q_n(t_{\leqslant m})\rangle_{\mathrm{T_I}} \tag{B.7}$$

for certain quadratic forms q_j. In fact these forms are rather related to the alternating forms ψ_j; if

$$\psi(x, y) = \sum_{k,l\leqslant m} a_{kl}x_l y_k \quad \text{then} \quad q(x) = \sum_{k<l} a_{kl}x_k x_l.$$

In terms of coordinates of the first kind, then, we see that

$$\Gamma = \{\langle r_1, \dots, r_m, r_{m+1} + q_{m+1}(r_{\leqslant m}), \dots, r_n + q_n(r_{\leqslant m}) : r_1, \dots, r_n \in \mathbb{Z}\}.$$

It follows that

$$\gamma_x = \langle -[x_1], \dots, -[x_m], -[x_{m+1} - \phi_{m+1}(x_{\leqslant m}, [x]_{\leqslant m}) + q_{m+1}([x]_{\leqslant m})], \dots, -[x_n - \phi_n(x_{\leqslant m}, [x]_{\leqslant m}) + q_n([x]_{\leqslant m})\rangle_{\mathrm{T_I}}$$

and that

$$\tau(x) = \langle \{x_1\}, \dots, \{x_m\}, \{x_{m+1} - \phi_{m+1}(x_{\leqslant m}, [x]_{\leqslant m}) + q_{m+1}([x]_{\leqslant m})\}, \\ \dots, \{x_n - \phi_n(x_{\leqslant m}, [x]_{\leqslant m}) + q_n([x]_{\leqslant m})\}\rangle_{\mathrm{T_I}}.$$

We remark that we have essentially provided an independent confirmation of Proposition B.5 for 2-step nilmanifolds. The proof in the s-step case merely involves more notation.

Combining this with (B.6) leads to the analogue of (B.3):

$$T_g^n x \equiv \langle n\beta_1, \dots, n\beta_m, \psi_{m+1}(n), \dots, \psi_n(n)\rangle_{\mathrm{T_I}} \pmod 1,$$

where each ψ_j has the form

$$\psi(n) = an + b + \sum_{i=1}^{m} c_i n[n\beta_i] + \sum_{l<k\leqslant m} c_{lk}\{n\beta_l\}\{n\beta_k\}.$$

The remainder of the proof of Proposition 2.3 is, from this point, almost identical to the special case of the Heisenberg nilmanifold. We leave the details to the reader. □

Appendix C. Divisor moment estimates

We collect some standard moment estimates for the divisor function $\tau(n) := \sum_{d|n} 1$. These are used to prove Proposition C.2, which is used in §11 to show that there are not too many "collisions" occuring in sets such as $\{dw : D < d \leqslant 2D; W < w \leqslant 2W\}$.

The basic estimate we need is

Lemma C.1. — *Let $m, N \geqslant 1$ be integers. Then we have the moment estimate*

$$\mathbb{E}_{n\in[N]}\tau(n)^m \ll_m (\log N)^{2^m-1}.$$

Proof. — This is very standard: see, for example, [4] or [18]. For our application, the precise value of exponent $2^m - 1$ does not need to be attained; any bound of the form $\log^{C_m} N$ would suffice. □

In particular, we have the second moment estimate

$$\mathbb{E}_{n\in[N]}\tau(n)^2 \ll \log^3 N$$

which by dyadic decomposition then implies

$$\sum_{n\in[N]} \frac{\tau(n)^2}{n} \ll \log^4 N. \tag{C.1}$$

Now if $A \subseteq \{1,\dots,N\}$ is a nonempty set of size αN and $m \geqslant 2$ is an integer, then from Hölder's inequality we have

$$\begin{aligned}\mathbb{E}_{n\in A}\tau(n)^2 &\leqslant (\mathbb{E}_{n\in A}\tau(n)^m)^{2/m}\\ &\leqslant \alpha^{-2/m}(\mathbb{E}_{n\in[N]}\tau(n)^m)^{2/m})\\ &\ll_m \alpha^{-2/m}(\log N)^{2(2^m-1)/m}.\end{aligned}$$

In particular, for any $\kappa < 1/2$ we have the moment estimate

$$\mathbb{E}_{n\in A}\tau(n)^2 \ll_\kappa \alpha^{-\kappa}\log^{2^{2/\kappa}} N. \tag{C.2}$$

This estimate has the following consequence.

LEMMA C.2 (Divisor packing lemma). — *Let $A \subseteq \{1,\dots,N\}$ be a nonempty set of size αN, and for each $d \geqslant 1$ let $A_d := \{n \in A : d|n\}$ denote those elements of A which are multiples of d. Suppose $\mathfrak{D} \subset \mathbb{Z}^+$ is a finite set of positive integers such that*

$$|A_d| \geqslant \delta|A|$$

for all $d \in \mathfrak{D}$ and some $\delta > 0$. Then for any positive $\kappa < 1/2$ we have

$$|\bigcup_{d\in\mathfrak{D}} A_d| \gg_\kappa \delta^2|\mathfrak{D}|^2|A|\alpha^\kappa \log^{-2^{2/\kappa}} N.$$

Proof. — From hypothesis we have

$$\mathbb{E}_{n\in A}\sum_{d\in\mathfrak{D}} 1_{A_d}(n) = \sum_{d\in\mathfrak{D}}\frac{|A_d|}{|A|} \geqslant \delta|\mathfrak{D}|.$$

By Cauchy-Schwarz we conclude that

$$\frac{|\bigcup_{d\in\mathfrak{D}} A_d|}{|A|}\mathbb{E}_{n\in A}(\sum_{d\in\mathfrak{D}} 1_{A_d}(n))^2 \geqslant \delta^2|\mathfrak{D}|^2.$$

From the trivial bound

$$\sum_{d\in\mathfrak{D}} 1_{A_d}(n) \leqslant \sum_{d|n} 1 = \tau(n)$$

and (C.2) we thus have

$$\frac{|\bigcup_{d\in\mathfrak{D}} A_d|}{|A|}\alpha^{-\kappa}\log^{2^{2/\kappa}} N \gg_\kappa \delta^2|\mathfrak{D}|^2$$

and the claim follows. □

BIBLIOGRAPHY

[1] L. Auslander, L. Green & F. Hahn, *Flows on homogeneous spaces*, With the assistance of L. Markus and W. Massey, and an appendix by L. Greenberg. Annals of Mathematics Studies, No. 53, Princeton University Press, Princeton, N.J., 1963, vii+107 pages.

[2] Y. Bilu, "Structure of sets with small sumset", *Astérisque* (1999), no. 258, p. xi, 77-108, Structure theory of set addition.

[3] N. Bourbaki, *Lie groups and Lie algebras. Chapters 1–3*, Elements of Mathematics (Berlin), Springer-Verlag, Berlin, 1998, Translated from the French, Reprint of the 1989 English translation, xviii+450 pages.

[4] J. Bourgain, "On $\Lambda(p)$-subsets of squares", *Israel J. Math.* **67** (1989), no. 3, p. 291-311.

[5] L. J. Corwin & F. 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, viii+269 pages.

[6] H. Davenport, "On some infinite series involving arithmetical functions. II", *Quart. J. Math. Oxf.* **8** (1937), p. 313-320.

[7] H. Davenport, *Multiplicative number theory*, third ed., Graduate Texts in Mathematics, vol. 74, Springer-Verlag, New York, 2000, Revised and with a preface by Hugh L. Montgomery, xiv+177 pages.

[8] H. Furstenberg, "Nonconventional ergodic averages", in *The legacy of John von Neumann (Hempstead, NY, 1988)*, Proc. Sympos. Pure Math., vol. 50, Amer. Math. Soc., Providence, RI, 1990, p. 43-56.

[9] W. T. Gowers, "A new proof of Szemerédi's theorem", *Geom. Funct. Anal.* **11** (2001), no. 3, p. 465-588.

[10] B. J. Green & T. C. Tao, "Linear equations in primes", to appear in Annals of Math.

[11] ———, "An inverse theorem for the Gowers U^3-norm", *Proc. Edinburgh Math. Soc.* **51** (2008), no. 1, p. 73-153.

[12] ———, "The primes contain arbitrarily long arithmetic progressions", *Annals of Math.* **167** (2008), p. 481-547.

[13] B. Green, "Finite field models in additive combinatorics", in *Surveys in combinatorics 2005*, London Math. Soc. Lecture Note Ser., vol. 327, Cambridge Univ. Press, Cambridge, 2005, p. 1-27.

[14] L. K. Hua, "Some results in the additive prime number theory", *Quart. J. Math. Oxford* **9** (1938), p. 68-80.

[15] H. Iwaniec & E. Kowalski, *Analytic number theory*, American Mathematical Society Colloquium Publications, vol. 53, American Mathematical Society, Providence, RI, 2004, xii+615 pages.

[16] A. I. Mal'cev, "On a class of homogeneous spaces", *Izvestiya Akad. Nauk. SSSR. Ser. Mat.* **13** (1949), p. 9-32.

[17] H. L. Montgomery, *Ten lectures on the interface between analytic number theory and harmonic analysis*, CBMS Regional Conference Series in Mathematics, vol. 84, Published for the Conference Board of the Mathematical Sciences, Washington, DC, 1994, xiv+220 pages.

[18] I. Z. Ruzsa, "On an additive property of squares and primes", *Acta Arith.* **49** (1988), no. 3, p. 281-289.

[19] T. Tao, "Arithmetic progressions and the primes", *Collect. Math.* (2006), no. Vol. Extra, p. 37-88.

[20] T. Tao & V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006, xviii+512 pages.

[21] R. C. Vaughan, *The Hardy-Littlewood method*, second ed., Cambridge Tracts in Mathematics, vol. 125, Cambridge University Press, Cambridge, 1997, xiv+232 pages.

[22] R.-C. Vaughan, "Sommes trigonométriques sur les nombres premiers", *C. R. Acad. Sci. Paris Sér. A-B* **285** (1977), no. 16, p. A981-A983.

[23] I. M. Vinogradov, "Some theorems concerning the primes", *Mat. Sbornik. N.S.* **2** (1937), p. 179-195.

Manuscrit reçu le 7 juin 2006,
accepté le 10 mai 2007.

Ben GREEN
Centre for Mathematical Sciences
Wilberforce Road
Cambridge CB3 0WA (England)
b.j.green@dpmms.cam.ac.uk

Terence TAO
UCLA Department of Mathematics
Los Angeles
CA 90095-1596 (USA)
tao@math.ucla.edu