



RESEARCH ARTICLE

Higher uniformity of arithmetic functions in short intervals

I. All intervals

Kaisa Matomäki¹ , Xuancheng Shao² , Terence Tao³ and Joni Teräväinen⁴

¹Department of Mathematics and Statistics, University of Turku, 20014, Turku, Finland; E-mail: ksmato@utu.fi.

²Department of Mathematics, University of Kentucky, 715 Patterson Office Tower, Lexington, 40506, USA;

E-mail: xuancheng.shao@uky.edu.

³Department of Mathematics, UCLA, 405 Hilgard Ave, Los Angeles, 90095, USA; E-mail: tao@math.ucla.edu.

⁴Department of Mathematics and Statistics, University of Turku, 20014, Turku, Finland; E-mail: joni.p.teravainen@gmail.com.

Received: 04 May 2020; Accepted: 12 May 2020

MSC Codes: Primary – 11N37; Secondary – 11B30

Abstract

We study higher uniformity properties of the Möbius function μ , the von Mangoldt function Λ , and the divisor functions d_k on short intervals $(X, X + H]$ with $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$ for a fixed constant $0 \leq \theta < 1$ and any $\varepsilon > 0$. More precisely, letting Λ^\sharp and d_k^\sharp be suitable approximants of Λ and d_k and $\mu^\sharp = 0$, we show for instance that, for any nilsequence $F(g(n)\Gamma)$, we have

$$\sum_{X < n \leq X + H} (f(n) - f^\sharp(n)) F(g(n)\Gamma) \ll H \log^{-A} X$$

when $\theta = 5/8$ and $f \in \{\Lambda, \mu, d_k\}$ or $\theta = 1/3$ and $f = d_2$.

As a consequence, we show that the short interval Gowers norms $\|f - f^\sharp\|_{U^s(X, X + H]}$ are also asymptotically small for any fixed s for these choices of f, θ . As applications, we prove an asymptotic formula for the number of solutions to linear equations in primes in short intervals, and show that multiple ergodic averages along primes in short intervals converge in L^2 .

Our innovations include the use of multi-parameter nilsequence equidistribution theorems to control type II sums, and an elementary decomposition of the neighborhood of a hyperbola into arithmetic progressions to control type I_2 sums.

Contents

1	Introduction	2
1.1	Gowers uniformity in short intervals	7
1.2	Applications	8
1.2.1	Polynomial phases	8
1.2.2	An application to ergodic theory	8
1.2.3	Linear equations in short intervals	9
1.3	Methods of proof	11
1.3.1	The result on Gowers norms	13
1.4	Notation	14
2	Basic tools	15

2.1	Total variation	15
2.2	Vinogradov lemma	17
2.3	Equidistribution on nilmanifolds	18
2.4	Combinatorial lemmas	29
3	Major arc estimates	33
3.1	Proof of Proposition 3.2	37
3.2	Proof of Lemma 3.3	42
3.3	Proof of Lemmas 3.5 and 3.6	42
3.4	Proof of Proposition 3.4	45
3.5	Major arc estimates with restricted prime factorization	49
4	Reduction to type I, type II, and type I₂ estimates	50
4.1	Combinatorial decompositions	52
4.2	Deduction of Theorem 1.1	56
5	The type I case	59
6	The non-abelian type II case	60
7	The abelian type II case	67
8	The type I₂ case	70
9	Controlling the Gowers uniformity norms	78
10	Applications	90
A	Variants of the main result	93
A.1	Results for the Liouville function	93
A.2	Results for the indicator function of the primes	94
A.3	Results for the counting function of sums of two squares	94
A.4	Potential result for the indicator function of the sums of two squares	95
A.5	Potential result for the indicator function of smooth numbers	96

1. Introduction

In this paper we shall study correlations of arithmetic functions $f: \mathbb{N} \rightarrow \mathbb{C}$ with arbitrary nilsequences $n \mapsto F(g(n)\Gamma)$ in short intervals. For simplicity, we will restrict attention to the following model examples of functions f :

- The *Möbius function* $\mu(n)$, defined to equal $(-1)^j$ when n is the product of j distinct primes, and 0 otherwise.
- The *von Mangoldt function* $\Lambda(n)$, defined to equal $\log p$ when n is a power p^j of a prime p for some $j \geq 1$, and 0 otherwise.
- The k^{th} *divisor function* $d_k(n)$, defined to equal the number of representations of n as the product $n = n_1 \cdots n_k$ of k natural numbers, where $k \geq 2$ is fixed. (In particular, all implied constants in our asymptotic notation are understood to depend on k .)

By a “nilsequence”, we mean a function of the form $n \mapsto F(g(n)\Gamma)$, where G/Γ is a filtered nilmanifold and $F: G/\Gamma \rightarrow \mathbb{C}$ is a Lipschitz function. The precise definitions of these terms will be given in Section 2.3, but a simple example of a nilsequence to keep in mind for now is $F(g(n)\Gamma) = e(\alpha n^d)$ for some real number α , some natural number $d \geq 0$, and with $e(\theta) := e^{2\pi i \theta}$.

When f is non-negative and $F(g(n)\Gamma)$ is “major arc” in some sense (e.g., if $F(g(n)\Gamma) = e(\alpha n^s)$ with α very close to a rational a/q with small denominator q), there is actually correlation between f and $F(g(n)\Gamma)$, but we shall deal with this by first subtracting off a suitable approximation $f^\#$ from f . In the case of the Möbius function μ , we may set $\mu^\# = 0$. On the other hand, the functions Λ, d_k are non-negative and one therefore needs to construct non-trivial approximants $\Lambda^\#, d_k^\#$ to such functions

before one can expect to obtain discorrelation; we shall choose

$$\Lambda^\sharp(n) := \frac{P(R)}{\varphi(P(R))} 1_{(n, P(R))=1}, \quad \text{where } P(w) := \prod_{p < w} p, \quad R := \exp((\log X)^{1/10}) \quad (1.1)$$

and

$$d_k^\sharp(n) := \sum_{\substack{m \leq R_k^{2k-2} \\ m|n}} P_m(\log n), \quad \text{where } R_k := X^\eta \text{ and } \eta = \frac{1}{10k} \quad (1.2)$$

and the polynomials $P_m(t)$ (which have degree $k-1$) are given by the formula

$$P_m(t) := \sum_{j=0}^{k-1} \binom{k}{j} \sum_{\substack{n_1, \dots, n_j \leq R_k < n_{j+1}, \dots, n_{k-1} \leq R_k^2 \\ n_1 \cdots n_{k-1} = m}} \frac{(t - \log(n_1 \cdots n_j R_k^{k-j}))^{k-j-1}}{(k-j-1)! \log^{k-j-1} R_k}. \quad (1.3)$$

We will discuss these choices of approximants more in Section 3.1 (which can be read independently of the rest of the paper), but let us already here note that the approximants lead to type I sums and are thus easier to handle than the original functions, and that the choice of the parameter R in Λ^\sharp allows for an arbitrary power of \log saving in (1.6) below. Moreover, the approximants are nonnegative, which is helpful for some applications (in particular in the proof of Theorem 1.5 below). For future use, we record the fact that our correlation estimates for $d_k - d_k^\sharp$ work for d_k^\sharp defined as in (1.2) with any fixed $0 < \eta \leq \frac{1}{10k}$, as long as we allow implied constants to depend on η .

For technical reasons, it can be beneficial to consider “maximal discorrelation” estimates. Loosely following Robert and Sargos [58] we adopt the convention¹ that, for an interval I ,

$$\left| \sum_{n \in I \cap \mathbb{Z}} f(n) \right|^* := \sup_{P \subset I \cap \mathbb{Z}} \left| \sum_{n \in P} f(n) \right|, \quad (1.4)$$

where P ranges over all arithmetic progressions in $I \cap \mathbb{Z}$.

Now we are ready to state our main theorem².

Theorem 1.1 (Discorrelation estimate). *Let $X \geq 3$, $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$ for some $0 < \theta < 1$ and $\varepsilon > 0$, and let $\delta \in (0, 1)$. Let G/Γ be a filtered nilmanifold of some degree d and dimension D , and complexity at most $1/\delta$, and let $F: G/\Gamma \rightarrow \mathbb{C}$ be a Lipschitz function of norm at most $1/\delta$.*

(i) *If $\theta = 5/8$, then for all $A > 0$,*

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X < n \leq X+H} \mu(n) \overline{F}(g(n)\Gamma) \right|^* \ll_{A, \varepsilon, d, D} \delta^{-O_{d, D}(1)} H \log^{-A} X \quad (1.5)$$

(ii) *If $\theta = 5/8$, then for all $A > 0$,*

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X < n \leq X+H} (\Lambda(n) - \Lambda^\sharp(n)) \overline{F}(g(n)\Gamma) \right|^* \ll_{A, \varepsilon, d, D} \delta^{-O_{d, D}(1)} H \log^{-A} X. \quad (1.6)$$

¹Strictly speaking, this is an abuse of notation, since the expression $|\sum_{n \in I \cap \mathbb{Z}} f(n)|^*$ depends not only on the value of the sum $\sum_{n \in I \cap \mathbb{Z}} f(n)$, but also on the individual summands $f(n)$ and the range $I \cap \mathbb{Z}$. In particular, we caution that $\sum_{n \in I \cap \mathbb{Z}} f(n) = \sum_{m \in J \cap \mathbb{Z}} g(m)$ does not necessarily imply that $|\sum_{n \in I \cap \mathbb{Z}} f(n)|^* = |\sum_{m \in J \cap \mathbb{Z}} g(m)|^*$.

²For definitions of undefined terms such as “filtered nilmanifold” and $\text{Poly}(\mathbb{Z} \rightarrow G)$, see Definitions 2.6 and 2.5 below. For our conventions for asymptotic notation such as \ll , see Section 1.4.

(iii) Let $k \geq 2$. Set $\theta = 1/3$ for $k = 2$, $\theta = 5/9$ for $k = 3$, and $\theta = 5/8$ for $k \geq 4$. Then

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X < n \leq X+H} (d_k(n) - d_k^\sharp(n)) \overline{F}(g(n)\Gamma) \right|^* \ll_{\varepsilon, d, D} \delta^{-O_{d, D}(1)} H X^{-c_{k, d, D}\varepsilon} \quad (1.7)$$

for some constant $c_{k, d, D} > 0$ depending only on k, d, D .

(iv) If $\theta = 3/5$, then

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X < n \leq X+H} \mu(n) \overline{F}(g(n)\Gamma) \right|^* \ll_{\varepsilon, d, D} \delta^{-O_{d, D}(1)} H \log^{-1/4} X. \quad (1.8)$$

(v) Let $k \geq 4$. If $\theta = 3/5$, then

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X < n \leq X+H} (d_k(n) - d_k^\sharp(n)) \overline{F}(g(n)\Gamma) \right|^* \ll_{\varepsilon, d, D} \delta^{-O_{d, D}(1)} H \log^{\frac{3}{4}k-1} X. \quad (1.9)$$

The dependency of the implied constants on A in (1.5) and (1.6) is ineffective due to the possible existence of Siegel zeros. All the other implied constants are effective.

Remark 1.2. One could extend the theorem to cover the range $X^{1-\varepsilon} \leq H \leq X$ without difficulty; however, this is not the most interesting regime and there are some places in the proof where the restriction to $H \leq X^{1-\varepsilon}$ is convenient. In the cases of (1.5), (1.8), the result for $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$ directly implies the result for $X^{1-\varepsilon} \leq H \leq X$ by splitting long sums into shorter ones. In the cases of (1.6), (1.7), (1.9), it turns out that there is some flexibility in the choice of the approximant (one can certainly vary R in (1.1) or R_k in (1.2) by a multiplicative factor $\asymp 1$), and then one can make a similar splitting argument. We leave the details to the interested reader.

In applications d, D, δ will often be fixed; however, the fact that the constants here depend in a polynomial fashion on δ will be useful for induction purposes.

Note that polynomial phases $F(g(n)\Gamma) = e(P(n))$, with $P: \mathbb{Z} \rightarrow \mathbb{R}$ a polynomial of degree d , are a special case of nilsequences — in this case the filtered nilmanifold is the unit circle \mathbb{R}/\mathbb{Z} (with $\mathbb{R} = (\mathbb{R}, +)$ being the filtered nilpotent group with $\mathbb{R}_i = \mathbb{R}$ for $i \leq d$ and $\mathbb{R}_i = \{0\}$ for $i > d$) and $F(\alpha) = e(\alpha)$ for all $\alpha \in \mathbb{R}/\mathbb{Z}$. In particular the results of Theorem 1.1 hold for polynomial phases, that is, with $G/\Gamma = \mathbb{R}/\mathbb{Z}$, $D = 1$, and with $\overline{F}(g(n)\Gamma)$ replaced with $e(P(n))$. Before moving on, let us for the convenience of the reader state the following corollary of our theorem in the polynomial phase case.

Corollary 1.3 (Discorrelation of μ and Λ with polynomial phases in short intervals). *Let $X \geq 3$ and let $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$ for some $0 < \theta < 1$ and $\varepsilon > 0$. Let $d \geq 1$ and let $P: \mathbb{Z} \rightarrow \mathbb{R}$ be any polynomial of degree d .*

(i) If $\theta = 5/8$, then, for all $A > 0$,

$$\left| \sum_{X < n \leq X+H} \mu(n) e(P(n)) \right| \ll_{d, A, \varepsilon} \frac{H}{\log^A X}$$

(ii) If $\theta = 5/8$ and $A > 0$, we have

$$\left| \sum_{X < n \leq X+H} \Lambda(n) e(P(n)) \right| \leq \frac{H}{\log^A X},$$

unless there exists $1 \leq q \leq (\log X)^{O_{d,A,\varepsilon}(1)}$ such that one has the “major arc” property

$$\max_{1 \leq j \leq d} H^j \|q\alpha_j\|_{\mathbb{R}/\mathbb{Z}} \leq (\log X)^{O_{d,A,\varepsilon}(1)}, \quad (1.10)$$

where α_j is the degree j coefficient of the polynomial $n \mapsto P(n+X)$ and $\|y\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance from y to the nearest integer(s).

(iii) If $\theta = 3/5$, then

$$\left| \sum_{X < n \leq X+H} \mu(n) e(P(n)) \right| \ll_{d,\varepsilon} \frac{H}{\log^{1/10} X}.$$

The claims (i) and (iii) are immediate from Theorem 1.1, but (ii) requires a short argument, provided in Section 10. One could state an analogous result in the case of d_k (with the same exponents as in Theorem 1.1).

Let us now discuss the literature on the topic, starting with results concerning the Möbius function. A discorrelation estimate such as Theorem 1.1(i) with arbitrary $F(g(n)\Gamma)$ was previously only known in case of long intervals due to the work of Green and the third author [18, Theorem 1.1]. Namely, they have shown that

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{n \leq X} \mu(n) \overline{F}(g(n)\Gamma) \right| \ll_{A,G/\Gamma,F} X \log^{-A} X \quad (1.11)$$

for any $X \geq 2$, $A > 0$, filtered nilmanifold G/Γ , and Lipschitz function $F: G/\Gamma \rightarrow \mathbb{C}$. This result of Green and the third author is a vast generalization of a classical result of Davenport [6], which states that

$$\sup_{\alpha \in \mathbb{R}} \left| \sum_{n \leq X} \mu(n) e(-\alpha n) \right| \ll_A X \log^{-A} X, \quad (1.12)$$

and of the Siegel–Walfisz theorem (see e.g. [37, Corollary 5.29]), which states that

$$\max_{a,q \in \mathbb{N}} \left| \sum_{\substack{n \leq X \\ n \equiv a \pmod{q}}} \mu(n) \right| \ll_A X \log^{-A} X. \quad (1.13)$$

As is well known, the bounds of $O_A(X \log^{-A} X)$ here cannot be improved unconditionally with current technology, due to the possible existence of Siegel zeroes (unless one subtracts a correction term to account for the contribution of such zero; see [61, Theorem 2.7]).

On the other hand, for short intervals there has been a lot of activity in the special case of polynomial phase twists.

Theorem 1.1(i) was previously only known in the linear phase case when $F(g(n)\Gamma) = e(\alpha n)$ for any $\alpha \in \mathbb{R}$ by work of Zhan [64]. More precisely Zhan [64, Theorem 5] established that

$$\sup_{\alpha \in \mathbb{R}} \left| \sum_{X < n \leq X+H} \mu(n) e(-\alpha n) \right| \ll_{A,\varepsilon} H \log^{-A} X \quad (1.14)$$

whenever $X^{5/8+\varepsilon} \leq H \leq X$ and $A \geq 1$. Hence Theorem 1.1(i) can be seen as a vast extension of Zhan’s work.

Concerning higher degree polynomials, the most recent result is due to the first two authors [49, Theorem 1.4] giving, for any polynomial $P(n)$ of degree $\leq d$,

$$\sum_{X < n \leq X+H} \mu(n) e(-P(n)) \ll_{A,d,\varepsilon} H \log^{-A} X \quad (1.15)$$

for all $A > 0$ and $X^{2/3+\varepsilon} \leq H \leq X$. In particular a special case of Theorem 1.1(i) (recorded here as Corollary 1.3(i)) supersedes this result by showing it with the exponent $2/3$ lowered to $5/8$.

All the previous results mentioned so far for the Möbius function exist also for the von Mangoldt function as long as $F(g(n)\Gamma)$ or $e(-P(n))$ is “minor arc” in certain sense (for results corresponding to (1.11), (1.12), (1.13), (1.14) and (1.15) see respectively [18, Section 7], [37, Theorem 13.6], [37, Corollary 5.29], [64, Theorems 2–3], and [49, Theorem 1.1]). It is very likely that with our choice of approximant these arguments also extend to cover major arc cases and maximal correlations, although we will not detail this here as such claims follow in any case from Theorem 1.1.

Theorem 1.1(iv) generalizes (albeit with a slightly weaker logarithmic saving) a result of the first and fourth authors [50, Theorem 1.5] that gave, for $0 < A < 1/3$,

$$\sup_{\alpha \in \mathbb{R}} \left| \sum_{X < n \leq X+H} \mu(n) e(-\alpha n) \right| \ll_{A, \varepsilon} H \log^{-A} X \quad (1.16)$$

in the regime $X \geq H \geq X^{3/5+\varepsilon}$ (actually [50, Remark 5.2] allows one to enlarge the range of A to $0 < A < 1$).

The literature on correlations between d_k and Fourier or higher order phases is sparse. A variant of the long interval case (1.11) (with a weaker error term) follows from work of Matthiesen [51, Theorem 6.1].

Furthermore, it should be possible to adapt the existing results on polynomial correlations of $\Lambda(n)$ also to the case of $d_k(n)$, but with power savings. More precisely, one should be able to follow the approach of Zhan [64] to obtain disconnection with linear phases $e(\alpha n)$ for $X \geq H \geq X^{5/8+\varepsilon}$ (for $k = 2$ one can replace $5/8$ by $1/2$ and for $k = 3$ one can replace $5/8$ by $3/5$) and the work of the first two authors [49] to obtain disconnection with polynomial phases for $X \geq H \geq X^{2/3+\varepsilon}$ (for $k = 2$ one can replace $2/3$ by $1/2$). We omit the details of these extensions of [64, 49] as they follow from our Theorem 1.1.

We note that in the case $k = 2$ the exponent $1/3$ in Theorem 1.1(iii) matches the classical Voronoi exponent for the error term in long sums of the divisor function without any twist, and the result seems to be new even in the case of linear phases.

In the most major arc case $F(g(n)\Gamma) = 1$, shorter intervals can be reached than in Theorem 1.1; see Theorem 3.1 below. Furthermore if one only wants disconnection in almost all intervals, for instance by seeking to bound

$$\int_X^{2X} \sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H} (f(n) - f^\sharp(n)) \overline{F}(g(n)\Gamma) \right|^* dx,$$

much shorter intervals can be reached with aid of additional ideas. We will return to this question and its applications in a follow-up paper [46].

Remark 1.4. It should be clear to experts from an inspection of our arguments that the methods used in this paper could also treat other arithmetic functions with similar structure to μ , Λ , or d_k . For instance, all of the results for the Möbius function μ here have counterparts for the Liouville function λ ; the results for the von Mangoldt function Λ have counterparts (with somewhat different normalizations) for the indicator function $1_{\mathbb{P}}$ of the primes \mathbb{P} , and the results for d_2 have counterparts for the function $r_2(n) := \sum_{a,b \in \mathbb{Z}: a^2+b^2=n} 1$ counting the number of representations of n as the sum of two squares. We sketch the modifications needed to establish these variants in Appendix A. We also conjecture that the methods can be extended to treat the indicator function 1_S of the set $S := \{a^2 + b^2 : a, b \in \mathbb{Z}\}$ of sums of two squares, or the indicator 1_{S_η} of X^η -smooth numbers, although in those two cases a technical difficulty arises that the construction of a sufficiently accurate approximant to these indicator functions is non-trivial. Again, see Appendix A for further discussion.

On the other hand, our arguments do not seem to easily extend to the Fourier coefficients $\lambda_f(n)$ of holomorphic cusp forms. The coefficients $\lambda_f(n)$ are analogous to $d_2(n)$ in many ways (though with vanishing approximant $\lambda_f^\sharp = 0$), and it is reasonable to conjecture parallel results for these two functions. For instance, in [10] it was established that

$$\sup_{\alpha} \left| \sum_{X < n \leq X+H} \lambda_f(n) e(\alpha n) \right| \ll H X^{-c_\varepsilon}$$

for $X^{2/5+\varepsilon} \leq H \leq X$. See also [25] for a result with general nilsequences but long intervals. Unfortunately, the methods we use in this paper rely heavily on the convolution structure of the functions involved and do not obviously extend to give results for λ_f .

1.1. Gowers uniformity in short intervals

Just as discorrelation estimates with polynomial phases are important for applications of the circle method, discorrelation estimates with nilsequences are important in higher order Fourier analysis due to the connection with the Gowers uniformity norms that we next discuss.

For any non-negative integer $s \geq 1$, and any function $f: \mathbb{Z} \rightarrow \mathbb{C}$ with finite support, define the (unnormalized) Gowers uniformity norm

$$\|f\|_{U^s(\mathbb{Z})} := \left(\sum_{x, h_1, \dots, h_s \in \mathbb{Z}} \prod_{\omega \in \{0,1\}^s} C^{|\omega|} f(x + \omega_1 h_1 + \dots + \omega_s h_s) \right)^{1/2^s}$$

where $\omega = (\omega_1, \dots, \omega_s)$, $|\omega| := \omega_1 + \dots + \omega_s$, and $C: z \mapsto \bar{z}$ is the complex conjugation map. Then for any interval $(X, X+H]$ with $H \geq 1$ and any $f: \mathbb{Z} \rightarrow \mathbb{C}$ (not necessarily of finite support), define the *Gowers uniformity norm over $(X, X+H]$* by

$$\|f\|_{U^s(X, X+H]} := \|f 1_{(X, X+H]}\|_{U^s(\mathbb{Z})} / \|1_{(X, X+H]}\|_{U^s(\mathbb{Z})} \quad (1.17)$$

where $1_{(X, X+H]}: \mathbb{Z} \rightarrow \mathbb{C}$ is the indicator function of $(X, X+H]$.

Using the inverse theorem for Gowers norms (see Proposition 9.4) we can deduce the following theorem from Theorem 1.1 and a construction of pseudorandom majorants in Section 9.

Theorem 1.5 (Gowers uniformity estimate). *Let $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$ for some fixed $0 < \theta < 1$ and $\varepsilon > 0$. Let $s \geq 1$ be a fixed integer. Also denote $\Lambda_w(n) := \frac{W}{\varphi(W)} 1_{(n, W)=1}$, where $W := \prod_{p \leq w} p$ and X is large enough in terms of w .*

(i) *If $\theta = 5/8$, then*

$$\|\Lambda - \Lambda_w\|_{U^s(X, X+H]} = o_{w \rightarrow \infty}(1), \quad (1.18)$$

and for any $1 \leq a \leq W$ with $(a, W) = 1$ we have

$$\left\| \frac{\varphi(W)}{W} \Lambda(W \cdot +a) - 1 \right\|_{U^s(X, X+H]} = o_{w \rightarrow \infty}(1). \quad (1.19)$$

(ii) *Let $k \geq 2$. Set $\theta = 1/3$ for $k = 2$, $\theta = 5/9$ for $k = 3$, and $\theta = 3/5$ for $k \geq 4$. Then*

$$\|d_k - d_k^\sharp\|_{U^s(X, X+H]} = o(\log^{k-1} X), \quad (1.20)$$

and for any W' satisfying $W \mid W' \mid W^{\lfloor w \rfloor}$ and for any $1 \leq a \leq W'$ with $(a, W') = 1$ we have

$$\|d_k(W' \cdot +a) - d_k^\sharp(W' \cdot +a)\|_{U^s(X, X+H]} = o_{w \rightarrow \infty} \left(\left(\frac{\varphi(W')}{W'} \right)^{k-1} \log^{k-1} X \right). \quad (1.21)$$

(iii) If $\theta = 3/5$, then

$$\|\mu\|_{U^s(X, X+H]} = o(1). \quad (1.22)$$

In all these estimates the $o(1)$ notation is with respect to the limit $X \rightarrow \infty$ (holding s, ε, k fixed).

Remarks.

- The model Λ_w with w fixed is simple to work with and arises in various applications of Gowers uniformity (e.g. to ergodic theory). This also motivates our choice of the Λ^\sharp model in (1.1) (although that is defined with a larger value of w to produce better error terms).
- Since the bounds in this theorem (unlike in Theorem 1.1) are qualitative in nature, it should be possible to use Heath-Brown's trick from [29] to extend the range of H from $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$ to $X^\theta \leq H \leq X^{1-\varepsilon}$. Also the range $X^{1-\varepsilon} \leq H \leq X$ could be covered, as in Remark 1.2. We leave the details to the interested reader.
- In the case $s = 2$, we obtain significantly stronger estimates thanks to the polynomial nature of the U^2 inverse theorem. Specifically, when $\theta = 5/8 + \varepsilon$, we have

$$\|\mu\|_{U^2(X, X+X^\theta]}, \|\Lambda - \Lambda^\sharp\|_{U^2(X, X+X^\theta]} \ll_{A, \varepsilon} \log^{-A} X$$

for all $A > 0$ and

$$\|d_k\|_{U^2(X, X+X^\theta]} \ll_\varepsilon X^{-c_k \varepsilon} \quad (1.23)$$

for some $c_k > 0$, with (1.23) also holding when $(k, \theta) = (3, 5/9), (2, 1/3)$, and finally

$$\|\mu\|_{U^2(X, X+X^\theta]} \ll_\varepsilon \log^{-1/20} X$$

when $\theta = 3/5$. All of these follow directly by combining Theorem 1.1 for $d = 1$ (that is, for Fourier phases in place of nilsequences) with the polynomial form of the U^2 inverse theorem, which states that if $f : [N] \rightarrow \mathbb{C}$ is 1-bounded and $\|f\|_{U^2[N]} \geq \delta$ for some $\delta > 0$, then $|\sum_{n \leq N} f(n)e(an)|^* \gg \delta^4 N$ for some $\alpha \in \mathbb{R}$. This form of the inverse theorem follows directly from the Fourier representation of the $U^2[N]$ norm and Parseval's theorem, where the Gowers norm $U^2[N]$ is defined analogously as in (1.17).

1.2. Applications

1.2.1. Polynomial phases

We already stated Corollary 1.3 concerning polynomial phases. But let us here mention that in a recent work of Kanigowski–Łemáńczyk–Radziwiłł [39] on the prime number theorem for analytic skew products, a key analytic input ([39, Theorem 9.1]) was that Corollary 1.3(ii) holds for $H = X^{2/3-\eta}$ (with a weaker error term of $o_{\eta \rightarrow 0}(H)$), thus going just beyond the range of validity of [49, Theorem 1.1]. Corollary 1.3 allows taking $\eta < 1/24$ with strongly logarithmic savings for the error terms. Similar remarks apply to the recent work of Kanigowski [38].

1.2.2. An application to ergodic theory

In a seminal work, Host and Kra [32] showed that, for any measure-preserving system (X, \mathcal{X}, μ, T) , any bounded functions $f_1, \dots, f_k : X \rightarrow \mathbb{C}$, and any intervals I_N whose lengths tend to infinity as $N \rightarrow \infty$,

the multiple ergodic averages

$$\frac{1}{|I_N|} \sum_{n \in I_N} f_1(T^n x) \cdots f_k(T^{kn} x)$$

converge in $L^2(\mu)$ as $N \rightarrow \infty$. Since this work, it has therefore become a natural and active question to determine for which sequences of intervals $(I_N)_N$ and weights $w : \mathbb{N} \rightarrow \mathbb{C}$ we have the L^2 -convergence of

$$\frac{1}{|I_N|} \sum_{n \in I_N} w(n) f_1(T^n x) \cdots f_k(T^{kn} x)$$

as $N \rightarrow \infty$. The case of $I_N = [1, N]$ and with the weight being the primes, that is $w(n) = 1_{\mathbb{P}}(n)$, was settled in the works of Frantzikinakis–Host–Kra [13] and Wooley–Ziegler [63] (the results of [13] in the cases $k \geq 4$ were originally conditional on the Gowers uniformity of the von Mangoldt function). Analogous results also exist for weights w supported on a sequence given by a Hardy field [12] or random sequences [14]; see also [42] for related results concerning correlation sequences $n \mapsto \int_X f_1(T^n x) \cdots f_k(T^{kn} x) d\mu(x)$. As an application of Theorem 1.5, we can extend the result on prime weights to short collections of intervals $(I_N)_N$.

Theorem 1.6 (Multiple ergodic averages over primes in short intervals). *Let $k \geq 1$, $\varepsilon > 0$ and $\kappa \in [5/8 + \varepsilon, 1 - \varepsilon]$. Let h_1, \dots, h_k be distinct positive integers. Let (X, \mathcal{X}, μ, T) be a measure-preserving system. Let $f_1, \dots, f_k : X \rightarrow \mathbb{C}$ be bounded and measurable. Then the multiple ergodic averages*

$$\mathbb{E}_{N < p \leq N + N^\kappa} f_1(T^{h_1 p} x) \cdots f_k(T^{h_k p} x)$$

converge in $L^2(\mu)$.

The results of [13] and [63] correspond to the case $\kappa = 1$. According to the best of our knowledge, Theorem 1.6 is the first result of its kind with $\kappa < 1$.

1.2.3. Linear equations in short intervals

The work of Green and the third author [17] on linear equations in primes (together with [18], [21]) provides for any finite complexity systems of linear forms $(\psi_1, \dots, \psi_t) : \mathbb{Z}^d \rightarrow \mathbb{Z}^t$ an asymptotic formula for

$$\sum_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{i=1}^t \Lambda(\psi_i(\mathbf{n})), \quad (1.24)$$

whenever $K \subset [-X, X]^d$ is a convex body containing a positive proportion of the whole cube $[-X, X]^d$, that is, $\text{vol}(K) \gg X^d$. One may ask if one can establish similar results when K is a smaller region in $[-X, X]^d$, of volume $\asymp X^{\theta d}$ with $\theta < 1$. Note that for a single linear form, this boils down to asymptotics for primes in short intervals (where the exponent $\theta = 7/12$ from [33], [29] is the best one known). Using Theorem 1.5, we can indeed give asymptotics for (1.24) in small regions.

Theorem 1.7 (Generalized Hardy–Littlewood conjecture in small boxes for finite complexity systems). *Let $X \geq 3$ and $X^{5/8+\varepsilon} \leq H \leq X^{1-\varepsilon}$ for some fixed $\varepsilon > 0$. Let $d, t, L \geq 1$. Let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms, where each $\psi_i : \mathbb{Z}^d \rightarrow \mathbb{Z}$ has the form $\psi_i(\mathbf{x}) = \dot{\psi}_i \cdot \mathbf{x} + \psi_i(0)$ with $\dot{\psi}_i \in \mathbb{Z}^d$ and $\psi_i(0) \in \mathbb{Z}$ satisfying $|\dot{\psi}_i| \leq L$ and $|\psi_i(0)| \leq LX$. Suppose that $\dot{\psi}_i$ and $\dot{\psi}_j$ are linearly independent*

whenever $i \neq j$. Let $K \subset (X, X + H]^d$ be a convex body. Then

$$\sum_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{i=1}^t \Lambda(\psi_i(\mathbf{n})) = \beta_\infty \prod_p \beta_p + o_{t,d,L}(H^d), \quad (1.25)$$

where Λ is extended as 0 to the nonpositive integers and the Archimedean factor is given by

$$\beta_\infty = \text{vol}(K \cap \Psi^{-1}(\mathbb{R}_{>0}^t))$$

and the local factors are given by

$$\beta_p = \mathbb{E}_{\mathbf{n} \in (\mathbb{Z}/p\mathbb{Z})^d} \prod_{i=1}^t \frac{p}{p-1} \mathbf{1}_{\psi_i(\mathbf{n}) \neq 0}.$$

Remark 1.8. From Theorem 1.5 and the proof method of Theorem 1.7, one can also deduce similar correlation results when in (1.25) one replaces Λ with μ or d_k (with the value of θ as in Theorem 1.5, and with no main term in the case of μ , and a different local product in the case of d_k). More specifically, under the assumption of Theorem 1.7, we have

$$\sum_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{i=1}^t \mu(\psi_i(\mathbf{n})) = o_{t,d,L}(H^d), \quad (1.26)$$

and, for a positive integer k ,

$$\sum_{\mathbf{n} \in K \cap \mathbb{Z}^d} \prod_{i=1}^t d_k(\psi_i(\mathbf{n})) = \beta_\infty \prod_p \beta_p + o_{t,d,L}(H^d \log^{t(k-1)} X),$$

where d_k is extended as 0 to the nonpositive integers and the Archimedean factor is given by

$$\beta_\infty = \int_K \prod_{i=1}^t \frac{\log_+^{k-1} \psi_i(\mathbf{x})}{(k-1)!} d\mathbf{x} = O_{t,d,L}(H^d \log^{t(k-1)} X),$$

and the local factors are given by

$$\beta_p = \frac{\mathbb{E}_{\mathbf{n} \in \mathbb{Z}_p^d} \prod_{i=1}^t d_{k,p}(\psi_i(\mathbf{n}))}{\prod_{i=1}^t \mathbb{E}_{m \in \mathbb{Z}_p} d_{k,p}(m)} = \mathbb{E}_{\mathbf{n} \in \mathbb{Z}_p^d} \prod_{i=1}^t \left(\frac{p-1}{p}\right)^{k-1} d_{k,p}(\psi_i(\mathbf{n})).$$

Here $\log_+ y := \log \max(y, 1)$, \mathbb{Z}_p is the p -adics (with the usual Haar probability measure),

$$d_{k,p}(m) = \binom{k-1 + v_p(m)}{k-1},$$

and $v_p(m)$ is the number of times p divides m . These local factors are natural extensions of the ones defined in [47, Remark 1.2] in the special case of two linear forms $\psi_1(n) = n, \psi_2(n) = n + h$.

We have the following immediate corollary to Theorem 1.7.

Corollary 1.9 (Linear equations in primes in short intervals). *Let $X \geq 3$ and $X^{5/8+\varepsilon} \leq H \leq X^{1-\varepsilon}$ for some fixed $\varepsilon > 0$. Let $d, t, L \geq 1$. Let $\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \rightarrow \mathbb{Z}^t$ be a system of affine-linear forms, where each ψ_i has the form $\psi_i(\mathbf{x}) = \dot{\psi}_i \cdot \mathbf{x} + \psi_i(0)$ with $\dot{\psi}_i \in \mathbb{Z}^d$ and $\psi_i(0) \in \mathbb{Z}$ satisfying $|\dot{\psi}_i| \leq L$ and $|\psi_i(0)| \leq LX$. Suppose that $\dot{\psi}_i$ and $\dot{\psi}_j$ are linearly independent whenever $i \neq j$. Suppose that, for every*

prime p , the system of equations $\Psi(\mathbf{n}) = 0$ is solvable with $\mathbf{n} \in ((\mathbb{Z}/p\mathbb{Z}) \setminus \{0\})^d$. Then the number of solutions to $\Psi(\mathbf{n}) = 0$ with $\mathbf{n} \in (\mathbb{P} \cap (X, X + H])^d$ is

$$\gg \frac{\text{vol}((X, X + H]^d \cap \Psi^{-1}(\mathbb{R}_{>0}^t))}{\log^d X} + o_{d,t,L} \left(\frac{H^d}{\log^d X} \right).$$

Thus, for example, for any $\varepsilon > 0$ and any large enough odd N there is a solution to

$$p_1 + p_2 + p_3 = N, \quad p_1, p_2, p_3, 2p_1 - p_2 \in \mathbb{P}$$

with $p_i \in [N/3 - N^{5/8+\varepsilon}, N/3 + N^{5/8+\varepsilon}]$. Without the condition $2p_1 - p_2 \in \mathbb{P}$, this is due to Zhan [64]. The exponent $5/8$ in Zhan's result has been improved using sieve methods (see e.g. [3]) and more recently using the transference principle [43]. It would probably be possible to use a sieve method also to improve on Corollary 1.9; it would suffice to find a suitable minorant function for $\Lambda(n)$ that has positive average and is Gowers uniform in shorter intervals. Such a minorant could be constructed with our arithmetic information using Harman's sieve method [24], but we do not do so here.

1.3. Methods of proof

We now describe (in somewhat informal terms) the general strategy of proof of our main theorems, although for various technical reasons the actual rigorous proof will not quite follow the intuitive plan that is outlined here.

To prove Theorem 1.1, the first step, which is standard, is to apply Heath–Brown's identity (Lemma 2.16) together with a combinatorial lemma regarding subsums of a finite number of non-negative reals summing to one (Lemma 2.20) to decompose μ, Λ, d_k (up to small errors) into three standard types of sums:

- (I) *Type I* sums, which are roughly of the form $\alpha * 1 = \alpha * d_1$ for some arithmetic function $\alpha: \mathbb{N} \rightarrow \mathbb{C}$ supported on some interval $[1, A_1]$ that is not too large, and with α bounded in an L^2 averaged sense.
- (I₂) *Type I₂* sums, which are roughly of the form $\alpha * d_2$ for some arithmetic function $\alpha: \mathbb{N} \rightarrow \mathbb{C}$ supported on some interval $[1, A_{I_2}]$ that is not too large, and with α bounded in an L^2 averaged sense.
- (II) *Type II* sums, which are roughly of the form $\alpha * \beta$ for some arithmetic functions $\alpha, \beta: \mathbb{N} \rightarrow \mathbb{C}$ with α supported on some interval $[A_{II}^-, A_{II}^+]$ that is neither too long nor too close to 1 or X , and with α, β bounded in an L^2 averaged sense.

This decomposition is detailed in Section 4. The precise ranges of parameters $A_1, A_{I_2}, A_{II}^-, A_{II}^+$ that arise in this decomposition depend on the choice of θ (and, in the case of d_k for small k , on the value of k); this is encoded in the combinatorial lemma given here as Lemma 2.20.

The treatment of these types of sums (in Theorem 4.2) depends on the behavior of the nilsequence $F(g(n)\Gamma)$, in particular whether it is “major arc” or “minor arc”. This splitting into different behaviors will be done somewhat differently for different types of sums.

In case of type I and type I_2 sums, one can use the equidistribution theory of nilmanifolds to essentially reduce to two cases, the *major arc case* in which the nilsequence $F(g(n)\Gamma)$ behaves like (or “pretends to be”) the constant function 1 (or some other function of small period), and the *minor arc case* in which F has mean zero and $g(n)\Gamma$ is highly equidistributed in the nilmanifold G/Γ . The contribution of type I and type I_2 major arc sums can be treated by standard methods, namely an application of Perron's formula and mean value theorems for Dirichlet series; see Section 3.

The contribution of type I minor arc sums can be treated by a slight modification of the arguments in [18], which are based on the “quantitative Leibman theorem” (Theorem 2.7 below) that characterizes when a nilsequence is equidistributed, as well as a classical lemma of Vinogradov (Lemma 2.3 below)

that characterizes when a polynomial modulo 1 is equidistributed. (Actually it will be convenient to rely primarily on a corollary of Lemma 2.3 that asserts that if typical dilates of a polynomial are equidistributed modulo 1, then the polynomial itself is equidistributed modulo 1: see Corollary 2.4 below.)

Our treatment of type I_2 minor arc sums is more novel. A model case is that of treating the d_2 -type correlation

$$\sum_{X < n \leq X+H} d_2(n) \overline{F}(g(n)\Gamma).$$

From the definition of the divisor function d_2 , we can expand this sum as a double sum

$$\sum_{n,m:X < nm \leq X+H} \overline{F}(g(nm)\Gamma). \quad (1.27)$$

We are not able to obtain non-trivial estimates on such sums in the regime $H \leq X^{1/3}$. However, when $H \geq X^{1/3+\varepsilon}$, it turns out by elementary geometry of numbers that the hyperbola neighborhood $\{(n, m) \in \mathbb{Z}^2 : X < nm \leq X + H\}$ may be partitioned³ into arithmetic progressions $P \subset \mathbb{Z}^2$ that mostly have non-trivial length; see Theorem 8.1 for a precise statement. This decomposition lets us efficiently decompose the sum (1.27) into short sums of the form

$$\sum_{(n,m) \in P} \overline{F}(g(nm)\Gamma)$$

that turn out to exhibit cancellation for most progressions P in the type I_2 minor arc case, mainly thanks to the quantitative Leibman theorem (Theorem 2.7) and a corollary of the Vinogradov lemma (Corollary 2.4); see Section 8.

It remains to handle the contribution of type II sums, which are of the form

$$\sum_{X < n \leq X+H} \alpha * \beta(n) \overline{F}(g(n)\Gamma)$$

which we can expand as

$$\sum_{A_{II}^- \leq a \leq A_{II}^+} \alpha(a) \sum_{X/a < b \leq X/a+H/a} \beta(b) \overline{F}(g(ab)\Gamma). \quad (1.28)$$

To treat these sums, we can use a Fourier decomposition and the equidistribution theory of nilmanifolds to reduce (roughly speaking) to treating the following three special cases of these sums:

- *Type II major arc sums* that are essentially of the form

$$\sum_{X < n \leq X+H} \alpha * \beta(n) n^{iT}$$

for some real number $T = X^{O(1)}$ of polynomial size (one can also consider generalizations of such sums when the n^{iT} factor is twisted by an additional Dirichlet character χ of bounded conductor).

- *Abelian Type II minor arc sums* in which $F(g(n)\Gamma) = e(P(n))$ is a polynomial phase that does not “pretend” to be a character n^{iT} (or more generally $\chi(n)n^{iT}$ for some Dirichlet character χ of bounded conductor) in the sense that the Taylor coefficients of $e(P(n))$ around X do not align with the corresponding coefficients of such characters.

³This partition is reminiscent of the classical Hardy–Littlewood partition of the unit circle into major and minor arcs, except that we are partitioning (a neighborhood of) a hyperbola rather than a circle.

- *Non-abelian Type II minor arc sums*, in which $g(n)\Gamma$ is highly equidistributed in a nilmanifold G/Γ arising from a non-abelian nilpotent group G , and F exhibits non-trivial oscillation in the direction of the center $Z(G)$ of G (which one can reduce to be one-dimensional).

One can treat the contribution of the type *II* major arc sums by applying Perron's formula and Dirichlet polynomial estimates of Baker–Harman–Pintz [4] in the regime, so long as one actually has a suitable triple convolution (with one of the subfactors having well-controlled correlations with n^{iT}); see Lemma 3.5. As already implicitly observed by Zhan [64], this case can be treated (with favorable choices of parameters) for any of the three functions μ, Λ, d_k in the case $\theta = 5/8$. As observed in [50], in the case of the Möbius function μ , it is possible to lower θ to $3/5$ and still obtain triple convolution structure after removing a small exceptional error term from μ (which is responsible for the final dis-correlation bounds not saving arbitrary powers of $\log X$); see Lemma 4.5.

It remains to treat the contribution of non-abelian and abelian type *II* minor arc sums. It turns out that we will be able to establish good estimates for such sums (1.28) in the regime

$$X^\varepsilon \frac{X}{H} \ll A_{II}^- < A_{II}^+ \ll X^{-\varepsilon} H.$$

In this regime, the inner intervals $(X/a, X/a + H/a]$ in (1.28) have non-negligible length (at least X^ε), and furthermore they exhibit non-trivial overlap with each other ($(X/a, X/a + H/a]$ will essentially be identical to $(X/a', X/a' + H/a']$ whenever $a' = (1 + O(X^{-\varepsilon} \frac{H}{X})) a$).

As a consequence, many of the dilated nilsequences $b \mapsto \overline{F}(g(ab)\Gamma)$ appearing in (1.28) will correlate with the same portion of the sequence β . To handle this situation we introduce a nilsequence version of the large sieve inequality in Proposition 2.15, which we establish with the aid of the equidistribution theory for nilsequences, as well as Goursat's lemma. The upshot of this large sieve inequality is that for many nearby pairs a', a there is an algebraic relation between the sequences $b \mapsto g(ab)$ and $b \mapsto g(a'b)$, namely that one has an identity of the form

$$g(a' \cdot) = \varepsilon_{aa'} g(a \cdot) \gamma_{aa'}$$

where $\varepsilon_{aa'}: \mathbb{Z} \rightarrow G$ is a “smooth” polynomial map and $\gamma_{aa'}: \mathbb{Z} \rightarrow G$ is a “rational” polynomial map; see (6.7) for a precise statement. This can be viewed as an assertion that the map g is “approximately dilation-invariant” in some weak sense. This turns out to imply a non-trivial lack of two-dimensional equidistribution for the map

$$(a, a', b, b') \mapsto (g(ab)\Gamma, g(ab')\Gamma, g(a'b)\Gamma, g(a'b')\Gamma)$$

which is incompatible with the non-abelian nature of G thanks to a commutator argument of Furstenberg and Weiss [15]; see Section 6. This resolves the non-abelian case. In the abelian case, one can replace the maps g by the ordinary polynomials P , and one can then proceed by adapting the arguments by the first two authors in [49] to show that $e(P(n))$ necessarily “pretends” to be like a character n^{iT} , which resolves the abelian type *II* minor arc case. Combining all these cases yields Theorem 1.1.

1.3.1. The result on Gowers norms

The proof of Theorem 1.5 (in Section 9) requires in addition to Theorem 1.1 and the inverse theorem for the Gowers norms also a construction of pseudorandom majorants for (W -tricked versions of) Λ and d_k over *short intervals* $(X, X + H]$. By this we mean functions ν_1, ν_2 that majorize the functions Λ, d_k (after W -tricking and suitable normalization), and such that $\nu_i - 1$ restricted to $(X, X + H]$ is Gowers uniform. In the case of long intervals (that is, $H = X$), the existence of such majorants is well known from works of Green and the third author [16] and Matthiesen [52]. Fortunately, it turns out that the structure of these well-known majorants as type I sums of small “level” enables us to show that they work as majorants also over short intervals $(X, X + H]$; see Lemmas 9.5 and 9.6. These lemmas

combined with the implementation of the W -trick (which in the case of d_k requires additionally two simple lemmas, namely Lemmas 9.8 and 9.9) leads to the proof of Theorem 1.5.

Remark 1.10. In this remark we discuss the obstructions to improving the value of θ in the various components of Theorem 1.1. In most of these results, the primary obstruction arises (roughly speaking) from portions of μ , Λ , or d_k that look something like

$$1_{(X^{\alpha_1}, 2X^{\alpha_1}]} * \cdots * 1_{(X^{\alpha_m}, 2X^{\alpha_m}]} \quad (1.29)$$

for various tuples $(\alpha_1, \dots, \alpha_m)$ of positive real numbers that add up to 1. More specifically:

- (a) For the $\theta = 5/8$ results in Theorem 1.1(i)–(iii), the primary obstruction arises from convolutions (1.29) with $(\alpha_1, \dots, \alpha_m)$ equal to $(1/4, 1/4, 1/4, 1/4)$, when correlated against characters n^T with $T \asymp X^{O(1)}$, as this lies just outside the reach of our twisted major arc type I and type II estimates when θ goes below $5/8$. This obstruction was already implicitly observed by Zhan [64].
- (b) For the $\theta = 3/5$ result in Theorem 1.1(iv), the primary obstruction are convolutions (1.29) with $(\alpha_1, \dots, \alpha_m)$ equal to $(2/5, 1/5, 1/5, 1/5)$ or $(1/5, 1/5, 1/5, 1/5, 1/5)$, when correlated against “minor arc” nilsequences, such as $e(\alpha n)$ for some minor arc α . Such convolutions become just out of reach of our type I, type II, and type I_2 estimates when θ goes below $3/5$. This obstruction was already observed in [50].
- (c) For the $\theta = 1/3$ result in Theorem 1.1(iii), the primary obstruction is of a different nature from the preceding cases: it is that our treatment of minor arcs in this case relies crucially on the ability to partition the neighborhood of a hyperbola into arithmetic progressions (see Theorem 8.1), and this partition is no longer available in any useful form once θ goes below $1/3$.
- (d) For the $\theta = 5/9$ result in Theorem 1.1(iii), the primary obstruction arises from convolutions (1.29) with $(\alpha_1, \dots, \alpha_m)$ equal to $(1/3, 1/3, 1/3)$, when correlated against minor arc nilsequences, for reasons similar to those in the previous case (c).

1.4. Notation

The parameter X should be thought of as being large.

We use $Y \ll Z$, $Y = O(Z)$, or $Z \gg Y$ to denote the estimate $|Y| \leq CZ$ for some constant C . If we wish to permit this constant to depend (possibly ineffectively) on one or more parameters we shall indicate this by appropriate subscripts, thus for instance $O_{\varepsilon, A}(Z)$ denotes a quantity bounded in magnitude by $C_{\varepsilon, A}Z$ for some quantity $C_{\varepsilon, A}$ depending only on ε, A . We write $Y \asymp Z$ for $Y \ll Z \ll Y$. When working with d_k , all implied constants are permitted to depend on k . We also write $y \sim Y$ to denote the assertion $Y < y \leq 2Y$.

If x is a real number (resp. an element of \mathbb{R}/\mathbb{Z}), we write $e(x) := e^{2\pi i x}$ and let $\|x\|_{\mathbb{R}/\mathbb{Z}}$ denote the distance of x to the nearest integer (resp. zero).

We use 1_E to denote the indicator of an event E , thus 1_E equals 1 when E is true and 0 otherwise. If S is a set, we write 1_S for the indicator function $1_S(n) := 1_{n \in S}$.

Unless otherwise specified, all sums range over natural number values, except for sums over p which are understood to range over primes. We use $d|n$ to denote the assertion that d divides n , (n, m) to denote the greatest common divisor of n and m , $n = a \pmod{q}$ to denote the assertion that n and a have the same residue mod q , and $f * g(n) := \sum_{d|n} f(d)g(n/d)$ to denote the Dirichlet convolution of two arithmetic functions $f, g : \mathbb{N} \rightarrow \mathbb{C}$.

The *height* of a rational number a/b with a, b coprime is defined as $\max(|a|, |b|)$.

2. Basic tools

2.1. Total variation

The notion of maximal summation defined in (1.4) interacts well with the notion of total variation, which we now define.

Definition 2.1 (Total variation). Given any function $f : P \rightarrow \mathbb{C}$ on an arithmetic progression P , the *total variation norm* $\|f\|_{\text{TV}(P)}$ is defined by the formula

$$\|f\|_{\text{TV}(P)} := \sup_{n \in P} |f(n)| + \sup_{n_1 < \dots < n_k} \sum_{j=1}^{k-1} |f(n_{j+1}) - f(n_j)|$$

where the second supremum ranges over all increasing finite sequences $n_1 < \dots < n_k$ in P and all $k \geq 1$. We remark that in this finitary setting one can simply take n_1, \dots, n_k to be the elements of P in increasing order, if one wishes. We adopt the convention that $\|f\|_{\text{TV}(P)} = 0$ when P is empty. For any natural number $q \geq 1$, we also define

$$\|f\|_{\text{TV}(P;q)} := \sum_{a \in \mathbb{Z}/q\mathbb{Z}} \|f\|_{\text{TV}(P \cap (a+q\mathbb{Z}))}.$$

Informally, if f is bounded in $\text{TV}(P; q)$ norm, then f does not vary much on each residue class modulo q in P . From the fundamental theorem of calculus we see that if $f : I \rightarrow \mathbb{C}$ is a continuously differentiable function then

$$\|f\|_{\text{TV}(P)} \ll \sup_{t \in I} |f(t)| + \int_I |f'(t)| dt \tag{2.1}$$

for all arithmetic progressions P in I . Also, from the identity $ab - a'b' = (a - a')b + (b - b')a'$ we see that

$$\|fg\|_{\text{TV}(P;q)} \ll \|f\|_{\text{TV}(P;q)} \|g\|_{\text{TV}(P;q)} \tag{2.2}$$

for any functions $f, g : P \rightarrow \mathbb{C}$ defined on an arithmetic progression, and any $q \geq 1$.

We can now record some basic properties of maximal summation:

Lemma 2.2 (Basic properties of maximal sums).

(i) (*Triangle inequalities*) For any subprogression P' of an arithmetic progression P , and any $f : P \rightarrow \mathbb{C}$ we have

$$\left| \sum_{n \in P} f(n) 1_{P'}(n) \right|^* = \left| \sum_{n \in P'} f(n) \right|^* \leq \left| \sum_{n \in P} f(n) \right|^*$$

and

$$\left| \sum_{n \in P} f(n) \right| \leq \left| \sum_{n \in P} f(n) \right|^* \leq \sum_{n \in P} |f(n)|.$$

If P can be partitioned into two subprogressions as $P = P_1 \uplus P_2$, then

$$\left| \sum_{n \in P} f(n) \right|^* \leq \left| \sum_{n \in P_1} f(n) \right|^* + \left| \sum_{n \in P_2} f(n) \right|^*. \tag{2.3}$$

Finally, the map $f \mapsto |\sum_{n \in P} f(n)|^*$ is a seminorm.

(ii) (Local stability) If $x_0 \in \mathbb{R}$, $H > 0$, and $f: \mathbb{Z} \rightarrow \mathbb{C}$, then

$$\left| \sum_{x_0 < n \leq x_0 + H} f(n) \right|^* \leq \frac{2}{H} \int_{x_0 - H/2}^{x_0 + H/2} \left| \sum_{x < n \leq x + H} f(n) \right|^* dx.$$

(iii) (Summation by parts) Let P be an arithmetic progression, and let $f, g: P \rightarrow \mathbb{C}$ be functions. Then we have

$$\left| \sum_{n \in P} f(n)g(n) \right|^* \leq \|g\|_{\text{TV}(P)} \left| \sum_{n \in P} f(n) \right|^* \quad (2.4)$$

and more generally

$$\left| \sum_{n \in P} f(n)g(n) \right|^* \leq \|g\|_{\text{TV}(P; q)} \left| \sum_{n \in P} f(n) \right|^* \quad (2.5)$$

for any $q \geq 1$.

Proof. The claims (i) all follow easily the triangle inequality and the observation that the intersection of two arithmetic progressions is again an arithmetic progression; for instance, (2.3) follows from the observation that any subprogression P' of P is partitioned into subprogressions $P' \cap P_1, P' \cap P_2$ of P_1, P_2 respectively. To prove (ii), we observe from (i) that for any $0 < t < H/2$ we have

$$\begin{aligned} \left| \sum_{x_0 < n \leq x_0 + H} f(n) \right|^* &\leq \left| \sum_{x_0 < n \leq x_0 + H/2} f(n) \right|^* + \left| \sum_{x_0 + H/2 < n \leq x_0 + H} f(n) \right|^* \\ &\leq \left| \sum_{x_0 - t < n \leq x_0 - t + H} f(n) \right|^* + \left| \sum_{x_0 + t < n \leq x_0 + t + H} f(n) \right|^* \end{aligned}$$

and the claim then follows by averaging in t .

To prove the first claim (2.4) of (iii), it will suffice by the monotonicity properties of total variation and maximal sums to show that

$$\left| \sum_{n \in P'} f(n)g(n) \right| \leq \|g\|_{\text{TV}(P')} \left| \sum_{n \in P'} f(n) \right|^* \quad (2.6)$$

for all subprogressions P' of P . Clearly we may assume P' is non-empty. If we order the elements of P' as $n_1 < n_2 < \dots < n_k$, then from summation by parts we have

$$\sum_{n \in P'} f(n)g(n) = \sum_{j=1}^{k-1} (g(n_j) - g(n_{j+1})) \sum_{i=1}^j f(n_i) + g(n_k) \sum_{i=1}^k f(n_i).$$

Since each segment $\{n_1, \dots, n_j\}$ of P' is again a subprogression of P' , we have from the triangle inequality that

$$\left| \sum_{n \in P'} f(n)g(n) \right| \leq \sum_{j=1}^{k-1} |g(n_j) - g(n_{j+1})| \left| \sum_{n \in P'} f(n) \right|^* + |g(n_k)| \left| \sum_{n \in P'} f(n) \right|^*$$

and the claim (2.6) now follows from Definition 2.1. Thus (2.4) holds. To prove the second claim (2.5), partition P into subprogressions $P \cap (a + q\mathbb{Z})$, apply (2.4) to each subprogression, and sum using (i). \square

2.2. Vinogradov lemma

If $P: \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ is a polynomial of degree d , and I is an interval of length $|I| \geq 1$, we define the *smoothness norm*

$$\|P\|_{C^\infty(I)} := \sup_{0 \leq j \leq d} \sup_{n \in I} |I|^j \|\partial_1^j P(n)\|_{\mathbb{R}/\mathbb{Z}}$$

where ∂_1 is the difference operator $\partial_1 P(n) := P(n) - P(n-1)$. We remark that this definition deviates very slightly from that in [19, Definition 2.7]; in particular, we allow the index j to equal zero and we allow n to range over I rather than being set to the origin. We use the same notation $\|P\|_{C^\infty(I)}$ for a polynomial $P: \mathbb{Z} \rightarrow \mathbb{R}$ after reducing its coefficients modulo 1.

The following lemma asserts, roughly speaking, that a polynomial P is (somewhat) equidistributed unless it is smooth.

Lemma 2.3 (Vinogradov lemma). *Let $0 < \varepsilon, \delta < 1/2$, $d \geq 0$, and let $P: \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ be a polynomial of degree at most d . Let I be an interval of length $|I| \geq 1$, and suppose that*

$$\|P(n)\|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon$$

for at least $\delta|I|$ integers $n \in I$. Then either $\delta \ll_d \varepsilon$, or else one has

$$\|qP\|_{C^\infty(I)} \ll_d \delta^{-O_d(1)} \varepsilon$$

for some integer $1 \leq q \ll_d \delta^{-O_d(1)}$.

Proof. By applying a translation, we may assume that I takes the form $(0, N]$ for some $N \geq 1$. We may also assume $\varepsilon \leq \delta/2$, since we are clearly done otherwise. We may now invoke [19, Lemma 4.5] to conclude that there exists $1 \leq q \ll_d \delta^{-O_d(1)} \varepsilon$ such that

$$\sup_{1 \leq j \leq d} \sup_{n \in I} |I|^j \|q\partial_1^j P(n)\|_{\mathbb{R}/\mathbb{Z}} \ll_d \delta^{-O_d(1)} \varepsilon. \quad (2.7)$$

This is almost what we want, except that we have to also control the $j = 0$ contribution. But from hypothesis we have at least one $n_0 \in I$ such that $\|P(n_0)\|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon$, and from (2.7) we have $\|q\partial_1 P(n)\|_{\mathbb{R}/\mathbb{Z}} \ll_d \delta^{-O_d(1)} |I|^{-1} \varepsilon$ for all $n \in I$. From the triangle inequality we then conclude that

$$\|qP(n)\|_{\mathbb{R}/\mathbb{Z}} \ll_d \delta^{-O_d(1)} \varepsilon$$

for all $n \in I$, and the claim follows. \square

The following handy corollary of Lemma 2.3 asserts, roughly speaking, that if many dilates of a polynomial are smooth, then the polynomial itself is smooth.

Corollary 2.4 (Concatenating dilated smoothness). *Let $0 < \delta < 1/2$, $d \geq 0$, and let $P: \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ be a polynomial of degree at most d . Let $A \geq 1$, let I be an interval with $|I| \geq 2A$, and suppose that*

$$\|P(a \cdot)\|_{C^\infty(\frac{1}{a}I)} \leq \frac{1}{\delta} \quad (2.8)$$

for at least δA integers a in $[A, 2A]$, where $\frac{1}{a}I := \{\frac{t}{a} : t \in I\}$ is the dilate of I by $\frac{1}{a}$. Then either $|I| \ll_d \delta^{-O_d(1)} A$, or else one has

$$\|qP\|_{C^\infty(I)} \ll_d \delta^{-O_d(1)}$$

for some integer $1 \leq q \ll_d \delta^{-O_d(1)}$.

Proof. We allow all implied constants to depend on d . We may assume that $|I| \geq C\delta^{-C}A$ for a large constant C depending on d , as the claim is immediate otherwise.

We now claim that for each $0 \leq j \leq d$ that there exists a decomposition

$$P = P_j + Q_j \quad (2.9)$$

where $P_j: \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ is a polynomial of degree at most d with

$$\|q_j P_j\|_{C^\infty(I)} \ll \delta^{-O(1)} \quad (2.10)$$

for some $1 \leq q_j \ll \delta^{-O(1)}$, and $Q_j: \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ is a polynomial of degree at most j . For $j = d$ one can simply set $P_d = 0$ and $Q_d = P$. Now suppose by downward induction that $0 \leq j < d$ and the claim has already been proven for $j + 1$. From (2.10) (for P_{j+1}) we have

$$\|q_{j+1} P_{j+1}\|_{C^\infty(I)} \ll \delta^{-O(1)}.$$

Routine Taylor expansion then gives

$$\|q_{j+1} P_{j+1}(a \cdot)\|_{C^\infty(\frac{1}{a}I)} \ll \delta^{-O(1)}$$

for all $a \in [A, 2A]$, thus by (2.8) and the triangle inequality we have

$$\|q_{j+1} Q_{j+1}(a \cdot)\|_{C^\infty(\frac{1}{a}I)} \ll \delta^{-O(1)}$$

for $\geq \delta A$ choices of $a \in [A, 2A]$.

Now write $Q_{j+1}(n) = \alpha_{j+1} \binom{n}{j+1} + Q_j(n)$ where Q_j is of degree at most j . Taking $j+1$ -fold derivatives, we see that

$$\|a^{j+1} q_{j+1} \alpha_{j+1}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)} (A/|I|)^{j+1}$$

for $\geq \delta A$ choices of $a \in [A, 2A]$. Applying Lemma 2.3 to the polynomial $a \rightarrow a^{j+1} q_{j+1} \alpha_{j+1}$ (and recalling that $|I|/A \geq C\delta^{-C}$ for a suitably large C by assumption), we conclude that there is $1 \leq q \ll \delta^{O(1)}$ such that

$$\|q(\cdot)^{j+1} q_{j+1} \alpha_{j+1}\|_{C^\infty([A, 2A])} \ll \delta^{-O(1)} (A/|I|)^{j+1}$$

and hence on taking $j+1$ -fold derivatives

$$\|(j+1)! q q_{j+1} \alpha_{j+1}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)} |I|^{-j-1}.$$

If one then sets $q_j := (j+1)! q q_{j+1}$ and $P_j(n) := P_{j+1}(n) + \alpha_{j+1} \binom{n}{j+1}$, we obtain the decomposition (2.9), and (2.10) follows from the triangle inequality. This closes the induction. Applying the claim with $j = 0$, we obtain the corollary. \square

2.3. Equidistribution on nilmanifolds

We now recall some of the basic notation and results from [19] concerning equidistribution of polynomial maps on nilmanifolds.

Definition 2.5 (Filtered group). Let $d \geq 1$. A *filtered group* is a group G (which we express in multiplicative notation $G = (G, \cdot)$ unless explicitly indicated otherwise) equipped with a filtration $G_\bullet = (G_i)_{i=0}^\infty$ of nested groups $G \geq G_0 \geq G_1 \geq \dots$ such that $[G_i, G_j] \leq G_{i+j}$ for all $i, j \geq 0$. We say that this group has degree at most d if G_i is trivial for all $i > d$. Given a filtered group of degree at most d , a *polynomial map* $g: \mathbb{Z} \rightarrow G$ from \mathbb{Z} to G is a map of the form $g(n) = g_0 g_1^{(n)} \cdots g_d^{(n)}$ where $g_i \in G_i$ for all $0 \leq i \leq d$; the collection of such maps will be denoted $\text{Poly}(\mathbb{Z} \rightarrow G)$.

The well-known Lazard–Leibman theorem (see e.g., [19, Proposition 6.2]) asserts that $\text{Poly}(\mathbb{Z} \rightarrow G)$ is a group under pointwise multiplication; also, from [19, Corollary 6.8] we see that if $g: \mathbb{Z} \rightarrow G$ is a polynomial map then so is $n \mapsto g(an + b)$ for any integers a, b .

If G is a simply connected nilpotent Lie group, we write $\log G$ for the Lie algebra. From the Baker–Campbell–Hausdorff formula⁴ (see e.g. [22, Theorem 3.3]) we see that the exponential map $\exp: \log G \rightarrow G$ is a homeomorphism and hence has an inverse $\log: G \rightarrow \log G$.

Definition 2.6 (Filtered nilmanifolds). Let $d, D \geq 1$ and $0 < \delta < 1$. A *filtered nilmanifold* G/Γ of degree at most d , dimension D , and complexity at most $1/\delta$ consists of the following data:

- A filtered simply connected nilpotent Lie group G of dimension D equipped with a filtration $G_\bullet = (G_i)_{i=0}^\infty$ of degree at most d , with $G_0 = G_1 = G$ and all G_i closed connected subgroups of G .
- A lattice (i.e., a discrete cocompact subgroup Γ) of G , with the property that $\Gamma_i := \Gamma \cap G_i$ is a lattice of G_i for all $i \geq 0$.
- A linear basis X_1, \dots, X_D (which we call a *Mal'cev basis*) of $\log G$.

Furthermore we assume the following axioms:

- For all $1 \leq i, j \leq D$ we have $[X_i, X_j] = \sum_{i,j < k \leq D} c_{ijk} X_k$ for some rational numbers c_{ijk} of height at most $1/\delta$.
- For all $0 \leq i \leq d$, the vector space G_i is spanned by the X_j with $D - \dim G_i < j \leq D$.
- We have $\Gamma = \{\exp(n_1 X_1) \cdots \exp(n_D X_D) : n_1, \dots, n_D \in \mathbb{Z}\}$.

It is easy to see that G/Γ has the structure of a smooth compact D -dimensional manifold, which we equip with a probability Haar measure $d\mu_{G/\Gamma}$. We define the metric d_G on G to be the largest right-invariant metric such that $d_G(\exp(t_1 X_1) \cdots \exp(t_D X_D), 1) \leq \sup_{1 \leq i \leq D} |t_i|$ for all $t_1, \dots, t_D \in \mathbb{R}$. We then define a metric $d_{G/\Gamma}$ on G/Γ by the formula $d_{G/\Gamma}(x, y) := \inf_{g \in \Gamma, h \in \Gamma} d_G(g, h)$. The Lipschitz norm of a function $F: G/\Gamma \rightarrow \mathbb{C}$ is defined to be the quantity

$$\sup_{x \in G/\Gamma} |F(x)| + \sup_{x, y \in G/\Gamma: x \neq y} \frac{|F(x) - F(y)|}{d_{G/\Gamma}(x, y)}.$$

A *horizontal character* η associated to a filtered nilmanifold is a continuous homomorphism $\eta: G \rightarrow \mathbb{R}$ that maps Γ to the integers.

An element γ of G is said to be *M-rational* for some $M \geq 1$ if one has $\gamma^r \in \Gamma$ for some natural number $1 \leq r \leq M$. A subnilmanifold G'/Γ' of G/Γ (thus G' is a closed connected subgroup of G with $\Gamma'_i := G'_i \cap \Gamma$ cocompact in G'_i for all i) is said to be *M-rational* if each element $X'_1, \dots, X'_{\dim G'}$ of the Mal'cev basis associated to G is a linear combination of the X_i with all coefficients rational of height at most M .

A *rational subgroup* G' of complexity at most $1/\delta$ is a closed connected subgroup of G with the property that $\log G'$ admits a linear basis consisting of $\dim G'$ vectors of the form $\sum_{i=1}^D a_i X_i$, where each a_i is a rational of height at most $1/\delta$.

It is easy to see that every horizontal character takes the form $\eta(g) = \lambda(\log g)$ for some linear functional $\lambda: \log G \rightarrow \mathbb{R}$ that annihilates $\log[G, G]$ and maps $\log \Gamma$ to the integers. From this one can verify that the number of horizontal characters of Lipschitz norm at most $1/\delta$ is at most $O_{d,D}(\delta^{-O_{d,D}(1)})$.

From several applications of Baker–Campbell–Hausdorff formula we see that if G has degree at most d and $\gamma_1, \gamma_2 \in G$ are *M*-rational, then $\gamma_1 \gamma_2$ is $O_d(M^{O_d(1)})$ -rational.

We have the following basic dichotomy between equidistribution and smoothness:

Theorem 2.7 (Quantitative Leibman theorem). Let $0 < \delta < 1/2$, let $d, D \geq 1$, let I be an interval with $|I| \geq 1$, and let G/Γ be a filtered nilmanifold of degree at most d , dimension at most D , and

⁴The reader may consult [48, Appendix B] for more details on the use of the Baker–Campbell–Hausdorff formula in the context of quantitative nilmanifold theory.

complexity at most $1/\delta$. Let $F: G/\Gamma \rightarrow \mathbb{C}$ be Lipschitz of norm at most $1/\delta$ and of mean zero (i.e., $\int_{G/\Gamma} F \, d\mu_{G/\Gamma} = 0$). Suppose that $g: \mathbb{Z} \rightarrow G$ is a polynomial map with

$$\left| \sum_{n \in I} F(g(n)\Gamma) \right|^* \geq \delta |I|.$$

Then there exists a non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm $O_{d,D}(\delta^{-O_{d,D}(1)})$ such that

$$\|\eta \circ g\|_{C^\infty(I)} \ll_{d,D} \delta^{-O_{d,D}(1)}.$$

Proof. By applying a translation we may assume $I = (0, N]$ for some $N \geq 1$. The claim now follows from [59, Theorem 3.5]. \square

Let G/Γ be a filtered nilmanifold of dimension D and complexity at most $1/\delta$, and let G' be a rational subgroup of complexity at most $1/\delta$. In [19, Proposition A.10] it is shown that G'/Γ' can be equipped with the structure of a filtered nilmanifold of complexity $O_{d,D}(\delta^{-O_{d,D}(1)})$, where $\Gamma' := \Gamma \cap G'$, $G'_i := G_i \cap G'$, and the metrics $d_G, d_{G'}$ are comparable on G' up to factors of $O_{d,D}(\delta^{-O_{d,D}(1)})$; one can view G'/Γ' as a subnilmanifold of G/Γ .

One can easily verify from basic linear algebra and the Baker–Campbell–Hausdorff formula that the following groups are rational subgroups of G of complexity $O_{d,D}(\delta^{-O_{d,D}(1)})$:

- The groups G_i in the filtration for $0 \leq i \leq d$.
- The kernel $\ker \eta$ of any horizontal character η of Lipschitz norm $O_{d,D}(\delta^{-O_{d,D}(1)})$.
- The center $Z(G) = \{\exp(X) : X \in \log G; [X, Y] = 0 \ \forall Y \in \log G\}$ of G .
- The intersection $G' \cap G''$ or commutator $[G', G'']$ of two rational subgroups G', G'' of G of complexity $O_{d,D}(\delta^{-O_{d,D}(1)})$.
- The product $G'N$ of two rational subgroups G', N of G of complexity $O_{d,D}(\delta^{-O_{d,D}(1)})$, with N normal.

We can quotient out a filtered nilmanifold by a normal subgroup to obtain another filtered nilmanifold, with polynomial control on complexity:

Lemma 2.8 (Quotienting by a normal subgroup). *Let G/Γ be a filtered nilmanifold of degree at most d , dimension D and complexity at most $1/\delta$. Let N be a normal rational subgroup of G of complexity at most $1/\delta$, and let $\pi: G \mapsto G/N$ be the quotient map. Then $\pi(G)/\pi(\Gamma)$ can be given the structure of a filtered nilmanifold of degree at most d , dimension $D - \dim N$, and complexity $O_{d,D}(\delta^{-O_{d,D}(1)})$, such that*

$$d_{\pi(G)}(\pi(g), \pi(h)) \asymp_{d,D} \delta^{-O_{d,D}(1)} \inf_{n \in \mathbb{N}} d_G(g, nh) \tag{2.11}$$

for any $g, h \in G$.

Proof. We allow all implied constants to depend on d, D . Let $\tilde{\pi}: \log G \rightarrow \log G/\log N \equiv \log(G/N)$ be the quotient map of $\log G$ by the Lie algebra ideal $\log N$, then $\pi \circ \exp = \exp \circ \tilde{\pi}$. For each $0 \leq i \leq d$, the vectors $\tilde{\pi}(X_j)$ for $D - \dim G_i < j \leq D$ span the linear subspace $\tilde{\pi}(\log G_i)$ of $\log(G/N)$, and the linear relations between those vectors are generated by $O(1)$ equations with coefficients rational of height $O(\delta^{-O(1)})$. From this and linear algebra we may find a basis $\tilde{X}_1, \dots, \tilde{X}_{\dim(G/N)}$ of $\log(G/N)$ such that for each $0 \leq i \leq d$, $\tilde{\pi}(\log G_i)$ is the span of \tilde{X}_j for $\dim(G/N) - \dim \tilde{\pi}(\log G_i) < j \leq \dim(G/N)$, and each \tilde{X}_j is a linear combination of the $\tilde{\pi}(X_1), \dots, \tilde{\pi}(X_D)$ with coefficients rational of height $O(\delta^{-O(1)})$. Meanwhile, $\pi(\Gamma)$ is generated by $\pi(X_1), \dots, \pi(X_D)$. From this and the Baker–Campbell–Hausdorff formula we see that the basis $\tilde{X}_1, \dots, \tilde{X}_{\dim(G/N)}$ is a $O(\delta^{-O(1)})$ -rational weak basis for $\pi(G)/\pi(\Gamma)$ in the sense of [19, Definition A.7]. Applying [19, Proposition A.9] to this weak basis, we obtain a Mal'cev basis that gives $\pi(G)/\pi(\Gamma)$ the structure of a filtered nilmanifold with the stated degree, dimension, and

complexity. It remains to establish the bound (2.11). By right translation invariance we can take g to be the identity. For the upper bound, it suffices (since π is N -invariant) to show that

$$d_{\pi(G)}(1, \pi(h)) \ll \delta^{-O(1)} d_G(1, h),$$

but this follows from the fact that $\tilde{\pi}: \log G \rightarrow \tilde{\pi}(\log G)$ has operator norm $O(\delta^{-O(1)})$ when using the X_1, \dots, X_D basis for $\log G$ and the $\tilde{X}_1, \dots, \tilde{X}_{\dim(G/N)}$ basis for $\tilde{\pi}(\log G)$ to define norms.

Now we need to establish the lower bound. By [19, Lemma A.4] it suffices to show that

$$\|Y\| \gg \delta^{-O(1)} \inf_{Y' \in \tilde{\pi}^{-1}(Y)} \|Y'\|$$

for any $Y \in \tilde{\pi}(\log G)$, where again we use the norm given by the X_1, \dots, X_D basis for $\log G$ and the $\tilde{X}_1, \dots, \tilde{X}_{\dim(G/N)}$. But this is easily verified for each $Y = \tilde{X}_i$, and the claim then follows by linearity. \square

A *central frequency* is a continuous homomorphism $\xi: Z(G) \rightarrow \mathbb{R}$ which maps $Z(G) \cap \Gamma$ to the integers \mathbb{Z} (that is to say, a horizontal character on $Z(G)$, or a Fourier character of the central torus $Z(G)/(Z(G) \cap \Gamma)$). A function $F: G/\Gamma \rightarrow \mathbb{C}$ is said to *oscillate with central frequency* ξ if one has the identity

$$F(zx) = e(\xi(z))F(x)$$

for all $x \in G/\Gamma$ and $z \in Z(G)$. As with horizontal characters, the number of central frequencies ξ of Lipschitz norm at most $1/\delta$ is $O_{d,D}(\delta^{-O_{d,D}(1)})$. If ξ is such a central frequency, one can readily verify that the kernel $\ker \xi$ is a rational normal subgroup of G of complexity $O_{d,D}(\delta^{-O_{d,D}(1)})$.

We have the following convenient decomposition⁵ (cf., [19, Lemma 3.7]):

Proposition 2.9 (Central Fourier approximation). *Let $d, D \geq 1$ and $0 < \delta < 1$. Let G/Γ be a filtered nilmanifold of degree at most d , dimension D , and complexity at most $1/\delta$. Let $F: G/\Gamma \rightarrow \mathbb{C}$ be a Lipschitz function of norm at most $1/\delta$. Then we can decompose*

$$F = \sum_{\xi} F_{\xi} + O(\delta)$$

where ξ ranges over central frequencies of Lipschitz norm at most $O_{d,D}(\delta^{-O_{d,D}(1)})$, and each F_{ξ} has Lipschitz norm $O_{d,D}(\delta^{-O_{d,D}(1)})$ and oscillates with central frequency ξ . Furthermore, if F has mean zero, then so do all of the F_{ξ} .

Proof. We allow all implied constants to depend on d, D . Since $Z(G)/(Z(G) \cap \Gamma)$ is an abelian filtered nilmanifold of complexity $O(\delta^{-O(1)})$, it can be identified with a torus $\mathbb{R}^m/\mathbb{Z}^m$, where $m = O(1)$ and the metric on $Z(G)$ is comparable to the metric on \mathbb{R}^m up to factors of $O(\delta^{-O(1)})$; the identification of $\log Z(G)$ with \mathbb{R}^m induces a logarithm map $\log: Z(G) \rightarrow \mathbb{R}^m$ and an exponential map $\exp: \mathbb{R}^m \rightarrow Z(G)$. Central frequencies ξ can then be identified with elements k_{ξ} of \mathbb{Z}^m , with $\xi(z) = k_{\xi} \cdot \log(z)$ for any $z \in Z(G)$.

Let $\varphi: \mathbb{R}^m \rightarrow \mathbb{R}$ be a fixed bump function (depending only on m) that equals 1 at the origin, and let $R > 1$ be a parameter to be chosen later. For any central frequency ξ , we set

$$F_{\xi}(x) := \varphi(k_{\xi}/R) \int_{\mathbb{R}^m/\mathbb{Z}^m} F(zx) e(-\xi(z)) dz$$

where dz is Haar probability measure on the torus $\mathbb{R}^m/\mathbb{Z}^m$, which acts centrally on G/Γ in the obvious fashion. It is easy to see that F_{ξ} has Lipschitz norm $O(\delta^{-O(1)})$, oscillates with central frequency ξ , and

⁵The decomposition in [19] uses the action of the vertical group G_d (which is a subgroup of the center $Z(G)$) rather than the entire center, but the arguments are otherwise nearly identical. One can think of Proposition 2.9 as a slight refinement of [19, Lemma 3.7], in that the components exhibit central oscillation rather than merely vertical oscillation.

vanishes unless ξ has Lipschitz norm $O(\delta^{-O(1)} R^{O(1)})$; also, if F has mean zero, then so do all of the F_ξ . From the Fourier inversion formula we have

$$\varphi(k_\xi/R) = \int_{\mathbb{R}^m} \hat{\varphi}(y) e(k_\xi \cdot y/R) dy = \int_{\mathbb{R}^m} \hat{\varphi}(y) e(\xi(\exp(y/R))) dy,$$

where $\hat{\varphi}(y) := \int_{\mathbb{R}^m} \varphi(\zeta) e(-\zeta \cdot y) d\zeta$, as well as the Fourier inversion formula on the torus,

$$\sum_\xi F_\xi(x) = \int_{\mathbb{R}^m} \hat{\varphi}(y) F(\exp(y/R)x) dy.$$

On the other hand, from the Lipschitz nature of F we have

$$F(\exp(y/R)x) = F(x) + O(\delta^{-O(1)} |y|/R).$$

Since $\hat{\varphi}$ is rapidly decreasing and has total integral 1, we obtain

$$F = \sum_\xi F_\xi + O(\delta^{-O(1)}/R),$$

and the claim follows by choosing $R = O(\delta^{-O(1)})$ suitably. \square

Next we shall recall a fundamental factorization theorem for polynomial sequences. Before we can state it, we need to define a few notions.

Definition 2.10 (Smoothness, total equidistribution, rationality). Let G/Γ be a filtered nilmanifold, $g \in \text{Poly}(\mathbb{Z} \rightarrow G)$ be a polynomial sequence, $I \subset \mathbb{R}$ be an interval of length $|I| \geq 1$, and $M > 0$.

(i) We say that g is (M, I) -smooth if one has

$$d_G(g(n), 1_G) \leq M; \quad d_G(g(n), g(n-1)) \leq M/|I|$$

for all $n \in I$.

(ii) We say that g is *totally $1/M$ -equidistributed* in G/Γ on I if one has

$$\left| \frac{1}{|P|} \sum_{n \in P} F(g(n)\Gamma) - \int_{G/\Gamma} F \right| \leq \frac{1}{M} \|F\|_{\text{Lip}}$$

whenever $F: G/\Gamma \rightarrow \mathbb{C}$ is Lipschitz and P is an arithmetic progression in I of cardinality at least $|I|/M$.

(iii) We say that g is M -rational if there exists $1 \leq r \leq M$ such that $g(n)^r \in \Gamma$ for all $n \in \mathbb{Z}$.

From Taylor expansion and the Baker–Campbell–Hausdorff formula it is not difficult to see that if G/Γ has degree at most d and g is M -rational, then the map $n \mapsto g(n)\Gamma$ is q -periodic for some period $1 \leq q \ll_d M^{O_d(1)}$.

Lemma 2.11. *Let $d, D \geq 1$ and $0 < \delta < 1$. Let G/Γ be a filtered nilmanifold of degree at most d , dimension D , and complexity at most $1/\delta$. Let $g \in \text{Poly}(\mathbb{Z} \rightarrow G)$, and let I be an interval with $|I| \geq 1$. Suppose that*

$$\|\eta \circ g\|_{C^\infty(I)} \leq 1/\delta \tag{2.12}$$

for some non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm at most $1/\delta$. Then there is a decomposition $g = \varepsilon g' \gamma$ into polynomial maps $\varepsilon, g', \gamma \in \text{Poly}(\mathbb{Z} \rightarrow G)$ such that

(i) ε is $(\delta^{-O_{d,D}(1)}, I)$ -smooth;

- (ii) g' takes values in $G' = \ker \eta$;
- (iii) γ is $\delta^{-O_{d,D}(1)}$ -rational.

Proof. This is a slight variant of [19, Lemma 7.9], the main difference being that our hypothesis (2.12) involves $\eta \circ g$ rather than $\eta \circ g_2$ (where g_2 is the nonlinear part of g). The argument in the proof of [19, Lemma 7.9] can be modified in an obvious manner as follows. By translation we may assume that $I = [1, |I|]$. Let $\psi : G \rightarrow \mathbb{R}^D$ be the Mal'cev coordinate map. Suppose that

$$\psi(g(n)) = t_0 + \binom{n}{1}t_1 + \binom{n}{2}t_2 + \cdots + \binom{n}{d}t_d$$

for some $t_0, t_1, \dots, t_d \in \mathbb{R}^D$ with $\psi^{-1}(t_i) \in G_i$. Our assumption on $\|\eta \circ g\|_{C^\infty(I)}$ implies that for some $k \in \mathbb{Z}^D$ with $|k| \leq \delta^{-1}$, we have

$$\|k \cdot t_i\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O_{d,D}(1)} / |I|$$

for each $1 \leq i \leq d$. Choose $u_i \in \mathbb{R}^D$ with $\psi^{-1}(u_i) \in G_i$, such that

$$k \cdot u_i \in \mathbb{Z}, \quad |t_i - u_i| \ll \delta^{-O_{d,D}(1)} / |I|.$$

Then choose $v_i \in \mathbb{R}^D$ with $\psi^{-1}(v_i) \in G_i$, all of whose coordinates are rationals over some denominator $\ll \delta^{-O_{d,D}(1)}$, such that

$$k \cdot u_i = k \cdot v_i$$

for each $1 \leq i \leq d$. Define ε, γ by

$$\psi(\varepsilon(n)) = t_0 + \sum_{i=1}^d \binom{n}{i} (t_i - u_i), \quad \psi(\gamma(n)) = \sum_{i=1}^d \binom{n}{i} v_i,$$

and then define g' by

$$g'(n) = \varepsilon(n)^{-1} g(n) \gamma(n)^{-1}.$$

One can verify that they satisfy the desired properties. \square

Theorem 2.12 (Factorization theorem). *Let $d, D \geq 1$ and $0 < \delta < 1$. Let G/Γ be a filtered nilmanifold of degree at most d , dimension D , and complexity at most $1/\delta$. Let $g \in \text{Poly}(\mathbb{Z} \rightarrow G)$ and $A > 0$, and let I be an interval with $|I| \geq 1$. Then there exists an integer $1/\delta \leq M \ll_{A,D,d} \delta^{-O_{A,D,d}(1)}$ and a decomposition $g = \varepsilon g' \gamma$ into polynomial maps $\varepsilon, g', \gamma \in \text{Poly}(\mathbb{Z} \rightarrow G)$ such that*

- (i) ε is (M, I) -smooth;
- (ii) There is an M -rational subnilmanifold G'/Γ' of G/Γ such that g' takes values in G' and is totally $1/M^A$ -equidistributed on I in G'/Γ' , and more generally in G'/Γ'' whenever Γ'' is a subgroup of Γ' of index at most M^A ; and
- (iii) γ is M -rational.

Proof. See [19, Theorem 1.19] (after rounding I to integer endpoints and translating to be of the form $[1, N]$). The additional requirement in (ii) that one has equidistribution in the larger nilmanifolds G'/Γ'' is not stated in [19, Theorem 1.19] but follows easily from the proof, the point being that if a sequence $g' \in \text{Poly}(\mathbb{Z} \rightarrow G')$ fails to be totally $1/M^A$ -equidistributed in G'/Γ'' , then one has $\|\eta \circ g'\|_{C^\infty(I)} \ll_{d,D} M^{O_{d,D}(A)}$ for some non-trivial horizontal character η on G'/Γ'' of Lipschitz norm $O_{d,D}(M^{O_{d,D}(A)})$, which on multiplying η by the index of Γ'' in Γ' also gives $\|\eta' \circ g'\|_{C^\infty(I)} \ll_{d,D} M^{O_{d,D}(A)}$ for some non-trivial horizontal character η' on G'/Γ' of Lipschitz norm $O_{d,D}(M^{O_{d,D}(A)})$. As a consequence, one can replace all occurrences of G'/Γ' in the proof of [19, Theorem 1.19] with G'/Γ'' with only negligible changes to the arguments. \square

We will also need a multidimensional version of this theorem.

Theorem 2.13 (Multidimensional factorization theorem). *Let $t, d, D \geq 1$ and $0 < \delta < 1$. Let G/Γ be a filtered nilmanifold of degree at most d , dimension D , and complexity at most $1/\delta$. Let $g \in \text{Poly}(\mathbb{Z}^t \rightarrow G)$ and $A > 0$, and let I_1, \dots, I_t intervals with $|I_1|, \dots, |I_t| \geq C\delta^{-C}$, for some C that is sufficiently large depending on t, d, D, A . Then there exists an integer $1/\delta \leq M \ll_{A,D,d,t} \delta^{-O_{A,D,d,t}(1)}$ and a decomposition $g = \varepsilon g' \gamma$ into polynomial maps $\varepsilon, g', \gamma \in \text{Poly}(\mathbb{Z}^t \rightarrow G)$ such that*

- (i) *ε is $(M, I_1 \times \dots \times I_t)$ -smooth, in the sense that $d_G(\varepsilon(n), 1_G) \leq M$ and $d_G(\varepsilon(n+e_i), 1_G) \leq M/|I_i|$ for all $n \in I_1 \times \dots \times I_t$ and $i = 1, \dots, t$, where e_1, \dots, e_t are the standard basis of \mathbb{Z}^d ;*
- (ii) *There is an M -rational subnilmanifold G'/Γ' of G/Γ such that g' takes values in G' and is totally $1/M^A$ -equidistributed in G'/Γ' , and more generally in G'/Γ'' whenever Γ'' is a subgroup of Γ' of index at most M^A , in the sense that*

$$\left| \frac{1}{|P_1 \times \dots \times P_t|} \sum_{n \in P_1 \times \dots \times P_t} F(g'(n)\Gamma) - \int_{G'/\Gamma''} F \right| \leq \frac{1}{M} \|F\|_{\text{Lip}}$$

whenever $F: G/\Gamma \rightarrow \mathbb{C}$ is Lipschitz and for each $i = 1, \dots, t$, P_i is an arithmetic progression in I_i of cardinality at least $|I_i|/M$; and

- (iii) *γ is M -rational, in the sense that there exists $1 \leq r \leq M$ such that $g(n)^r \in \Gamma$ for all $n \in \mathbb{Z}^t$.*

Proof. This follows from [19, Theorem 10.2], after implementing the corrections in [20], and the modifications indicated in the proof of Theorem 2.12. \square

As a first application of Theorem 2.12, we can obtain a criterion for correlation between nilsequences with a non-trivial central frequency:

Proposition 2.14 (Correlation criterion). *Let $d, D \geq 1$ and $0 < \delta < 1$. Let G/Γ be a filtered nilmanifold of degree at most d , dimension D , and complexity at most $1/\delta$, whose center $Z(G)$ is one-dimensional. Let $g_1, g_2 \in \text{Poly}(\mathbb{Z} \rightarrow G)$, let I be an interval with $|I| \geq 1$, and let $F: G/\Gamma \rightarrow \mathbb{C}$ be Lipschitz of norm at most $1/\delta$ and having a non-zero central frequency ξ . Suppose that one has the correlation*

$$\left| \sum_{n \in I} F(g_1(n)\Gamma) \overline{F}(g_2(n)\Gamma) \right|^* \geq \delta |I|.$$

Then at least one of the following holds:

- (i) *There exists a non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm $O_{d,D}(\delta^{-O_{d,D}(1)})$ such that $\|\eta \circ g_i\|_{C^\infty(I)} \ll_{d,D} \delta^{-O_{d,D}(1)}$ for some $i \in \{1, 2\}$.*
- (ii) *There exists a factorization*

$$g_2 = \varepsilon(\phi \circ g_1)\gamma$$

where ε is $(O_{d,D}(\delta^{-O_{d,D}(1)}), I)$ -smooth, $\phi: G \rightarrow G$ is a Lie group automorphism whose associated Lie algebra isomorphism $\log \phi: \log G \rightarrow \log G$ has matrix coefficients that are all rational of height $O_{d,D}(\delta^{-O_{d,D}(1)})$ in the Mal'cev basis X_1, \dots, X_D of $\log G$, and γ is $O_{d,D}(\delta^{-O_{d,D}(1)})$ -rational.

Proof. We allow all implied constants to depend on d, D . The product of the filtered nilmanifold G/Γ with itself is again a filtered nilmanifold $(G \times G)/(\Gamma \times \Gamma)$, with the obvious filtration $(G \times G)_i := G_i \times G_i$ and Mal'cev basis $(X_i, 0), (0, X_i)$, $i = 1, \dots, D$. This product filtered nilmanifold has degree at most d , dimension $2D$, and complexity at most $O(\delta^{-O(1)})$. The pair (g_1, g_2) can be then viewed as an element of $\text{Poly}(\mathbb{Z} \rightarrow G \times G)$. If we let $F \otimes \overline{F}: (G \times G)/(\Gamma \times \Gamma) \rightarrow \mathbb{C}$ be the function

$$F \otimes \overline{F}(x_1, x_2) := F(x_1)\overline{F}(x_2)$$

then F is Lipschitz with norm $O(\delta^{-O(1)})$ and one has

$$\left| \sum_{n \in I} F \otimes \overline{F}((g_1, g_2)(n)(\Gamma \times \Gamma)) \right|^* \geq \delta |I|. \quad (2.13)$$

Let $A > 1$ be sufficiently large depending on d, D . Applying Theorem 2.12 to (g_1, g_2) (with δ replaced by δ^A) we can find $\delta^{-A} \leq M \ll_A \delta^{-O_A(1)}$ and a factorization

$$(g_1, g_2) = (\varepsilon_1, \varepsilon_2)(g'_1, g'_2)(\gamma_1, \gamma_2) \quad (2.14)$$

where $\varepsilon_1, g'_1, \gamma_1 \in \text{Poly}(\mathbb{Z} \rightarrow G_1)$, $\varepsilon_2, g'_2, \gamma_2 \in \text{Poly}(\mathbb{Z} \rightarrow G_2)$ such that

- (i) $(\varepsilon_1, \varepsilon_2)$ is (M, I) -smooth;
- (ii) There is an M -rational subnilmanifold G'/Γ' of $(G \times G)/(\Gamma \times \Gamma)$ such that (g'_1, g'_2) takes values in G' and is totally $1/M^A$ -equidistributed in G'/Γ'' for any subgroup Γ'' of Γ' of index at most M^A ; and
- (iii) (γ_1, γ_2) is M -rational.

We caution that G' is a subgroup of $G \times G$ rather than G . From (2.13) we thus have

$$\left| \sum_{n \in I} F \otimes \overline{F}((\varepsilon_1, \varepsilon_2)(n)(g'_1, g'_2)(n)(\gamma_1, \gamma_2)(n)(\Gamma \times \Gamma)) \right|^* \geq \delta |I|.$$

Since (γ_1, γ_2) is M -rational, it is $O(M^{O(1)})$ -periodic, and then by the pigeonhole principle (and Lemma 2.2(i)) we can thus find M -rational $(\gamma_1^0, \gamma_2^0) \in G \times G$ such that

$$\left| \sum_{n \in I} F \otimes \overline{F}((\varepsilon_1, \varepsilon_2)(n)(g'_1, g'_2)(n)(\gamma_1^0, \gamma_2^0)(\Gamma \times \Gamma)) \right|^* \gg M^{-O(1)} |I|.$$

By shifting γ_1^0, γ_2^0 by elements of Γ if necessary we may assume that they lie at distance $O(M^{O(1)})$ from the identity. If we partition I into subintervals J of length $\asymp M^{-C} |I|$ for some large constant C , we see from the pigeonhole principle (and Lemma 2.2(i)) that we can find one such J for which

$$\left| \sum_{n \in J} F \otimes \overline{F}((\varepsilon_1, \varepsilon_2)(n)(g'_1, g'_2)(n)(\gamma_1^0, \gamma_2^0)(\Gamma \times \Gamma)) \right|^* \gg M^{-O(1)} |J|.$$

As $(\varepsilon_1, \varepsilon_2)$ is (M, I) -smooth, it fluctuates by $O(M^{1-C})$ on J and stays a distance $O(M)$ from the identity, hence by the Lipschitz nature of $F \otimes \overline{F}$ we conclude (for $C = O(1)$ large enough) that there exists $(\varepsilon_1^0, \varepsilon_2^0) \in G \times G$ at distance $O(M)$ from the identity such that

$$\left| \sum_{n \in J} F \otimes \overline{F}((\varepsilon_1^0, \varepsilon_2^0)(g'_1, g'_2)(n)(\gamma_1^0, \gamma_2^0)(\Gamma \times \Gamma)) \right|^* \gg M^{-O(1)} |J|.$$

Allowing implied constants to depend on C , we conclude that

$$\left| \sum_{n \in I} F \otimes \overline{F}((\varepsilon_1^0, \varepsilon_2^0)(g'_1, g'_2)(n)(\gamma_1^0, \gamma_2^0)(\Gamma \times \Gamma)) \right|^* \gg M^{-O(1)} |I|.$$

From the Baker–Campbell–Hausdorff formula and the M -rationality of (γ_1^0, γ_2^0) , we see that $(\gamma_1^0, \gamma_2^0)(\Gamma \times \Gamma)(\gamma_1^0, \gamma_2^0)^{-1}$ can be covered by $O(M^{O(1)})$ cosets of $\Gamma \times \Gamma$, and conversely. Thus if we set

$$\Gamma'' := G' \cap (\Gamma \times \Gamma) \cap (\gamma_1^0, \gamma_2^0)(\Gamma \times \Gamma)(\gamma_1^0, \gamma_2^0)^{-1}$$

then $G' \cap (\Gamma \times \Gamma)$ can be covered by $O(M^{O(1)})$ cosets of Γ'' , thus Γ'' is a subgroup of $G' \cap (\Gamma \times \Gamma)$ of index $O(M^{O(1)})$ such that

$$\Gamma''(\gamma_1^0, \gamma_2^0) \subset (\gamma_1^0, \gamma_2^0)(\Gamma \times \Gamma). \quad (2.15)$$

Indeed, one can take Γ'' to be the intersection of $G' \cap (\Gamma \times \Gamma)$ and $(\gamma_1^0, \gamma_2^0)(\Gamma \times \Gamma)(\gamma_1^0, \gamma_2^0)^{-1}$. One can then write the above claim as

$$\left| \sum_{n \in I} F'((g'_1, g'_2)(n)\Gamma'') \right|^* \gg M^{-O(1)} |I|$$

where $F' \colon G'/\Gamma'' \rightarrow \mathbb{C}$ is defined by

$$F'((g'_1, g'_2)\Gamma'') := F(\varepsilon_1^0 g'_1 \gamma_1^0 \Gamma) \overline{F}(\varepsilon_2^0 g'_2 \gamma_2^0 \Gamma)$$

for any $(g'_1, g'_2) \in G'$, with the inclusion (2.15) ensuring that this function is well-defined. Since F is Lipschitz with norm $1/\delta \leq M$, and $\varepsilon_1^0, \gamma_1^0, \varepsilon_2^0, \gamma_2^0$ are at distance $O(M^{O(1)})$ from the identity, this function is Lipschitz with norm $O(M^{O(1)})$, hence by total equidistribution of (g'_1, g'_2) we conclude (for A large enough) that

$$\left| \int_{G'/\Gamma''} F' \right| \gg M^{-O(1)}. \quad (2.16)$$

Suppose that the slice $H := \{g \in G : (g, 1) \in G'\}$ is non-trivial. This is a non-trivial closed connected subgroup of G ; by considering the final non-trivial element of the series $H, [H, G], [[H, G], G], \dots$, we conclude that H contains a non-trivial closed connected *central* subgroup of G . Since $Z(G)$ is one-dimensional, we conclude that H contains $Z(G)$. In particular, G' contains $Z(G) \times \{1\}$.

Since F has central frequency ξ , we see that

$$F'((z, 1)(g_1, g_2)) = e(\xi \cdot z) F'(g_1, g_2)$$

for all $z \in Z(G)$. By invariance of Haar measure, this implies that

$$\int_{G'/\Gamma''} F' = e(\xi \cdot z) \int_{G'/\Gamma''} F'.$$

Since ξ is non-trivial, this implies that $\int_{G'/\Gamma''} F' = 0$, contradicting (2.16). Thus the slice $\{g \in G : (g, 1) \in G'\}$ is trivial. Similarly the slice $\{g \in G : (1, g) \in G'\}$ is trivial.

Now suppose that the projection $K := \{g_1 \in G : (g_1, g_2) \in G' \text{ for some } g_2 \in G\}$ is not all of G . This is a proper closed connected subgroup of G with

$$\log K = \{X \in \log G : (X, Y) \in \log G' \text{ for some } Y \in \log G\};$$

thus $\log K$ is the projection of $\log G'$ to $\log G$. Since $\log G'$ is $M^{O(1)}$ -rational, $\log K$ is also. Hence there exists a non-trivial horizontal character $\eta \colon G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm $O(M^{O(1)})$ that annihilates K , so in particular $\eta(g'_1(n)) = 0$ for all n . From (2.14) we then have

$$\eta(g_1(n)) = \eta(\varepsilon_1(n)) + \eta(\gamma_1(n)).$$

Since γ_1 is M -rational, $M\eta(\gamma_1(n)) = 0$. Thus if we replace η by $M\eta$ we have

$$\eta(g_1(n)) = \eta(\varepsilon_1(n)).$$

Since $(\varepsilon_1, \varepsilon_2)$ is (M, I) smooth we thus conclude that

$$\|\eta \circ g_1\|_{C^\infty(I)} \ll M^{O(1)}$$

and we are in conclusion (i) of the proposition. Thus we may assume that the projection $\{g_1 \in G : (g_1, g_2) \in G'\}$ for some $g_2 \in G$ is all of G . Similarly we may assume that $\{g_2 \in G : (g_1, g_2) \in G'\}$ for some $g_1 \in G$ is all of G .

Applying Goursat's lemma, we now conclude that G' takes the form

$$G' = \{(g_1, \phi(g_1)) : g_1 \in G\}$$

for some group automorphism $\phi: G \rightarrow G$. Since G' is a $O(M^{O(1)})$ -rational subgroup of $G \times G$, ϕ must be a Lie group automorphism whose associated Lie algebra automorphism $\log \phi: \log G \rightarrow \log G$ has coefficients that are rational of height $O(M^{O(1)})$ in the Mal'cev basis. Since $(g'_1(n), g'_2(n))$ takes values in G' , we have

$$g'_2(n) = \phi(g'_1(n))$$

and hence by (2.14) and some rearranging

$$g_2(n) = \varepsilon_2(n) \phi(\varepsilon_1(n))^{-1} \phi(g_1(n)) \phi(\gamma_1(n))^{-1} \gamma_2(n).$$

It is then routine to verify that conclusion (ii) of the proposition holds. \square

As a consequence of this criterion, we can establish the following large sieve inequality for nilsequences, which is a more quantitative variant of the one in [48, Proposition 4.11].

Proposition 2.15 (Large sieve). *Let $d, D \geq 1$ and $0 < \delta < 1$. Let G/Γ be a filtered nilmanifold of degree at most d , dimension D , and complexity at most $1/\delta$, whose center $Z(G)$ is one-dimensional. Let $g_1, \dots, g_K \in \text{Poly}(\mathbb{Z} \rightarrow G)$, let I be an interval with $|I| \geq 1$, and let $F: G/\Gamma \rightarrow \mathbb{C}$ be Lipschitz of norm at most $1/\delta$ and having a non-zero central frequency ξ . Suppose that there is a function $f: \mathbb{Z} \rightarrow \mathbb{C}$ with $\sum_{n \in I} |f(n)|^2 \leq \frac{1}{\delta} |I|$ such that*

$$\left| \sum_{n \in I} f(n) \overline{F}(g_i(n)\Gamma) \right|^* \geq \delta |I| \quad (2.17)$$

for all $i = 1, \dots, K$. Then at least one of the following holds:

- (i) *There exists a non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm $O_{d,D}(\delta^{-O_{d,D}(1)})$ such that $\|\eta \circ g_i\|_{C^\infty(I)} \ll_{d,D} \delta^{-O_{d,D}(1)}$ for $\gg_{d,D} \delta^{O_{d,D}(1)} K$ values of $i = 1, \dots, K$.*
- (ii) *For $\gg_{d,D} \delta^{O_{d,D}(1)} K^2$ pairs $(i, j) \in \{1, \dots, K\}^2$, there exists a factorization*

$$g_i = \varepsilon_{ij} g_j \gamma_{ij}$$

where ε_{ij} is $(O_{d,D}(\delta^{-O_{d,D}(1)}), I)$ -smooth and γ_{ij} is $O_{d,D}(\delta^{-O_{d,D}(1)})$ -rational.

Proof. We allow implied constants to depend on d, D . From (2.17) one can find progressions $P_i \subset I$ for $i = 1, \dots, K$ such that

$$\left| \sum_{n \in I} f(n) 1_{P_i}(n) \overline{F}(g_i(n)\Gamma) \right| \geq \delta |I|$$

and thus

$$\left| \sum_{i=1}^K \theta_i \sum_{n \in I} f(n) 1_{P_i}(n) \overline{F}(g_i(n) \Gamma) \right| \geq \delta K |I|$$

for some complex numbers θ_i with $|\theta_i| \leq 1$. By interchanging the sums and applying Cauchy–Schwarz, we have

$$\left| \sum_{i=1}^K \theta_i \sum_{n \in I} f(n) 1_{P_i}(n) \overline{F}(g_i(n) \Gamma) \right|^2 \leq \frac{1}{\delta} |I| \sum_{n \in I} \left| \sum_{i=1}^K \theta_i 1_{P_i}(n) \overline{F}(g_i(n) \Gamma) \right|^2$$

and thus

$$\sum_{n \in I} \left| \sum_{i=1}^K \theta_i 1_{P_i}(n) \overline{F}(g_i(n) \Gamma) \right|^2 \geq \delta^3 K^2 |I|.$$

From the triangle inequality we have

$$\sum_{n \in I} \left| \sum_{i=1}^K \theta_i 1_{P_i}(n) \overline{F}(g_i(n) \Gamma) \right|^2 \leq \sum_{1 \leq i, j \leq K} \left| \sum_{n \in I} F(g_i(n) \Gamma) \overline{F}(g_j(n) \Gamma) \right|^2$$

and thus

$$\sum_{1 \leq i, j \leq K} \left| \sum_{n \in I} F(g_i(n) \Gamma) \overline{F}(g_j(n) \Gamma) \right|^2 \geq \delta^3 K^2 |I|.$$

The inner sum is $O(\delta^{-2} |I|)$, thus we have

$$\left| \sum_{n \in I} F(g_i(n) \Gamma) \overline{F}(g_j(n) \Gamma) \right|^2 \gg \delta^{O(1)} |I|$$

for $\gg \delta^{O(1)} K^2$ pairs $(i, j) \in \{1, \dots, K\}^2$. For each such pair, we apply Proposition 2.14. If conclusion (i) of that proposition holds for $\gg \delta^{O(1)} K^2$ pairs (i, j) , then by the pigeonhole principle (noting that there are only $O(\delta^{-O(1)})$ choices for η) we obtain conclusion (i) of the current proposition. Thus we may assume that conclusion (ii) of Proposition 2.14 holds for $\gg \delta^{O(1)} K^2$ pairs $(i, j) \in \{1, \dots, K\}^2$, thus we have

$$g_i = \varepsilon_{ij} \phi_{ij}(g_j) \gamma_{ij}$$

for all such pairs (i, j) , where ε_{ij} is $(O(\delta^{-O(1)}), I)$ -smooth, γ_{ij} is $O(\delta^{-O(1)})$ -rational, and $\phi_{ij}: G \rightarrow G$ is a Lie group automorphism whose associated Lie algebra isomorphism $\log \phi: \log G \rightarrow \log G$ has matrix coefficients that are all rational of height $O(\delta^{-O(1)})$ in the Mal'cev basis X_1, \dots, X_D of $\log G$. The total number of choices for ϕ_{ij} is $O(\delta^{-O(1)})$, so by the pigeonhole principle we may assume that $\phi_{ij} = \phi$ is independent of i, j . By Cauchy–Schwarz, we may thus find $\gg \delta^{O(1)} K^3$ triples $(i, i', j) \in \{1, \dots, K\}^3$ such that

$$g_i = \varepsilon_{ij} \phi(g_j) \gamma_{ij}; \quad g_{i'} = \varepsilon_{i'j} \phi(g_j) \gamma_{i'j}$$

where $\varepsilon_{ij}, \varepsilon_{i'j}, \gamma_{ij}, \gamma_{i'j}$ are as above. This implies that

$$g_i = \varepsilon_{ij} \varepsilon_{i'j}^{-1} g_{i'} \gamma_{i'j}^{-1} \gamma_{ij}.$$

Pigeonholing in j and relabeling i, i' as i, j , we obtain conclusion (ii) of the current proposition. \square

2.4. Combinatorial lemmas

The following lemma is a standard consequence of Heath-Brown's identity.

Lemma 2.16. *Let $X \geq 2$, and let $L \in \mathbb{N}$ be fixed. We may find a collection \mathcal{F} of $(\log X)^{O(1)}$ functions $f: \mathbb{N} \rightarrow \mathbb{R}$, such that*

$$\Lambda(n) = \sum_{f \in \mathcal{F}} f(n)$$

for each $X/2 \leq n \leq 4X$, and each $f \in \mathcal{F}$ takes the form

$$f = a^{(1)} * \cdots * a^{(\ell)}$$

for some $\ell \leq 2L$, where $a^{(i)}$ is supported on $(N_i, 2N_i]$ for some $N_i \geq 1/2$, and each $a^{(i)}(n)$ is either $1_{(N_i, 2N_i]}(n)$, $(\log n)1_{(N_i, 2N_i]}(n)$, or $\mu(n)1_{(N_i, 2N_i]}$. Moreover, $N_1 N_2 \cdots N_\ell \asymp X$, and $N_i \ll X^{1/L}$ for each i with $a^{(i)}(n) = \mu(n)1_{(N_i, 2N_i]}(n)$. The same statement holds for μ in place of Λ (but $(\log n)1_{(N_i, 2N_i]}(n)$ does not appear).

Proof. Using Heath-Brown's identity (see [37, (13.37), (13.38)] with $K = L$ and $z = (2X)^{1/L}$), we have

$$\Lambda(n) = \sum_{1 \leq j \leq L} (-1)^{j-1} \binom{L}{j} \sum_{m_1, \dots, m_j \leq (2X)^{1/L}} \mu(m_1) \cdots \mu(m_j) \sum_{m_1 \cdots m_j n_1 \cdots n_j = n} \log n_1$$

and

$$\mu(n) = \sum_{1 \leq j \leq L} (-1)^{j-1} \binom{L}{j} \sum_{m_1, \dots, m_j \leq (2X)^{1/L}} \mu(m_1) \cdots \mu(m_j) \sum_{m_1 \cdots m_j n_1 \cdots n_{j-1} = n} 1.$$

The conclusion follows after dyadic division of the ranges of variables. \square

The following Shiu's bound [60, Theorem 1] will be used multiple times to control sums of divisor functions in short intervals in arithmetic progressions.

Lemma 2.17. *Let $A \geq 1$ and $\varepsilon > 0$ be fixed. Let $X \geq H \geq X^\varepsilon$ and $1 \leq q \leq H^{1-\varepsilon}$. Let f be a non-negative multiplicative function such that $f(p^\ell) \leq A^\ell$ for every prime power p^ℓ and $f(n) \ll_c n^c$ for every $c > 0$. Then, for any integer a coprime to q , we have*

$$\sum_{\substack{X < n \leq X+H \\ n \equiv a \pmod{q}}} f(n) \ll \frac{H}{\varphi(q) \log X} \exp \left(\sum_{\substack{p \leq 2X \\ p \nmid q}} \frac{f(p)}{p} \right).$$

For proving Theorem 1.1(iv)–(v), we need a more flexible combinatorial decomposition of the multiplicative functions μ, d_k , where we introduce an extra variable $p \in (P, Q]$ in the factorization. Before stating this, let us quickly prove a lemma that will in particular allow us to write, for $P < Q \leq X^{1/(\log \log X)^2}$,

$$1_{(n, \prod_{P < p \leq Q} p) = 1} = \sum_{\substack{d \mid (n, \prod_{P < p \leq Q} p) \\ d \leq X^\varepsilon}} \mu(d) + \text{acceptable error}$$

in our sums. This can be seen as a simple version of the fundamental lemma of the sieve that is sufficient to our needs.

Lemma 2.18. *Let $k, r \geq 1$ and $\varepsilon > 0$ be fixed. Let $X \geq H \geq X^\varepsilon$ and $X \geq D \geq Q > P \geq 2$. Then, for any $C \geq 1$,*

$$\sum_{\substack{X < mn \leq X+H \\ p|m \implies p \in (P, Q] \\ m > D}} d_k(mn)^r \ll_C H \frac{(\log X)^{2k^r e^C}}{\exp(C \frac{\log D}{\log Q})}. \quad (2.18)$$

Proof. Write $\ell = mn$ and note that since $m > D$, we have $\Omega(\ell) \geq \frac{\log D}{\log Q}$. Hence the left hand side of (2.18) is

$$\leq \sum_{\substack{X < \ell \leq X+H \\ \Omega(\ell) \geq \frac{\log D}{\log Q}}} d_2(\ell) d_k(\ell)^r \leq e^{-C \frac{\log D}{\log Q}} \sum_{X < \ell \leq X+H} e^{C\Omega(\ell)} d_2(\ell) d_k(\ell)^r \ll_C H \frac{(\log X)^{2k^r e^C}}{\exp(C \frac{\log D}{\log Q})}$$

by Lemma 2.17. \square

Now we state the lemma allowing us to introduce an extra variable $p \in (P, Q]$ in the factorization. It is a slight variant of [50, Lemma 3.1] (see also [50, Remark 3.2]).

Lemma 2.19. *Let $\varepsilon > 0$ and $k \geq 1$ be fixed. Let $X \geq 3$, $X^\varepsilon \leq H \leq X$, and let $2 \leq P < Q \leq X^{1/(\log \log X)^2}$. Write $\mathcal{P}(P, Q) = \prod_{P < p \leq Q} p$. Let f be any multiplicative function satisfying $|f(n)| \leq d_k(n)$. Then for any sequence $\{\omega_n\}$ with $|\omega_n| \leq 1$, we have*

$$\sum_{\substack{X < n \leq X+H \\ (n, \mathcal{P}(P, Q)) > 1}} f(n) \omega_n = \sum_{\substack{X < prn \leq X+H \\ P < p \leq Q \\ r \leq X^{\varepsilon/2}}} a_r f(p) f(n) \omega_{prn} + O\left(\frac{H(\log X)^{4k}}{P} + \frac{H}{\exp((\log \log X)^2)}\right),$$

where $\{a_r\}$ is an explicit sequence satisfying $|a_r| \leq d_{k+1}(r)$.

Proof. This is very similar to [50, Remark 3.2] but for completeness we provide the proof in a somewhat simpler form.

By Ramaré's identity

$$f(n) \omega_n 1_{(n, \mathcal{P}(P, Q)) > 1} = \sum_{P < p \leq Q} \sum_{pm=n} \frac{f(pm) \omega_{pm}}{\omega_{(P, Q]}(pm)} \quad (2.19)$$

where $\omega_{(P, Q]}(m)$ is the number of distinct prime divisors of m on $(P, Q]$; this identity follows directly since the number of representations $n = pm$ with $P < p \leq Q$ is $\omega_{(P, Q]}(n)$.

We write m uniquely as $m = m_1 m_2$ with m_1 having all of its prime factors from $(P, Q]$ and m_2 having no prime factors from that interval. Summing over n and then spotting the condition $(m_2, \mathcal{P}(P, Q)) = 1$ using Möbius inversion, we see that

$$\begin{aligned} \sum_{\substack{X < n \leq X+H \\ (n, \mathcal{P}(P, Q)) > 1}} f(n) \omega_n &= \sum_{P < p \leq Q} \sum_{\substack{X/p \leq m_1 m_2 \leq (X+H)/p \\ p' | m_1 \implies p' \in (P, Q] \\ (m_2, \mathcal{P}(P, Q)) = 1}} \frac{f(pm_1 m_2)}{\omega_{(P, Q]}(pm_1)} \omega_{m_1 m_2 p} \\ &= \sum_{P < p \leq Q} \sum_{\substack{X/p \leq m_1 d m_2 \leq (X+H)/p \\ d | \mathcal{P}(P, Q) \\ p' | m_1 \implies p' \in (P, Q]}} \frac{\mu(d) f(pm_1 d m_2)}{\omega_{(P, Q]}(pm_1)} \omega_{m_1 d m_2 p}. \end{aligned} \quad (2.20)$$

Let us show that we can restrict the summation to $dm_1 \leq X^{\varepsilon/2}$. Writing $m = dm_1$ and $n = pm_2$, we see that by Lemma 2.18 with $C = 4/\varepsilon$ the contribution of $dm_1 > X^{\varepsilon/2}$ is bounded by

$$\leq \sum_{\substack{X < mn \leq X+H \\ p|m \implies p \in (P, Q] \\ m > X^{\varepsilon/2}}} d_2(m)d_2(n)d_k(mn) \leq \sum_{\substack{X < mn \leq X+H \\ p|m \implies p \in (P, Q] \\ m > X^{\varepsilon/2}}} d_{2k}(mn)^3 \ll \frac{H}{\exp((\log \log X)^2)}.$$

Furthermore, since in (2.20) all prime factors of pdm_1 are from $(P, Q]$, we have

$$f(pdm_1) = f(p)f(dm_1)f(m_2) \quad \text{and} \quad \omega_{(P, Q]}(pm_1) = \omega_{(P, Q]}(m_1) + 1 \quad (2.21)$$

unless there exists a prime $q \in (P, Q]$ such that $q^2 \mid pm_1dm_2 =: \ell$. Applying Lemma 2.17, the error introduced by making the changes (2.21) to (2.20) is

$$\ll \sum_{P < q \leq Q} \sum_{\substack{X < \ell \leq X+H \\ q^2 \mid \ell}} d_4(\ell)d_k(\ell) \ll \sum_{P < q \leq Q} \frac{H}{q^2} (\log X)^{4k-1} \ll \frac{H}{P} (\log X)^{4k-1}.$$

Thus (2.20) equals

$$\sum_{\substack{X \leq pm_1dm_2 \leq X+H \\ p' \mid dm_1 \implies p' \in (P, Q] \\ P < p \leq Q, dm_1 \leq X^{\varepsilon/2}}} \frac{\mu(d)f(p)f(dm_1)f(m_2)}{\omega_{(P, Q]}(m_1) + 1} \omega_{m_1dm_2p} + O\left(\frac{H}{\exp((\log \log X)^2)} + \frac{H}{P} (\log X)^{4k-1}\right),$$

and the claim follows with

$$a_r := f(r)1_{p|r \implies p \in (P, Q]} \sum_{r=dm_1} \frac{\mu(d)}{\omega_{(P, Q]}(m_1) + 1},$$

□

The following combinatorial lemma will be used to arrange each component arising from Lemma 2.16 into a desired form, such as a type I sum, a type II sum, or a type I_2 sum.

Lemma 2.20. *Let $\alpha_1, \dots, \alpha_k$ be nonnegative real numbers with $\sum_{i=1}^k \alpha_i = 1$ and let $\frac{1}{3} \leq \theta \leq 1$. For any $I \subset \{1, \dots, k\}$, write $\alpha_I := \sum_{i \in I} \alpha_i$. Consider the following statements:*

- (I) *One has $\alpha_i \geq 1 - \theta$ for some $1 \leq i \leq k$.*
- (I_2^{maj}) *One has $\alpha_{\{i, j\}} \geq 1 - \theta$ for some $1 \leq i < j \leq k$.*
- (I_2) *One has $\alpha_{\{i, j\}} \geq \frac{3}{2}(1 - \theta)$ for some $1 \leq i < j \leq k$.*
- (II^{maj}) *There exists a partition $\{1, \dots, k\} = I \uplus J \uplus J'$ such that $2\theta - 1 \leq \alpha_I \leq 4\theta - 2$ and $|\alpha_J - \alpha_{J'}| \leq 2\theta - 1$.*
- (II^{min}) *There exists a partition $\{1, \dots, k\} = J \uplus J'$ such that $|\alpha_J - \alpha_{J'}| \leq 2\theta - 1$ (or equivalently, $\alpha_J, \alpha_{J'} \in [1 - \theta, \theta]$; or equivalently, $\alpha_J \in [1 - \theta, \theta]$).*

Then the following claims hold.

- (i) *Suppose that $\theta = 5/8$. Then at least one of (I) or (II^{maj}) holds.*
- (ii) *Suppose that $\theta \geq 3/5$. Then at least one of (I), (I_2), or (II^{min}) holds.*
- (iii) *Suppose that $\theta = 7/12$. Then at least one of (I), (I_2^{maj}), or (II^{maj}) holds.*
- (iv) *Suppose that $k = 5$ and $\theta = 11/20$. Then at least one of (I_2^{maj}) or (II^{maj}) holds.*
- (v) *Suppose that $k \in \{3, 4\}$ and $\theta \geq 1/2$. Then (I_2^{maj}) holds.*
- (vi) *Suppose that $k = 3$ and $\theta \geq 5/9$ or $k = 2$ and $\theta \geq 1/3$. Then (I_2) holds.*

Remark 2.21. The different conclusions (I), (I_2^{maj}) , (I_2) , (II^{maj}) , (II^{min}) in Lemma 2.20 correspond to different types of sums that behave well on intervals $(X, X + H]$ with H much larger than X^θ :

- Exponents obeying (I) correspond to “type I sums” which behave well for both major and minor arc correlations.
- Exponents obeying (I_2^{maj}) correspond to “type I_2 sums” which behave well for major arc correlations.
- Exponents obeying (I_2) correspond to “type I_2 sums” which behave well for both major and minor arc correlations.
- Exponents obeying (II^{maj}) correspond to “type II sums” which behave well for major arc correlations.
- Exponents obeying (II^{min}) correspond to “type II sums” which behave well for minor arc correlations, or for major arc correlations when one can extract a medium-sized prime factor from the sum.

Proof. We first handle the easy case (vi). If $k = 2$ and $\theta \geq 1/3$, then $\frac{3}{2}(1 - \theta) \leq 1$ and (I_2) follows simply by taking $\{i, j\} = \{1, 2\}$. If $k = 3$ and $\theta \geq \frac{5}{9}$, then $\frac{3}{2}(1 - \theta) \leq \frac{2}{3}$ and (I_2) follows by noting that the sum of the two largest of the reals $\alpha_1, \alpha_2, \alpha_3$ is necessarily at least $\frac{2}{3}$.

Now we prove (v). If $k = 4$ and $\theta \geq 1/2$, then by the pigeonhole principle one of $\alpha_{\{1,2\}}, \alpha_{\{3,4\}}$ is at least $\frac{1}{2} \geq 1 - \theta$, and we obtain (I_2^{maj}) in this case. The case $k = 3$ follows similarly, with some room to spare.

In a similar spirit in case (iv), when $k = 5$ and $\theta = \frac{11}{20}$, then one of the α_i must be at most $\frac{1}{5}$; without loss of generality $\alpha_5 \leq \frac{1}{5}$. Since $1 - \theta = \frac{9}{20}$, we obtain (I_2^{maj}) except when $\alpha_{\{1,2\}}, \alpha_{\{3,4\}} \leq \frac{9}{20}$, which by $\sum_{i=1}^5 \alpha_i = 1$ forces $\alpha_{\{3,4\}}, \alpha_{\{1,2\}} \geq 1 - \frac{9}{20} - \frac{1}{5} = \frac{7}{20}$. Thus $|\alpha_{\{1,2\}} - \alpha_{\{3,4\}}| \leq \frac{9}{20} - \frac{7}{20} = \frac{1}{10} = 2\theta - 1$. Also we have

$$\alpha_5 = 1 - \alpha_{\{1,2\}} - \alpha_{\{3,4\}} \geq 1 - \frac{9}{20} - \frac{9}{20} = \frac{1}{10} = 2\theta - 1$$

and

$$\alpha_5 \leq \frac{1}{5} = 4\theta - 2$$

and so we obtain (II^{maj}) in this case. This establishes (iv).

In the remaining cases (i)–(iii) we assume, without loss of generality, that

$$\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_k.$$

In case (ii) when $\theta \geq 3/5$ we obtain (I) unless $\alpha_j < 1 - \theta$ for each j and (I_2) unless $\alpha_{\{i,j\}} < \frac{3}{2}(1 - \theta) \leq \theta$ for any distinct i, j . But if $\alpha_{\{i,j\}} \in [1 - \theta, \theta]$ for some distinct i, j , then we have (II^{min}) . Hence we can assume that $\alpha_{i,j} < 1 - \theta$ for any distinct i, j . In particular, for any $j \neq 1$ we have

$$\alpha_j \leq \frac{\alpha_1 + \alpha_j}{2} \leq \frac{1 - \theta}{2} \leq 2\theta - 1.$$

Consequently there must be an index $r \in \{3, \dots, k\}$ such that $\alpha_1 + \sum_{j=2}^r \alpha_j \in [1 - \theta, \theta]$, and hence (II^{min}) holds.

Let us now consider (i). Now $\theta = 5/8$ and we obtain (I) unless $\alpha_j < 3/8$ for every j (and in particular we can assume that $k \geq 3$). Note that $2\theta - 1 = 1/4$ in this case. If now $\alpha_3 > 1/4$, then $\alpha_1, \alpha_2 \in [1/4, 3/8]$ and we have (II^{maj}) with $J = \{1\}, J' = \{2\}$, and $I = \{3, \dots, k\}$.

On the other hand, if $\alpha_3 \leq 1/4$, we set $J_0 = \{1\}$ and $J'_0 = \{2, \dots, r\}$ with $r \geq 2$ the greatest integer such that $\alpha_{J'_0} < \alpha_{J_0}$. Then necessarily $|\alpha_{J_0} - \alpha_{J'_0}| \leq 1/4 = 2\theta - 1$. Furthermore $\alpha_{J'_0} + \alpha_{J_0} \leq 2 \cdot \alpha_1 \leq 3/4$. If also $\alpha_{J'_0} + \alpha_{J_0} \geq 1/2$ then we have (II^{maj}) with $J = J_0, J' = J'_0$, and $I = \{1, \dots, k\} \setminus (J_0 \cup J'_0)$.

Otherwise we add indices $j \geq r + 1$ one by one to J_0 or J'_0 depending on whether $\alpha_{J_0} < \alpha_{J'_0}$ or not. We continue this process until $\alpha_{J_0} + \alpha_{J'_0} \in [1/2, 3/4]$, and we again obtain (II^{maj}) .

Let us finally turn to (iii). Now $\theta = 7/12$ and $2\theta - 1 = 1/6$. We obtain (I_2^{maj}) unless $\alpha_{\{i,j\}} < 1 - \theta = 5/12$ for any distinct i, j . In particular we can assume that $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 < 5/6 < 1$ and thus $k \geq 5$.

If $\alpha_5 > 1/6$, then $\alpha_{\{2,3\}}, \alpha_{\{1,4\}} \in [1/3, 5/12]$. Consequently $1 - \alpha_{\{1,4\}} - \alpha_{\{2,3\}} \in [1/6, 1/3]$ and we obtain (II^{maj}) with $J = \{1, 4\}, J' = \{2, 3\}$, and $I = \{1, \dots, k\} \setminus \{1, 2, 3, 4\}$.

On the other hand if $\alpha_5 \leq 2\theta - 1 = 1/6$, we can argue similarly to case (i): We set $J_0 = \{1, 2\}$ and $J'_0 = \{3, \dots, r\}$ with $r \geq 4$ the greatest integer such that $\alpha_{J'_0} \leq \alpha_{J_0}$. Then necessarily $|\alpha_{J_0} - \alpha_{J'_0}| \leq 1/6 = 2\theta - 1$. Furthermore $\alpha_{J_0} + \alpha_{J'_0} \leq 2\alpha_{1,2} \leq 5/6$. If also $\alpha_{J_0} + \alpha_{J'_0} \geq 2/3$ then we have (II^{maj}) with $J = J_0$ and $J' = J'_0$. Otherwise we add indices $j \geq r + 1$ one by one to J_0 or J'_0 depending on whether $\alpha_{J_0} < \alpha_{J'_0}$ or not. We continue this process until $\alpha_{J_0} + \alpha_{J'_0} \in [2/3, 5/6]$, and we again obtain (II^{maj}) . \square

Remark 2.22. The following counterexamples, with ε small, show that θ in the various components of Lemma 2.20 cannot be decreased (apart from the $k = 3$ case of (v)):

- $\theta = 5/8 - \varepsilon, (\alpha_1, \dots, \alpha_k) = (1/4, 1/4, 1/4, 1/4)$;
- $\theta = 3/5 - \varepsilon, (\alpha_1, \dots, \alpha_k) \in \{(2/5, 1/5, 1/5, 1/5), (1/5, 1/5, 1/5, 1/5, 1/5)\}$;
- $\theta = 7/12 - \varepsilon, (\alpha_1, \dots, \alpha_k) = (1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$;
- $\theta = 11/20 - \varepsilon, (\alpha_1, \dots, \alpha_k) = (1/5, 1/5, 1/5, 1/5, 1/5)$;
- $\theta = 1/2 - \varepsilon, (\alpha_1, \dots, \alpha_k) = (1/4, 1/4, 1/4, 1/4)$;
- $\theta = 5/9 - \varepsilon, (\alpha_1, \dots, \alpha_k) = (1/3, 1/3, 1/3)$;
- $\theta = 1/3 - \varepsilon, (\alpha_1, \dots, \alpha_k) = (\alpha, 1 - \alpha)$ for any $\alpha \in (0, 1)$.

3. Major arc estimates

In the proof of Theorem 1.1 we shall use Theorem 4.2 below to reduce to “major arc” cases where more-or-less $F(g(n)\Gamma) = 1$ (or $F(g(n)\Gamma) = n^{it}$ in case of type II sums). The purpose of this section is to establish the following estimates corresponding to the case $F(g(n)\Gamma) = 1$ as well as an auxiliary result (Lemma 3.5 below) on trilinear sums in case $F(g(n)\Gamma) = n^{it}$.

Theorem 3.1 (Major arc estimate). *Let $X \geq 3$ and $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$ for some $0 < \theta < 1$ and $\varepsilon > 0$.*

(i) *(Huxley type estimates) Set $\theta = 7/12$. Then, for all $A > 0$,*

$$\left| \sum_{X < n \leq X+H} \mu(n) \right|^* \ll_{A, \varepsilon} \frac{H}{\log^A X}$$

and

$$\left| \sum_{X < n \leq X+H} (\Lambda(n) - \Lambda^\sharp(n)) \right|^* \ll_{A, \varepsilon} \frac{H}{\log^A X}.$$

(ii) *Let $k \geq 2$. Set $\theta = 1/3$ for $k = 2$, $\theta = 1/2$ for $k = 3, 4$, $\theta = 11/20$ for $k = 5$, and $\theta = 7/12$ for $k \geq 6$. Then*

$$\left| \sum_{X < n \leq X+H} (d_k(n) - d_k^\sharp(n)) \right|^* \ll_{\varepsilon} \frac{H}{X^{c_k}} + \frac{H}{X^{\varepsilon/1000}}$$

for some constant $c_k > 0$ depending only on k .

We remark that if we replace the maximal sums $|\cdot|^*$ here by the ordinary sums $|\cdot|$, then the $\theta = 7/12$ case of Theorem 3.1 can also be extracted after some computation from the work of Ramachandra [56] (see in particular Remarks 4, 5 of that paper), with a pseudopolynomial gain

$O(\exp(-c(\log X)^{1/3}/(\log \log X)^{1/3}))$, while the cases $k = 4, 5$ of Theorem 3.1(ii) follow from [23, (4.23)] and [27]. Here we will provide the proofs from our viewpoint. It may be possible to improve the error terms in (i) to be pseudopolynomial in nature even for the maximal sums, if one adjusts the approximants $\mu^\sharp, \Lambda^\sharp$ to take into account the possibility of a Siegel zero, in the spirit of [61, Proposition 2.2].

For the $\theta = 7/12$ result, the primary obstruction arises from convolutions (1.29) with $(\alpha_1, \dots, \alpha_m)$ equal to $(1/6, 1/6, 1/6, 1/6, 1/6, 1/6)$, as this lies just outside the reach of our untwisted major arc type *I* and type *II* estimates when θ goes below $7/12$ (cf., the third item of Remark 2.22). This obstruction has long been known; see e.g., [29]. Note that this obstruction does not arise for $k < 6$, which explains the fact that better exponents than $7/12$ are available for d_2, d_3, d_4, d_5 . The corresponding obstructions can be found in the other items of Remark 2.22.

It would probably be possible to obtain Theorem 3.1(ii) for $\theta = 131/416 \approx 0.315$ when $k = 2$ and for $\theta = 43/96 \approx 0.448$ when $k = 3$ — corresponding to the progress in the Dirichlet divisor problem [34, 41] — but we do not attempt to compute this here (it requires checking that the arguments in the literature, when adapted to the Dirichlet divisor problem in an arithmetic progression, give a polynomial dependence on the common difference of the arithmetic progression, and it also does not directly improve the exponents in Theorem 1.1).

Let us now explain the strategy of the proof of Theorem 3.1. Let $f \in \{\mu, \Lambda, d_k\}$. By adjusting the implied constants, it suffices to show the claims with

$$\left| \sum_{X < n \leq X+H} (f(n) - f^\sharp(n)) \right|^* \text{ replaced by } \max_{a, q \in \mathbb{N}} \left| \sum_{\substack{X < n \leq X+H \\ n \equiv a \pmod{q}}} (f(n) - f^\sharp(n)) \right|.$$

In the cases $f = \mu, \Lambda$ we take $H' := X/\log^{20A} X$ and in the case $f = d_k$ we take $H' := X^{1-1/100k}$. We use the triangle inequality to write

$$\begin{aligned} \left| \frac{1}{H} \sum_{\substack{X < n \leq X+H \\ n \equiv a \pmod{q}}} (f(n) - f^\sharp(n)) \right| &\leq \left| \frac{1}{H} \sum_{\substack{X < n \leq X+H \\ n \equiv a \pmod{q}}} f(n) - \frac{1}{H'} \sum_{\substack{X < n \leq X+H' \\ n \equiv a \pmod{q}}} f(n) \right| \\ &+ \left| \frac{1}{H'} \sum_{\substack{X < n \leq X+H' \\ n \equiv a \pmod{q}}} (f(n) - f^\sharp(n)) \right| + \left| \frac{1}{H} \sum_{\substack{X < n \leq X+H \\ n \equiv a \pmod{q}}} f^\sharp(n) - \frac{1}{H'} \sum_{\substack{X < n \leq X+H' \\ n \equiv a \pmod{q}}} f^\sharp(n) \right|. \end{aligned} \tag{3.1}$$

Then we show that each of the three differences on the right-hand side is small. Let us next state the required results.

To attack the second difference in (3.1), we show in Section 3.1 that Theorem 3.1 holds in long intervals.

Proposition 3.2 (Long intervals). *Let $X \geq H_2 \geq 2$.*

(i) *Let $A > 0$ and⁶ $X/\log^A X \leq H_2 \leq X$. Then*

$$\max_{a, q \in \mathbb{N}} \left| \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} \mu(n) \right| \ll_A \frac{H_2}{\log^A X}. \tag{3.2}$$

⁶Actually, thanks to Lemma 2.2(i), it would suffice to consider the case $H_2 = X$ here.

and

$$\max_{a,q \in \mathbb{N}} \left| \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} (\Lambda(n) - \Lambda^\sharp(n)) \right| \ll_A \frac{H_2}{\log^A X}. \quad (3.3)$$

(ii) Let $k \geq 2$ and $X^{1-\frac{1}{50k}} \leq H_2 \leq X$. Then

$$\max_{a,q \in \mathbb{N}} \left| \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} (d_k(n) - d_k^\sharp(n)) \right| \ll \frac{H_2^2}{X} \log^{k-2} X. \quad (3.4)$$

Furthermore, using the definitions of our approximants $\Lambda^\sharp(n)$ and $d_k^\sharp(n)$ as type I sums, it will be straightforward to show that the third difference on the right of (3.1) is small; in Section 3.2 we shall show the following.

Lemma 3.3 (Long and short averages of approximant). *Let $X \geq H_2 \geq H_1 \geq X^{1/4} \geq 2$.*

(i) *One has*

$$\max_{a,q \in \mathbb{N}} \left| \frac{1}{H_1} \sum_{\substack{X < n \leq X+H_1 \\ n \equiv a \pmod{q}}} \Lambda^\sharp(n) - \frac{1}{H_2} \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} \Lambda^\sharp(n) \right| \ll \exp(-(\log X)^{1/10}). \quad (3.5)$$

(ii) *Let $k \geq 2$. Then*

$$\max_{a,q \in \mathbb{N}} \left| \frac{1}{H_1} \sum_{\substack{X < n \leq X+H_1 \\ n \equiv a \pmod{q}}} d_k^\sharp(n) - \frac{1}{H_2} \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} d_k^\sharp(n) \right| \ll \frac{1}{X^{1/100}} + \frac{H_2}{X} \log^{k-2} X. \quad (3.6)$$

Our ability to handle the first difference in (3.1) is what determines the exponent θ . Concerning the first difference we prove the following proposition in Section 3.4.

Proposition 3.4 (Long and short averages of arithmetic function).

(i) *Let $X/\log^{20A} X \geq H_2 \geq H_1 \geq X^{7/12+\varepsilon}$. Then*

$$\max_{a,q \in \mathbb{N}} \left| \frac{1}{H_1} \sum_{\substack{X < n \leq X+H_1 \\ n \equiv a \pmod{q}}} \Lambda(n) - \frac{1}{H_2} \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} \Lambda(n) \right| \ll_{A,\varepsilon} \frac{1}{\log^A X}$$

and

$$\max_{a,q \in \mathbb{N}} \left| \frac{1}{H_1} \sum_{\substack{X < n \leq X+H_1 \\ n \equiv a \pmod{q}}} \mu(n) \right| \ll_{A,\varepsilon} \frac{1}{\log^A X}.$$

(ii) Let $k \geq 2$. Set $\theta = 1/3$ for $k = 2$, $\theta = 1/2$ for $k = 3, 4$, $\theta = 11/20$ for $k = 5$, and $\theta = 7/12$ for $k \geq 6$. There exists $c_k > 0$ such that if $X^{1-1/(100k)} \geq H_2 \geq H_1 \geq X^{\theta+\varepsilon}$, then

$$\max_{a,q \in \mathbb{N}} \left| \frac{1}{H_1} \sum_{\substack{X < n \leq X+H_1 \\ n \equiv a \pmod{q}}} d_k(n) - \frac{1}{H_2} \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} d_k(n) \right| \ll_{\varepsilon, k} \frac{1}{X^{c_k}} + \frac{1}{X^{\varepsilon/1000}}$$

Theorem 3.1 now follows from (3.1) together with Propositions 3.4 and 3.2 and Lemma 3.3.

The case $k = 2$ of Proposition 3.4(ii) can be treated using classical methods on the Dirichlet divisor problem. In $k \geq 3$ cases of Proposition 3.4(ii), we write $d_k(n) = \sum_{n=m_1 \dots m_k} 1$, split m_j into dyadic intervals $m_j \sim M_j \asymp X^{\alpha_j}$ and classify resulting dyadic sums using Lemma 2.20(iii). On the other hand in case of Proposition 3.4(i) we first use Heath-Brown's identity and then Lemma 2.20(iii) to classify the resulting sums.

For trilinear sums satisfying (II^{maj}) from Lemma 2.20 we shall deduce in Section 3.3 the following consequence of the work of Baker, Harman and Pintz [4]. Part (ii) of the lemma will be used in handling certain type II sums in Section 4.

Lemma 3.5. *Let $1/2 \leq \theta < 1$ and $\varepsilon > 0$. Let also $W \leq X^{\varepsilon/200}$ and $X^{\theta+\varepsilon} \leq H_1 \leq H_2 \leq X/W^4$. Let $L, M_1, M_2 \geq 1$ be such that $M_j = X^{\alpha_j}$ and $LM_1 M_2 \asymp X$. Let a_{m_1}, b_{m_2}, v_ℓ be bounded by d_2^C for some $C \geq 1$.*

Assume that $a, q \in \mathbb{N}$, $\theta \in \{11/20, 7/12, 3/5, 5/8\}$ and that $\alpha_1, \alpha_2 > 0$ obey the bounds

$$|\alpha_1 - \alpha_2| \leq 2\theta - 1 + \frac{\varepsilon}{100} \quad \text{and} \quad 1 - \alpha_1 - \alpha_2 \leq 4\theta - 2 + \frac{\varepsilon}{100}.$$

(i) *If*

$$\max_{r|(a,q)} \max_{\chi \pmod{\frac{q}{(a,q)}}} \sup_{W \leq |t| \leq \frac{XW^4}{H_1}} \left| \sum_{\ell \sim L/r} \frac{v_\ell r \chi(\ell)}{\ell^{1/2+it}} \right| \ll_C \frac{(L/r)^{1/2}}{W^{1/3}}, \quad (3.7)$$

then

$$\left| \frac{1}{H_1} \sum_{\substack{X < m_1 m_2 \ell \leq X+H_1 \\ m_j \sim M_j, \ell \sim L \\ m_1 m_2 \ell \equiv a \pmod{q}}} a_{m_1} b_{m_2} v_\ell - \frac{1}{H_2} \sum_{\substack{X < m_1 m_2 \ell \leq X+H_2 \\ m_j \sim M_j, \ell \sim L \\ m_1 m_2 \ell \equiv a \pmod{q}}} a_{m_1} b_{m_2} v_\ell \right| \ll d_3(q) \frac{\log^{O_C(1)} X}{W^{1/3}}.$$

(ii) *If*

$$\max_{r|(a,q)} \max_{\chi \pmod{\frac{q}{(a,q)}}} \sup_{|t| \leq \frac{XW^4}{H_1}} \left| \sum_{\ell \sim L/r} \frac{v_\ell r \chi(\ell)}{\ell^{1/2+it}} \right| \ll_C \frac{(L/r)^{1/2}}{W^{1/3}}, \quad (3.8)$$

then

$$\left| \frac{1}{H_1} \sum_{\substack{X < m_1 m_2 \ell \leq X+H_1 \\ m_j \sim M_j, \ell \sim L \\ m_1 m_2 \ell \equiv a \pmod{q}}} a_{m_1} b_{m_2} v_\ell \right| \ll d_3(q) \frac{\log^{O_C(1)} X}{W^{1/3}}.$$

For sums satisfying (I_2^{maj}) from Lemma 2.20 we shall use standard methods to deduce in Section 3.3 the following lemma.

Lemma 3.6. Let $\theta \in [1/2, 1)$ and $\varepsilon > 0$. Let $W \leq X^{\varepsilon/4}$ and let $X^{\theta+\varepsilon} \leq H_1 \leq H_2 \leq X/W^4$. Let $L, M_1, M_2 \geq 1$ be such that $M_j = X^{\alpha_j}$ and $LM_1M_2 \asymp X$. Let v_ℓ be bounded by $d_2^C(\ell)$. Assume that $a, q \in \mathbb{N}$ and

$$\alpha_1 + \alpha_2 \geq 1 - \theta. \quad (3.9)$$

Then

$$\left| \frac{1}{H_1} \sum_{\substack{X < m_1 m_2 \ell \leq X+H_1 \\ m_j \sim X^{\alpha_j} \\ m_1 m_2 \ell \equiv a \pmod{q}}} v_\ell - \frac{1}{H_2} \sum_{\substack{X < m_1 m_2 \ell \leq X+H_2 \\ m_j \sim X^{\alpha_j} \\ m_1 m_2 \ell \equiv a \pmod{q}}} v_\ell \right| \ll d_3(q) \frac{\log^{O_C(1)} X}{W^{1/6}}.$$

3.1. Proof of Proposition 3.2

The bound (3.2) follows immediately from the Siegel–Walfisz theorem (1.13) and the triangle inequality.

Before turning to the proof of (3.3), let us discuss the choice of Λ^\sharp . The prime number theorem with classical error term (see, e.g., [54, Theorem 6.9]) gives

$$\sum_{n \leq X} \Lambda(n) = X + O(X \exp(-c\sqrt{\log X})), \quad (3.10)$$

so that if one is interested only in the correlation of $\Lambda(n)$ with a constant function, one can select the simple approximant 1. However, this is not sufficient even for the maximal correlation with the constant function. There is some flexibility⁷ in how to select the approximant, but (following [61]) we use the Cramér–Granville model (1.1), which has the benefits of being a nonnegative model function and one that is known to be pseudorandom (which will be helpful in Section 9).

Proof of (3.3). It suffices to show that, for any $a, q \in \mathbb{N}$ and any $H_2 \in [X/\log^A X, X]$, we have

$$\left| \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} (\Lambda(n) - \Lambda^\sharp(n)) \right| \ll \frac{H_2}{\log^A X}.$$

We can clearly assume that $q < R$ and $(a, q) = 1$.

Let $D = \exp((\log X)^{3/5})$. By the fundamental lemma of the sieve (see e.g. [37, Fundamental Lemma 6.3 with $y = D, z = R$, and $\kappa = 1$]), there exist real numbers $\lambda_d^+ \in [-1, 1]$ such that, for any $H \geq 2$, $q < R$, and $a \in \mathbb{N}$ with $(a, q) = 1$, we have

$$\begin{aligned} \sum_{\substack{X < n \leq X+H \\ n \equiv a \pmod{q}}} \Lambda^\sharp(n) &\leq \frac{P(R)}{\varphi(P(R))} \sum_{\substack{d \leq D \\ d \mid P(R)}} \lambda_d^+ \sum_{\substack{X < n \leq X+H \\ n \equiv a \pmod{q} \\ d \mid n}} 1 \\ &= \prod_{p < R} \left(1 - \frac{1}{p}\right)^{-1} \sum_{\substack{d \leq D \\ d \mid P(R) \\ (d, q) = 1}} \lambda_d^+ \frac{H}{dq} + O(D \log R) \\ &= \frac{H}{\varphi(q)} \left(1 + O\left(\exp\left(-\frac{\log D}{\log R}\right)\right)\right) + O(D \log R), \end{aligned}$$

⁷For instance, a Fourier-analytic approximant $\Lambda^\sharp(n) := \sum_{q \leq Q} \frac{\mu(q)c_q(n)}{\phi(q)}$ is used in [28], where $c_q(n) := \sum_{1 \leq a \leq q: (a, q) = 1} e(an/q)$ denotes the Ramanujan sum. Another option is to use a truncated convolution sum, $\Lambda^\sharp(n) := -\sum_{d \mid n, d \leq R} \mu(d) \log d$, following e.g. [37, §19.2].

and also by the fundamental lemma we have a lower bound of the same shape. Hence, for $H \geq X^\varepsilon$ we have

$$\sum_{\substack{X < n \leq X+H \\ n=a(q)}} \Lambda^\sharp(n) = \frac{H}{\varphi(q)} + O_\varepsilon(H \exp(-(\log X)^{1/2})), \quad (3.11)$$

so (3.3) follows by the Siegel–Walfisz theorem and the triangle inequality. \square

Remark 3.7. One could improve the error term in (3.3) by adjusting the approximant Λ^\sharp to account for a potential Siegel zero; see for instance [37, Theorem 5.27] or [61, Proposition 2.2]. However, we will not do so here.

Before turning to the proof of (3.4) let us discuss the construction of the approximant d_k^\sharp which is a somewhat non-trivial task. The classical Dirichlet hyperbola method gives the asymptotic

$$\sum_{\substack{n \leq X \\ n=a(q)}} d_k(n) = X P_{k,a,q}(\log X) + O_{q,\varepsilon}(X^{1-1/k+\varepsilon}) \quad (3.12)$$

for any fixed a, q , any $\varepsilon > 0$, and some explicit polynomial $P_{k,a,q}$ of degree $k-1$ with coefficients depending only on k, a, q . Better error terms are known here; see e.g., [36, Section 13].

From (3.12), the triangle inequality, and Taylor expansion one has

$$\sum_{\substack{X < n \leq X+H \\ n=a(q)}} d_k(n) = H \left(P_{k,a,q}(\log X) + P'_{k,a,q}(\log X) + O_{q,\varepsilon} \left(\frac{X^{1-1/k+\varepsilon}}{H} + \frac{H}{X^{1-\varepsilon}} \right) \right)$$

for any $\varepsilon > 0$ whenever $2 \leq H \leq X$.

Hence we have to choose the approximant d_k^\sharp to also obey estimates such as

$$\sum_{\substack{X \leq n < X+H \\ n=a(q)}} d_k^\sharp(n) = H \left(P_{k,a,q}(\log X) + P'_{k,a,q}(\log X) + O_\varepsilon(X^{-\kappa_k} + H X^{\varepsilon-1}) \right) \quad (3.13)$$

for some $\kappa_k > 0$, with exactly the same choice of polynomial $P_{k,a,q}$.

The delta method of Duke, Friedlander and Iwaniec [9] can be used to build an approximant of a Fourier-analytic nature, basically by isolating the major arc components of d_k ; see [35], [5], [55], and [47, Proposition 4.2] for relevant calculations in this direction. However, the approximant that is (implicitly) constructed in these papers is very complicated, and somewhat difficult to deal with for our purposes (for instance, it is not evident whether it is non-negative).

The simpler approximant

$$d_k(n, A) := A^{1-k} \sum_{\substack{m|n \\ m \leq n^A}} d_{k-1}(m)$$

was recently proposed by Andrade and Smith [1] for various choices of parameter $0 < A < 1$. Unfortunately the polynomial $P_{k,a,q,A}(\log X)$ associated to this approximant usually only agrees with $P_{k,a,q}(\log X)$ to leading order (see [1, Theorem 2.1]), and so with this approximant one cannot hope to get polynomial saving like in our Theorem 1.1(iii).

Our approximant (1.2) with $P_m(t)$ as in (1.3) can be seen as a more complicated variant of the Andrade–Smith approximant. Note that the constraint $m \leq R_k^{2k-2}$ in (1.2) is redundant, as P_m vanishes for $m > R_k^{2k-2}$. Note also that (by adjusting the value of $c_{k,d,D}$ in Theorem 1.1) one could take R_k to

be any sufficiently small power of X , and that, for any $n \ll X$,

$$\begin{aligned} d_k^\sharp(n) &= \sum_{\substack{m \leq R_k^{2k-2} \\ m|n}} \sum_{j=0}^{k-1} \binom{k}{j} \sum_{\substack{n_1, \dots, n_j \leq R_k \\ n_{j+1}, \dots, n_{k-1} \leq R_k^2 \\ n_1 \cdots n_{k-1} = m}} \frac{\left(\log n - \log(n_1 \cdots n_j R_k^{k-j}) \right)^{k-j-1}}{(k-j-1)! \log^{k-j-1} R_k} \\ &\ll \sum_{m|n} d_{k-1}(m) = d_k(n) \end{aligned} \quad (3.14)$$

Recall we chose $R_k = X^{\frac{1}{10k}}$ in (1.2). The motivation for our approximant d_k^\sharp can be seen by noting that, sorting a factorization $n = n_1 \cdots n_k$ into terms $n_1, \dots, n_j \leq R_k$ and terms $n_{j+1}, \dots, n_k > R_k$, we get the generalized Dirichlet hyperbola identity

$$d_k(n) = \sum_{j=0}^{k-1} \binom{k}{j} \sum_{n_1, \dots, n_j \leq R_k} \sum_{\substack{n_{j+1}, \dots, n_{k-1} > R_k \\ \frac{n}{n_1 \cdots n_{k-1}} > R_k}} 1_{n_1 \cdots n_{k-1} | n}. \quad (3.15)$$

The polynomials $P_m(t)$ are chosen to match with the contribution from the sum over n_{j+1}, \dots, n_{k-1} as can be seen from the proof of (3.4) that we now give.

Proof of (3.4). It suffices to show that, for any $k \geq 2$, any $a, q \in \mathbb{N}$, and any $H_2 \in [X^{1-1/(50k)}, X]$, we have

$$\left| \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} (d_k(n) - d_k^\sharp(n)) \right| \ll \frac{H_2^2}{X} \log^{k-2} X.$$

Since $d_k(n) = O_\varepsilon(n^\varepsilon)$, we can clearly assume that $q \leq X^{\frac{1}{40k}}$. Using (3.15) we obtain

$$\begin{aligned} \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} d_k(n) &= \sum_{\substack{a_i \pmod{q} \\ a_1 \cdots a_k \equiv a \pmod{q}}} \sum_{j=0}^{k-1} \binom{k}{j} \sum_{\substack{n_1, \dots, n_j \leq R_k \\ n_i \equiv a_i \pmod{q}}} \sum_{\substack{n_{j+1}, \dots, n_{k-1} > R_k \\ \frac{X}{n_1 \cdots n_{k-1}} > R_k \\ n_i \equiv a_i \pmod{q}}} \left(\frac{H_2}{qn_1 \cdots n_{k-1}} + O(1) \right) \\ &\quad + O \left(\sum_{n_1, \dots, n_j \leq R_k} \sum_{\substack{n_{j+1}, \dots, n_{k-1} > R_k \\ \frac{X+H_2}{n_1 \cdots n_{k-1}} > R_k \\ \frac{X}{n_1 \cdots n_{k-1}} > R_k}} \left(\frac{H_2}{n_1 \cdots n_{k-1}} + 1 \right) \right). \end{aligned}$$

Let us consider the two error terms. The first error term contributes, using the inequality $1 < X/(R_k n_1 \cdots n_{k-1})$,

$$\ll \sum_{a_k \pmod{q}} \sum_{n_1, \dots, n_{k-1} \leq X} \frac{X}{R_k n_1 \cdots n_{k-1}} \ll q \frac{X}{R_k} \log^{k-1} X \ll \frac{H_2^2}{X} \log^{k-2} X$$

since $q \leq X^{\frac{1}{40k}}$, $R_k = X^{\frac{1}{10k}}$, and $H_2 \geq X^{1 - \frac{1}{50k}}$. The second error term contributes, using $n_1 \cdots n_{k-1} \asymp X/R_k$ and Shiu's bound (Lemma 2.17),

$$\ll \sum_{\substack{n_1, \dots, n_{k-1} \leq 2X \\ \frac{X}{R_k} < n_1 \cdots n_{k-1} \leq \frac{X+H_2}{R_k}}} \frac{R_k H_2}{X} = \frac{R_k H_2}{X} \sum_{\frac{X}{R_k} < n < \frac{X+H_2}{R_k}} d_{k-1}(n) \ll \frac{H_2^2}{X} \log^{k-2} X.$$

Hence

$$\begin{aligned} & \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} d_k(n) \\ &= \frac{H_2}{q} \sum_{\substack{a_i \pmod{q} \\ a_1 \cdots a_k \equiv a \pmod{q}}} \sum_{j=0}^{k-1} \binom{k}{j} \sum_{\substack{n_1, \dots, n_j \leq R_k \\ n_i \equiv a_i \pmod{q}}} \frac{1}{n_1 \cdots n_j} \sum_{\substack{n_{j+1}, \dots, n_{k-1} > R_k \\ \frac{X}{n_1 \cdots n_{k-1}} > R_k \\ n_i \equiv a_i \pmod{q}}} \frac{1}{n_{j+1} \cdots n_{k-1}} \\ & \quad + O\left(\frac{H_2^2}{X} \log^{k-2} X\right). \end{aligned} \quad (3.16)$$

For any $B \geq A \geq 1$, we have

$$\sum_{\substack{A < n < B \\ n \equiv a \pmod{q}}} \frac{1}{n} = \frac{1}{q} \int_A^B \frac{1}{t} dt + O\left(\frac{1}{A}\right).$$

Applying this $k-1-j$ times, we see that⁸

$$\begin{aligned} & \sum_{\substack{n_{j+1}, \dots, n_{k-1} > R_k \\ \frac{X}{n_1 \cdots n_{k-1}} > R_k \\ n_i \equiv a_i \pmod{q}}} \frac{1}{n_{j+1} \cdots n_{k-1}} \\ &= \frac{1}{q^{k-1-j}} \int_{\substack{t_{j+1}, \dots, t_{k-1} > R_k \\ t_{j+1} \cdots t_{k-1} \leq \frac{X}{n_1 \cdots n_j R_k}}} \frac{dt_{j+1} \cdots dt_{k-1}}{t_{j+1} \cdots t_{k-1}} + O\left(\frac{(\log X)^{k-1-j-1}}{q^{k-1-j-1}} \cdot \frac{1}{R_k}\right) \quad (3.17) \\ &= \frac{1}{q^{k-1-j}} \frac{\log^{k-j-1} \frac{X}{n_1 \cdots n_j R_k^{k-j}}}{(k-j-1)!} + O\left(\frac{(\log X)^{k-j-2}}{q^{k-j-2}} \cdot \frac{1}{R_k}\right). \end{aligned}$$

Since $R_k = X^{\frac{1}{10k}}$, $q \leq X^{\frac{1}{40k}}$, and $H_2 \geq X^{1 - \frac{1}{50k}}$, the error term contributes to (3.16)

$$\begin{aligned} & \ll \frac{H_2}{q} \sum_{j=0}^{k-1} \sum_{a_{j+1}, \dots, a_k \pmod{q}} \sum_{n_1, \dots, n_j \leq R_k} \frac{1}{n_1 \cdots n_j} \cdot \frac{(\log X)^{k-j-2}}{q^{k-j-2}} \cdot \frac{1}{R_k} \\ & \ll \frac{H_2}{q} \sum_{j=0}^{k-1} q^{k-j} (\log X)^j \cdot \frac{(\log X)^{k-j-2}}{q^{k-j-2}} \cdot \frac{1}{R_k} \ll \frac{H_2^2}{X} \log^{k-2} X. \end{aligned}$$

⁸To obtain the second equality we use the classical formula $\int_{x_1, \dots, x_d \geq 0: x_1 + \cdots + x_d \leq L} 1 dx_1 \cdots dx_d = \frac{L^d}{d!}$ for the volume of a simplex (easily proven by induction on d and the Fubini–Tonelli theorem combined with the change of variables $x_i = \log \frac{t_{i+1}}{R}$ for $i = 1, \dots, k-j-1$).

Hence (3.16) and (3.17) give

$$\begin{aligned} \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} d_k(n) &= \frac{H_2}{q^{k-j}} \sum_{\substack{a_i \pmod{q} \\ a_1 \cdots a_k \equiv a \pmod{q}}} \sum_{j=0}^{k-1} \binom{k}{j} \sum_{\substack{n_1, \dots, n_j \leq R_k \\ n_i \equiv a_i \pmod{q}}} \frac{\log^{k-j-1} \frac{X}{n_1 \cdots n_j R_k^{k-j}}}{(k-j-1)! n_1 \cdots n_j} \\ &\quad + O\left(\frac{H_2^2}{X} \log^{k-2} X\right). \end{aligned}$$

On the other hand, by definition,

$$\begin{aligned} &\sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} d_k^\sharp(n) \\ &= \sum_{\substack{a_i \pmod{q} \\ a_1 \cdots a_k \equiv a \pmod{q}}} \sum_{j=0}^{k-1} \binom{k}{j} \sum_{\substack{n_1, \dots, n_j \leq R_k \\ n_i \equiv a_i \pmod{q}}} \frac{\log^{k-j-1} \frac{X}{n_1 \cdots n_j R_k^{k-j}} + O\left(\frac{H_2}{X} \log^{k-j-2} X\right)}{(k-j-1)! \log^{k-j-1} R_k} \\ &\quad \cdot \sum_{\substack{R_k < n_{j+1}, \dots, n_{k-1} \leq R_k^2 \\ n_i \equiv a_i \pmod{q}}} \left(\frac{H_2}{qn_1 \cdots n_{k-1}} + O(1) \right) \\ &= \frac{H_2}{q} \sum_{\substack{a_i \pmod{q} \\ a_1 \cdots a_k \equiv a \pmod{q}}} \sum_{j=0}^{k-1} \binom{k}{j} \sum_{\substack{n_1, \dots, n_j \leq R_k \\ n_i \equiv a_i \pmod{q}}} \frac{\log^{k-j-1} \frac{X}{n_1 \cdots n_j R_k^{k-j}}}{(k-j-1)! \log^{k-j-1} R_k} \\ &\quad \sum_{\substack{R_k < n_{j+1}, \dots, n_{k-1} \leq R_k^2 \\ n_i \equiv a_i \pmod{q}}} \frac{1}{n_1 \cdots n_{k-1}} \\ &\quad + O\left(\sum_{a_k \pmod{q}} \sum_{j=0}^{k-1} \sum_{n_1, \dots, n_j \leq R_k} \frac{H_2}{X \log X} \sum_{R_k < n_{j+1}, \dots, n_{k-1} \leq R_k^2} \left(\frac{H_2}{qn_1 \cdots n_{k-1}} + 1 \right) \right) \\ &\quad + O\left(\sum_{a_k \pmod{q}} \sum_{j=0}^{k-1} \sum_{n_1, \dots, n_j \leq R_k} \sum_{R_k < n_{j+1}, \dots, n_{k-1} \leq R_k^2} 1 \right). \end{aligned}$$

The error terms contribute

$$\ll \frac{H_2^2}{X} \log^{k-2} X + q \frac{H_2}{X \log X} R_k^{2(k-1)} + q R_k^{2(k-1)} \ll \frac{H_2^2}{X} \log^{k-2} X + q X^{1/2}$$

and in the main term

$$\frac{1}{\log^{k-j-1} R_k} \sum_{\substack{R_k < n_{j+1}, \dots, n_{k-1} \leq R_k^2 \\ n_i \equiv a_i \pmod{q}}} \frac{1}{n_{j+1} \cdots n_{k-1}} = \left(\frac{1}{q} + O\left(\frac{1}{R_k}\right) \right)^{k-j-1}.$$

The claim follows since $R_k = X^{\frac{1}{10k}}$ and $q \leq X^{\frac{1}{40k}}$. \square

3.2. Proof of Lemma 3.3

Note first that the claims are trivial unless $q \leq X^{1/80}$. For part (ii), note that, for $j = 1, 2$,

$$\begin{aligned}
& \frac{1}{H_j} \sum_{\substack{X < n \leq X+H_j \\ n \equiv a \pmod{q}}} d_k^\sharp(n) \\
&= \frac{1}{H_j} \sum_{\substack{b,c \pmod{q} \\ bc \equiv a \pmod{q}}} \sum_{\substack{m \leq X^{\frac{2k-2}{10k}} \\ m \equiv b \pmod{q}}} \left(P_m(\log X) + O\left(d_{k-1}(m) \frac{H_j}{X \log X}\right) \right) \sum_{\substack{X/m < n \leq (X+H_j)/m \\ n \equiv c \pmod{q}}} 1 \\
&= \frac{1}{H_j} \sum_{\substack{b,c \pmod{q} \\ bc \equiv a \pmod{q}}} \sum_{\substack{m \leq X^{\frac{k-1}{5k}} \\ m \equiv b \pmod{q}}} \left(P_m(\log X) + O\left(d_{k-1}(m) \frac{H_j}{X \log X}\right) \right) \left(\frac{H_j}{mq} + O(1) \right) \\
&= \sum_{\substack{b,c \pmod{q} \\ bc \equiv a \pmod{q}}} \sum_{\substack{m \leq X^{\frac{k-1}{5k}} \\ m \equiv b \pmod{q}}} \frac{P_m(\log X)}{mq} + O\left(\frac{H_j \log^{k-2} X}{X} + \frac{qX^{1/5}}{H_j} \right).
\end{aligned}$$

The claim follows by subtracting this for $j = 1, 2$. Part (i) follows directly from (3.11) applied with $H \in \{H_1, H_2\}$ and the triangle inequality.

3.3. Proof of Lemmas 3.5 and 3.6

We first make a standard reduction to studying averages of Dirichlet polynomials.

Lemma 3.8. *Let $W \leq X^{1/100}$. Let $|a_n| \leq d_2(n)^C$ for some $C \geq 1$ and let $A(s, \chi) := \sum_{c_1 X < n \leq c_2 X} a_n \chi(n) n^{-s}$ for some fixed $c_2 > c_1 > 0$. Let $X^{1/2} \leq H_1 \leq H_2 \leq X/W^4$ and $(a, q) = 1$.*

(i) *One has*

$$\begin{aligned}
& \left| \frac{1}{H_1} \sum_{\substack{X < n \leq X+H_1 \\ n \equiv a \pmod{q}}} a_n - \frac{1}{H_2} \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} a_n \right| \ll \frac{\log^{O_C(1)} X}{W^2} \\
&+ \frac{\log X}{X^{1/2}} \max_{\frac{X}{H_1} \leq T \leq \frac{XW^4}{H_1}} \frac{1}{\varphi(q)} \sum_{\chi \pmod{q}} \frac{X/H_1}{T} \int_{W \leq |t| \leq T} |A(\frac{1}{2} + it, \chi)| dt.
\end{aligned}$$

(ii) *One has*

$$\begin{aligned}
& \left| \frac{1}{H_1} \sum_{\substack{X < n \leq X+H_1 \\ n \equiv a \pmod{q}}} a_n \right| \ll \frac{\log^{O_C(1)} X}{W^2} \\
&+ \frac{\log X}{X^{1/2}} \max_{\frac{X}{H_1} \leq T \leq \frac{XW^4}{H_1}} \frac{1}{\varphi(q)} \sum_{\chi \pmod{q}} \frac{X/H_1}{T} \int_{|t| \leq T} |A(\frac{1}{2} + it, \chi)| dt.
\end{aligned}$$

Proof. Let us first consider part (i). We begin by using the orthogonality of characters and Perron's formula (see e.g. [54, Corollary 5.3]) to get that, for $j = 1, 2$,

$$\begin{aligned} \frac{1}{H_j} \sum_{\substack{X < n \leq X+H_j \\ n \equiv a \pmod{q}}} a_n &= \frac{1}{\varphi(q)H_j} \sum_{\chi \pmod{q}} \bar{\chi}(a) \int_{-\frac{XW^4}{H_j}}^{\frac{XW^4}{H_j}} A\left(\frac{1}{2} + it, \chi\right) \frac{(X+H_j)^{1/2+it} - X^{1/2+it}}{\frac{1}{2} + it} dt \\ &\quad + O\left(\frac{\log^{O_C(1)} X}{W^4}\right). \end{aligned}$$

The “main term” comes from (only χ_0 contributes to actual main terms)

$$\begin{aligned} \frac{1}{\varphi(q)H_j} \sum_{\chi \pmod{q}} \bar{\chi}(a) \int_{-W}^W A\left(\frac{1}{2} + it, \chi\right) \frac{(X+H_j)^{1/2+it} - X^{1/2+it}}{\frac{1}{2} + it} dt \\ = \frac{1}{\varphi(q)} \sum_{\chi \pmod{q}} \bar{\chi}(a) \int_{-W}^W A\left(\frac{1}{2} + it, \chi\right) X^{-1/2+it} dt + O\left(\frac{H_j W^2}{X} \log^{O_C(1)} X\right). \end{aligned}$$

The error term is $O(\log^{O_C(1)} X/W^2)$ while the main term is independent of j . Hence

$$\begin{aligned} \left| \frac{1}{H_1} \sum_{\substack{X < n \leq X+H_1 \\ n \equiv a \pmod{q}}} a_n - \frac{1}{H_2} \sum_{\substack{X < n \leq X+H_2 \\ n \equiv a \pmod{q}}} a_n \right| &\ll \frac{\log^{O_C(1)} X}{W^2} \\ &\quad + \sum_{j=1}^2 \frac{1}{\varphi(q)H_j} \sum_{\chi \pmod{q}} \int_{W \leq |t| \leq \frac{XW^4}{H_j}} |A\left(\frac{1}{2} + it, \chi\right)| \left| \frac{(X+H_j)^{1/2+it} - X^{1/2+it}}{\frac{1}{2} + it} \right| dt \end{aligned}$$

Since $\left| \frac{(X+H_j)^{1/2+it} - X^{1/2+it}}{1/2+it} \right| \ll \min\{H_j X^{-1/2}, X^{1/2}/(1+|t|)\}$, the second line contributes

$$\begin{aligned} &\ll \sum_{j=1}^2 \frac{1}{\varphi(q)H_j} \sum_{\chi \pmod{q}} \frac{H_j}{X^{1/2}} \int_{W \leq |t| \leq \frac{X}{H_j}} |A\left(\frac{1}{2} + it, \chi\right)| dt \\ &\quad + \sum_{j=1}^2 \frac{1}{\varphi(q)H_j} \sum_{\chi \pmod{q}} \int_{\frac{X}{H_j} \leq |t| \leq \frac{XW^4}{H_j}} |A\left(\frac{1}{2} + it, \chi\right)| \frac{X^{1/2}}{1+|t|} dt. \end{aligned}$$

Splitting the second integral dyadically, we see that this is

$$\ll \frac{\log X}{X^{1/2}} \sum_{j=1}^2 \max_{\frac{X}{H_j} \leq T \leq \frac{XW^4}{H_j}} \frac{1}{\varphi(q)} \sum_{\chi \pmod{q}} \frac{X/H_j}{T} \int_{W \leq |t| \leq T} |A\left(\frac{1}{2} + it, \chi\right)| dt.$$

Since $H_2 \geq H_1$, the contribution of the part with $j = 1$ is larger than the contribution of the part with $j = 2$. Hence part (i) follows.

Part (ii) follows similarly, except there is no need to handle a main term separately. \square

Proof of Lemma 3.5. By Shiu's bound (Lemma 2.17) we can clearly assume that $q \leq W^{1/2} \leq X^{\varepsilon/400}$. Let us consider, for $j = 1, 2$,

$$\frac{1}{H_j} \sum_{\substack{X < m_1 m_2 \ell \leq X+H_j \\ m_1 m_2 \ell \equiv a \pmod{q} \\ m_j \sim M_j, \ell \sim L}} a_{m_1} b_{m_2} v_\ell.$$

We first split the sums according to $r_1 = (m_1, q)$, $r_2 = (m_2, q/r_1)$ and $r_3 = (\ell, q/(r_1 r_2))$, writing $m_j = r_j m'_j$ and $\ell = r_3 \ell'$. Then $m'_1 m'_2 \ell' r_1 r_2 r_3 \equiv a \pmod{\frac{q}{r_1 r_2 r_3} r_1 r_2 r_3}$ and necessarily $r_1 r_2 r_3 = (a, q)$. We have

$$\begin{aligned} & \frac{1}{H_j} \sum_{\substack{X < m_1 m_2 \ell \leq X + H_j \\ m_1 m_2 \ell \equiv a \pmod{q} \\ m_j \sim M_j, \ell \sim L}} a_{m_1} b_{m_2} v_\ell \\ &= \sum_{r_1 r_2 r_3 = (a, q)} \frac{1}{H_j} \sum_{\substack{X/(r_1 r_2 r_3) < m'_1 m'_2 \ell' \leq (X + H_j)/(r_1 r_2 r_3) \\ m'_1 m'_2 \ell' \equiv \frac{a}{r_1 r_2 r_3} \pmod{\frac{q}{r_1 r_2 r_3}} \\ (m'_1, q/r_1) = (m'_2, q/(r_1 r_2)) = (\ell', q/(r_1 r_2 r_3)) = 1 \\ m'_j \sim M_j/r_j, \ell' \sim L/r_3}} a_{m'_1 r_1} b_{m'_2 r_2} v_{\ell' r_3}. \end{aligned}$$

Part (i) follows from Lemma 3.8 (with $X/(a, q)$, $H_j/(a, q)$, $q/(a, q)$, and $a/(a, q)$ in place of X , H_j , q , and a) if, for any $T \in [X/H_1, XW^4/H_1]$ and any $r_1 r_2 r_3 = (a, q)$ and any $\chi \pmod{q/(a, q)}$, one has

$$\begin{aligned} & \int_{W \leq |t| \leq T} \left| \sum_{\substack{m'_1 \sim M_1/r_1 \\ (m_1, q/r_1) = 1}} \frac{a_{m'_1 r_1} \chi(m'_1)}{m'_1^{1/2+it}} \sum_{\substack{m'_2 \sim M_2/r_2 \\ (m'_2, q/(r_1 r_2)) = 1}} \frac{b_{m'_2 r_2} \chi(m'_2)}{m'_2^{1/2+it}} \sum_{\substack{\ell' \sim L/r_3 \\ (\ell', q/(r_1 r_2 r_3)) = 1}} \frac{v_{\ell' r_3} \chi(\ell')}{\ell'^{1/2+it}} \right| dt \\ & \ll \frac{\log^{O_C(1)} X}{W^{1/3}} \frac{T}{X/H_1} \left(\frac{X}{(a, q)} \right)^{1/2}. \end{aligned}$$

But, using the assumption (3.7), this follows from a slight variant of [4, Lemma 9] with $g = 1$ in cases $\theta \in \{7/12, 3/5, 5/8\}$ and with $g = 2$ in case $\theta = 11/20$ (alternatively see [24, Lemma 7.3]). The idea in the proofs of these lemmas is to first split the integral to level sets according to the absolute values of the three Dirichlet polynomials appearing, and then to apply appropriate mean and large value results individually for the three Dirichlet polynomials to obtain upper bounds for the sizes of the level sets. Combining these upper bounds using case-by-case study and Hölder's inequality leads to the lemmas.

Part (ii) follows similarly. \square

In fact, one can establish Lemma 3.5 for $\theta \in [7/12, 5/8]$ by using [4, Lemma 9] with $g = 1$, and for $\theta \in [11/20, 9/16]$ by using [4, Lemma 9] with $g = 2$ (see [24, end of Section 7.2]), but we shall not need this more general result.

Proof of Lemma 3.6. By Shiu's bound (Lemma 2.17) we can assume that $q \leq W^{1/6}$. Notice first that if for either $i = 1$ or $i = 2$, we have $\theta + \varepsilon - (1 - \alpha_i) \geq \varepsilon$, then we can obtain the claim by simply moving the sum over m_i inside. Hence we can assume that $\alpha_1, \alpha_2 < 1 - \theta$.

Arguing as in proof of Lemma 3.5 and doing a dyadic splitting it suffices to show that, for any $T \in [W, XW^4/H_1]$ and any $r_1 r_2 r_3 = (a, q)$,

$$\begin{aligned} & \frac{1}{\varphi(\frac{q}{(a, q)})} \sum_{\chi \pmod{\frac{q}{(a, q)}}} \int_T^{2T} \left| \sum_{\substack{m_1 \sim M_1/r_1 \\ (m_1, q/r_1) = 1}} \frac{\chi(m_1)}{m_1^{1/2+it}} \sum_{\substack{m_2 \sim M_2/r_2 \\ (m_2, q/(r_1 r_2)) = 1}} \frac{\chi(m_2)}{m_2^{1/2+it}} \sum_{\substack{\ell \sim L/r_3 \\ (\ell, q/(r_1 r_2 r_3)) = 1}} \frac{\chi(\ell) v_{\ell r_3}}{\ell^{1/2+it}} \right| dt \\ & \ll \frac{\log^{O(1)} X}{W^{1/6}} \max \left\{ \frac{T}{X/H_1}, 1 \right\} \left(\frac{X}{(a, q)} \right)^{1/2}. \end{aligned} \tag{3.18}$$

By the fourth moment estimate for Dirichlet L -functions we have (see [24, Lemma 10.11]), for any $M, T \geq 2$ and $d \mid (a, q)$,

$$\begin{aligned} \sum_{\substack{\chi \pmod{\frac{q}{(a,q)}} \\ (m, q/d) = 1}} \int_T^{2T} \left| \sum_{m \sim M} \frac{\chi(m)}{m^{1/2+it}} \right|^4 dt &\ll \sum_{\substack{\chi \pmod{\frac{q}{d}} \\ (m, q/d) = 1}} \int_T^{2T} \left| \sum_{m \sim M} \frac{\chi(m)}{m^{1/2+it}} \right|^4 dt \\ &\ll \left(q^3 T + \frac{qM^2}{T^3} \right) \log^{O(1)}(MT). \end{aligned}$$

Hence, using also Hölder and the mean value theorem (see e.g. [37, Theorem 9.12 with $k = q$ and $Q = 1$]), the left-hand side of (3.18) is

$$\begin{aligned} &\ll \log^{O(1)} X \left(q^2 T + \frac{X^{2\alpha_1}}{T^3} \right)^{1/4} \left(q^2 T + \frac{X^{2\alpha_2}}{T^3} \right)^{1/4} \left(T + \frac{X^{1-\alpha_1-\alpha_2}}{q} \right)^{1/2} \\ &\ll q \log^{O(1)} X \left(T + T^{1/2} X^{1/2-\alpha_1/2-\alpha_2/2} + X^{\alpha_1/2} + X^{\alpha_2/2} + \frac{X^{1/2-\alpha_1/2}}{T^{1/2}} + \frac{X^{1/2-\alpha_2/2}}{T^{1/2}} + \frac{X^{1/2}}{T^{3/2}} \right). \end{aligned}$$

One can see that this is always at most the right-hand side of (3.18) by considering each term separately — depending on the term, the worst case is either $T = W$ or $T = X/H_1$. \square

3.4. Proof of Proposition 3.4

Let us first show the $k = 2$ case of Proposition 3.4(ii). It follows from classical arguments leading to the exponent $1/3 + \varepsilon$ in the Dirichlet divisor problem (see e.g. [62, Section I.6.4]). For completeness, we provide the proof here. By a trivial bound we can assume that $q \leq X^{\varepsilon/4}$.

First note that

$$\frac{1}{H_j} \sum_{\substack{X < n \leq X+H_j \\ n \equiv a \pmod{q}}} d_2(n) = \frac{2}{H_j} \sum_{\substack{X < mn \leq X+H_j \\ m \leq X^{1/2} \\ mn \equiv a \pmod{q}}} 1 + O\left(\frac{1}{H_j} \sum_{m \in (X^{1/2}, (X+H_j)^{1/2}]} \sum_{\substack{X/m < n \leq (X+H_j)/m \\ mn \equiv a \pmod{q}}} 1 \right).$$

The error term contributes

$$\ll \frac{1}{H_j} \cdot \left(\frac{H_j}{X^{1/2}} + 1 \right) \cdot \left(\frac{H_j}{X^{1/2}} + 1 \right) \ll \frac{H_j}{X} + \frac{1}{H_j}.$$

Hence it suffices to show that, for any $M \in [1/2, X^{1/2}]$, we have

$$\frac{1}{H_1} \sum_{\substack{X < mn \leq X+H_1 \\ m \sim M \\ mn \equiv a \pmod{q}}} 1 = \frac{1}{H_2} \sum_{\substack{X < mn \leq X+H_2 \\ m \sim M \\ mn \equiv a \pmod{q}}} 1 + O\left(\frac{1}{X^{\varepsilon/5}} \right).$$

Now, for $j = 1, 2$,

$$\begin{aligned}
\sum_{\substack{X < mn \leq X + H_j \\ m \sim M \\ mn \equiv a \pmod{q}}} 1 &= \sum_{\substack{0 \leq b, c < q \\ bc \equiv a \pmod{q}}} \sum_{\substack{m \sim M \\ m \equiv b \pmod{q}}} \left(\sum_{\substack{1 \leq n \leq \frac{X+H_j}{m} \\ n \equiv c \pmod{q}}} 1 - \sum_{\substack{1 \leq n \leq \frac{X}{m} \\ n \equiv c \pmod{q}}} 1 \right) \\
&= \sum_{\substack{0 \leq b, c < q \\ bc \equiv a \pmod{q}}} \sum_{\substack{m \sim M \\ m \equiv b \pmod{q}}} \left(\left\lfloor \frac{X+H_j}{mq} - \frac{c}{q} \right\rfloor - \left\lfloor \frac{X}{mq} - \frac{c}{q} \right\rfloor \right) \\
&= \sum_{\substack{0 \leq b, c < q \\ bc \equiv a \pmod{q}}} \sum_{\substack{m \sim M \\ m \equiv b \pmod{q}}} \left(\frac{H_j}{mq} + \left(\frac{1}{2} - \left\{ \frac{X+H_j}{mq} - \frac{c}{q} \right\} \right) - \left(\frac{1}{2} - \left\{ \frac{X}{mq} - \frac{c}{q} \right\} \right) \right).
\end{aligned}$$

Hence it suffices to show that, for $j = 1, 2$ and $\xi \in \{X/q, (X + H_j)/q\}$,

$$\sum_{\substack{0 \leq b, c < q \\ bc \equiv a \pmod{q}}} \sum_{\substack{m \sim M \\ m \equiv b \pmod{q}}} \left(\frac{1}{2} - \left\{ \frac{\xi}{m} - \frac{c}{q} \right\} \right) = O \left(\frac{H_j}{X^{\varepsilon/5}} \right). \quad (3.19)$$

The left-hand side is trivially $O(qM) = O(X^{\varepsilon/4}M)$ and so (3.19) is immediate in case $M \leq H_j/X^{\varepsilon/2}$, and so we can concentrate on showing (3.19) for j and M for which $M > H_j/X^{\varepsilon/2}$.

For any $K \geq 1$ we have the Fourier expansion (see e.g. [62, Section I.6.4])

$$\frac{1}{2} - \{y\} = \sum_{k \neq 0} v_k e(ky) + O(1/K) \quad \text{with} \quad v_k \ll \min\{1/k, K/k^2\}.$$

Taking $K_j = MX^{\varepsilon/2}/H_j$ (which is ≥ 1) and writing $m = b + rq$, it suffices to show that, for $j = 1, 2$ and $\xi \in \{X/q, (X + H_j)/q\}$,

$$\sum_{|k| > 0} \min \left\{ \frac{1}{k}, \frac{MX^{\varepsilon/2}/H_j}{k^2} \right\} \left| \sum_{(M-b)/q < r \leq (2M-b)/q} e(k\xi/(b + rq)) \right| = O(X^{-\varepsilon/2}H_j/q^2).$$

The second derivative of the phase has size $\asymp kXq/M^3$, so that by van der Corput's exponential sum bound (see e.g. [62, Theorem 5 in Section I.6.3] or [37, Corollary 8.13]), the left-hand side is

$$\begin{aligned}
&\ll \sum_{0 < |k| \leq MX^{\varepsilon/2}/H_j} \frac{1}{k} \left(\left(\frac{kXq}{M^3} \right)^{1/2} \frac{M}{q} + \left(\frac{M^3}{kXq} \right)^{1/2} \right) \\
&\quad + \sum_{|k| > MX^{\varepsilon/2}/H_j} \frac{MX^{\varepsilon/2}/H_j}{k^2} \left(\left(\frac{kXq}{M^3} \right)^{1/2} \frac{M}{q} + \left(\frac{M^3}{kXq} \right)^{1/2} \right) \\
&\ll \frac{X^{1/2+\varepsilon/4}}{H_j^{1/2}q^{1/2}} + \frac{M^{3/2}}{q^{1/2}X^{1/2}}.
\end{aligned}$$

This is $\ll X^{-\varepsilon/2}H_j/q^2$ since $H_2 \geq H_1 \geq X^{1/3+\varepsilon}$, $q \leq X^{\varepsilon/4}$, and $M \leq X^{1/2}$. This establishes the $k = 2$ case of Proposition 3.4.

The cases $k = 3, 4$ of Proposition 3.4(ii) follow from dyadic splitting, Lemma 2.20(v), and Lemma 3.6 with $W = \min\{X^{\frac{1}{400k}}, X^{\varepsilon/4}\}$, so we can concentrate on Proposition 3.4(i) and cases $k \geq 5$ of Proposition 3.4(ii). To apply Lemma 3.5 we need parts (i) and (ii) of the following lemma (part (iii) will be used in the proof of Lemma 4.5 below):

Lemma 3.9 (Dirichlet polynomial bounds). *Let $0 \leq T_0 \leq X$ and $\alpha \in (0, 1]$.*

(i) *There exists $\delta = \delta(\alpha)$ such that, for any character χ of modulus $q \leq X^{\alpha/2}$ and any $L \in [X^\alpha, X]$,*

$$\sup_{T_0 \leq |t| \leq X} \sup_{I \subset [L, 2L]} \left| \sum_{\ell \in I} \frac{\chi(\ell)}{\ell^{1/2+it}} \right| \ll_\alpha L^{1/2} X^{-\delta} + L^{1/2} \frac{\log X}{(T_0 + 1)^{1/2}}.$$

(ii) *For any $A > 0$, any $1 \leq r \leq X$, and any character χ of modulus $q \leq \log^A X$, one has*

$$\sup_{|t| \leq X} \sup_{I \subset [X^\alpha, 2X^\alpha]} \left| \sum_{\ell \in I} \frac{\mu(r\ell)\chi(\ell)}{\ell^{1/2+it}} \right| \ll_{\alpha, A} \frac{X^{\alpha/2}}{\log^A X}.$$

(iii) *Let $\varepsilon > 0$. For any $A > 0$, any $P \in [\exp((\log X)^{2/3+\varepsilon}), X^2]$ and any character χ of modulus $q \leq \log^A X$,*

$$\sup_{T_0 \leq |t| \leq X} \sup_{I \subset [P, 2P]} \left| \sum_{p \in I} \frac{\chi(p)}{p^{1/2+it}} \right| \ll_{\varepsilon, A} \frac{P^{1/2}}{T_0} + \frac{P^{1/2}}{\log^A X}.$$

Proof. Parts (ii) and (iii) follow by standard contour integration arguments, using the known zero-free region for $L(s, \chi)$ (see e.g., [44, Lemma 2] for a similar argument without the character).

Let us concentrate on part (i). By partial summation, splitting into residue classes $a \pmod{q}$ and writing $\ell = mq + a$, it suffices to show that, for any $a \in \{1, \dots, q\}$ and $|t| \in [T_0, X]$, we have

$$\sum_{m \in \frac{1}{q}I} e\left(\frac{t}{2\pi} \log(mq + a)\right) \ll L \frac{X^{-\delta}}{q} + L \frac{\log X}{q(T_0 + 1)^{1/2}}. \quad (3.20)$$

The ν th derivative of the phase $g(m) = \frac{t}{2\pi} \log(mq + a)$ satisfies

$$|g^{(\nu)}(m)| \frac{m^\nu}{\nu!} \asymp_\nu |t|$$

for any $\nu \geq 1$. We apply the Weyl bound in the form of [37, Theorem 8.4]. When $T_0 \leq |t| \leq L/q$, we use [37, Theorem 8.4] with $k = 2$, obtaining

$$\sum_{m \in \frac{1}{q}I} e\left(\frac{t}{2\pi} \log(mq + a)\right) \ll \left(\frac{|t|}{L^2/q^2} + \frac{1}{|t|} \right)^{1/2} \frac{L}{q} \log X \ll \frac{L^{1/2}}{q^{1/2}} \log X + \frac{L \log X}{q(T_0 + 1)^{1/2}}.$$

Recalling that $q \leq L^{1/2}$, the bound (3.20) follows with $\delta = \alpha/5$.

On the other hand, when $L/q < |t| \leq X$, we use [37, Theorem 8.4] with $k = \lfloor \frac{2}{\alpha} + 2 \rfloor$, obtaining

$$\begin{aligned} \sum_{m \in \frac{1}{q}I} e\left(\frac{t}{2\pi} \log(mq + a)\right) &\ll_{\alpha} \left(\frac{|t|}{(L/q)^k} + \frac{1}{|t|}\right)^{\frac{4}{k^2k}} \frac{L}{q} \log X \\ &\ll_{\alpha} \left(\frac{X}{(L^{1/2})^k} + \frac{1}{L^{1/2}}\right)^{\frac{4}{k^2k}} \frac{L}{q} \log X \\ &\ll_{\alpha} \frac{L^{1-\frac{2}{k^2k}}}{q} \log X \end{aligned} \quad (3.21)$$

and (3.20) follows. \square

Let us now get back to the proof of Proposition 3.4(ii). Recall that we can assume that $k \geq 5$. The claim follows trivially unless $q \leq \min\{X^{2c_k}, X^{\varepsilon/900}\}$. We can request that $c_k \leq \frac{1}{4000k}$. By dyadic splitting it suffices to show that, for any $N_j \in [1/2, X]$ with $N_1 \cdots N_k \asymp X$, one has

$$\max_{\substack{a, q \in \mathbb{N} \\ q \leq X^{1/(2000k)}}} \left| \frac{1}{H_1} \sum_{\substack{X < n_1 \cdots n_k \leq X+H_1 \\ n_i \sim N_i \\ n_1 \cdots n_k \equiv a \pmod{q}}} 1 - \frac{1}{H_2} \sum_{\substack{X < n_1 \cdots n_k \leq X+H_2 \\ n_i \sim N_i \\ n_1 \cdots n_k \equiv a \pmod{q}}} 1 \right| \ll \frac{1}{X^{2c_k}} + \frac{1}{X^{\varepsilon/800}}. \quad (3.22)$$

We can find $\alpha_1, \dots, \alpha_k \in [0, 1]$ with $\alpha_1 + \cdots + \alpha_k = 1$ such that $N_i \asymp X^{\alpha_i}$ for each $i = 1, \dots, k$.

In case $k = 5$ and $\theta = 11/20$ we start by applying Lemma 2.20(iv). In case (I_2^{maj}) holds we apply Lemma 3.6 with $W = \min\{X^{\varepsilon/4}, X^{8c_k}\}$ to obtain (3.22). In case (II^{maj}) holds we wish to apply Lemma 3.5. In order to do this, we need to show that (3.7) holds with

$$v_m = \sum_{\substack{m = \prod_{i \in I} m_i \\ m_i \sim N_i}} 1 \quad (3.23)$$

and $W = \min\{X^{\varepsilon/200}, X^{20c_k}\}$ for any $L \asymp \prod_{i \in I} N_i$. Now there exists $i_0 \in I$ such that $\alpha_{i_0} \geq (2\theta - 1)/k = \frac{1}{10k}$. We have (using $d(r)d_{|I|-1}(m) \ll W^{1/100}$)

$$\begin{aligned} \left| \sum_{\ell \sim L/r} \frac{v_{\ell r} \chi(\ell)}{\ell^{1/2+it}} \right| &\leq \sum_{r=r_1 r_2} \sum_{\frac{L}{2r_2 X^{\alpha_{i_0}}} < m \leq \frac{2L}{r_2 X^{\alpha_{i_0}}}} \frac{d_{|I|-1}(m)}{m^{1/2}} \left| \sum_{\substack{m_{i_0} \sim X^{\alpha_{i_0}}/r_1 \\ m_{i_0} \sim L/(mr)}} \frac{\chi(m_{i_0})}{m_{i_0}^{1/2+it}} \right| \\ &\ll \left(\frac{L}{X^{\alpha_{i_0}}} \right)^{1/2} W^{1/100} \max_{r=r_1 r_2} \frac{1}{r_2^{1/2}} \max_{y \sim X^{\alpha_{i_0}}/r_1} \left| \sum_{X^{\alpha_{i_0}}/r_1 < m \leq y} \frac{\chi(m)}{m^{1/2+it}} \right|. \end{aligned}$$

Hence (3.7) follows for (3.23) if we show that

$$\max_{\substack{r_1 r_2 | q \\ \chi \pmod{\frac{q}{(a, q)}}}} \sup_{W \leq |t| \leq \frac{XW^4}{H_1}} \max_{y \sim X^{\alpha_{i_0}}/r_1} \left| \sum_{X^{\alpha_{i_0}}/r_1 < m \leq y} \frac{\chi(m)}{m^{1/2+it}} \right| \ll \frac{(X^{\alpha_{i_0}}/r_1)^{1/2}}{W^{1/3+1/100}}, \quad (3.24)$$

Note that $X^{\alpha_{i_0}}/r_1 \geq X^{\frac{1}{10k}-2c_k} \geq X^{\frac{1}{20k}}$. We apply Lemma 3.9(i) with $T_0 = W$. Taking $c_k \leq \delta(\frac{1}{20k})/30$ we obtain that the left-hand-side of (3.24) is

$$\ll \left(\frac{X^{\alpha_{i_0}}}{r_1} \right)^{1/2} \cdot \frac{\log X}{W^{1/2}} \ll \frac{(X^{\alpha_{i_0}}/r_1)^{1/2}}{W^{1/3+1/100}}.$$

Hence (3.22) follows from Lemma 3.5. The case $k \geq 6$ and $\theta = 7/12$ follows similarly using Lemma 2.20(iii).

A similar method allows us to establish Proposition 3.4(i). We start by applying Heath-Brown's identity (Lemma 2.16) with $L = \lceil 2/\varepsilon \rceil$, writing $N_i = X^{\alpha_i}$. Then we apply Lemma 2.20(iii) to these α_i .

In case (II^{maj}) holds we argue as above but with $W = \log^A X$ for some large $A > 0$. On the other hand, in case $\alpha_{i_0} \geq 1 - \theta - \varepsilon/2$ for some i_0 , we write $M = \frac{1}{N_{i_0}} \prod_{j=1}^{\ell} N_j$ and move the summation over $n_{i_0} \sim X^{\alpha_{i_0}}$ inside. Then it suffices to show in this case that, for any $B \geq 1$,

$$\max_{a,q \in \mathbb{N}} \sum_{M < m \leq 2^{\ell} M} d_{\ell-1}(m) \left| \frac{1}{H_1} \sum_{\substack{X/m < n_{i_0} \leq (X+H_1)/m \\ n_{i_0} \sim N_{i_0} \\ n_{i_0} m \equiv a \pmod{q}}} a_{n_{i_0}} - \frac{1}{H_2} \sum_{\substack{X/m < n_{i_0} \leq (X+H_2)/m \\ n_{i_0} \sim N_{i_0} \\ n_{i_0} m \equiv a \pmod{q}}} a_{n_{i_0}} \right| \ll \frac{1}{(\log X)^B}$$

for $a_{n_{i_0}} = \mathbf{1}_{(N_{i_0}, 2N_{i_0})}(n_{i_0})$ and $a_{n_{i_0}} = \mathbf{1}_{(N_{i_0}, 2N_{i_0})}(n_{i_0}) \log n_{i_0}$. But here $H_2/M \geq H_1/M \geq X^{\varepsilon/2}$, so the claim is easy to establish.

In the remaining case (I_2^{maj}) holds and $\alpha_i, \alpha_j > \varepsilon/2$. Thus the corresponding coefficients from Heath-Brown's identity are either $\mathbf{1}_{(N_i, 2N_i)}(n)$ or $(\log n)\mathbf{1}_{(N_i, 2N_i)}(n)$ and the claim follows from Lemma 3.6 (and partial summation if needed).

3.5. Major arc estimates with restricted prime factorization

When proving Theorem 1.1(iv)–(v) we need the following quick consequence of Theorem 3.1. One could obtain stronger results, but this is sufficient for our needs.

Corollary 3.10. *Let $X \geq 3$ and $X^{7/12+\varepsilon} \leq H \leq X^{1-\varepsilon}$ for some $\varepsilon > 0$. Let $2 \leq P < Q \leq X^{1/(\log \log X)^2}$ and write $\mathcal{P}(P, Q) = \prod_{P < p \leq Q} p$.*

(i) *For all $A > 0$,*

$$\left| \sum_{X < n \leq X+H} \mathbf{1}_{(n, \mathcal{P}(P, Q)) > 1} \mu(n) \right|^* \ll_{A, \varepsilon} \frac{H}{\log^A X} + \frac{H(\log X)^4}{P}.$$

(ii) *Let $k \geq 2$. For all $A > 0$,*

$$\left| \sum_{X < n \leq X+H} \mathbf{1}_{(n, \mathcal{P}(P, Q)) > 1} (d_k(n) - d_k^{\#}(n)) \right|^* \ll_{A, \varepsilon} \frac{H}{\log^A X} + \frac{H(\log X)^{4k}}{P}$$

Proof. Let us first show (i). By Lemma 2.19 it suffices to show that

$$\left| \sum_{\substack{X < prn \leq X+H \\ P < p \leq Q \\ r \leq X^{\varepsilon/2}}} a_r \mu(n) \right|^* \ll_{A, \varepsilon} \frac{H}{\log^A X}$$

whenever $|a_r| \leq d_2(r)$. By the triangle inequality and Theorem 3.1 the left-hand side is

$$\begin{aligned} & \ll \sum_{P < p \leq Q} \sum_{r \leq X^{\varepsilon/2}} d_2(r) \left| \sum_{X/(pr) < n \leq (X+H)/(pr)} \mu(n) \right|^* \\ & \ll_{A, \varepsilon} \sum_{P < p \leq Q} \sum_{r \leq X^{\varepsilon/2}} d_2(r) \frac{H}{pr(\log X)^{A+3}} \ll \frac{H}{\log^A X}. \end{aligned}$$

Let us now turn to (ii). By Theorem 3.1 and the triangle inequality it suffices to show the claim with $1_{(n, \mathcal{P}(P, Q)) > 1}$ replaced by $1_{(n, \mathcal{P}(P, Q)) = 1}$. Hence by Möbius inversion we need to show that

$$\left| \sum_{X < n \leq X+H} \sum_{d \mid (n, \mathcal{P}(P, Q))} \mu(d) (d_k(n) - d_k^\sharp(n)) \right|^* \ll_A \frac{H}{\log^A X}. \quad (3.25)$$

Write $D := \min\{X^{\varepsilon/2000}, X^{c_k/2}\}$. Since $d_k^\sharp(m) \ll d_k(m)$ (see (3.14)), the contribution of $d > D$ to the left-hand side of (3.25) is by Lemma 2.18 at most

$$\ll \sum_{\substack{X < dn \leq X+H \\ d > D \\ d \mid \mathcal{P}(P, Q)}} d_k(dn) \ll_A \frac{H}{\log^A X}.$$

On the other hand, the contribution of $d \leq D$ to the left-hand side of (3.25) is by the triangle inequality and Theorem 3.1

$$\begin{aligned} & \left| \sum_{X < n \leq X+H} \sum_{\substack{d \leq D \\ d \mid \mathcal{P}(P, Q)}} \mu(d) 1_{n \equiv 0 \pmod{d}} (d_k(n) - d_k^\sharp(n)) \right|^* \\ & \ll \sum_{d \leq D} \left| \sum_{X < n \leq X+H} (d_k(n) - d_k^\sharp(n)) \right|^* \ll_\varepsilon \frac{H}{X^{\varepsilon/2000}} + \frac{H}{X^{c_k/2}}. \end{aligned}$$

□

4. Reduction to type I, type II, and type I_2 estimates

To complement the major arc estimates in Theorem 3.1, we will establish later in the paper some “inverse theorems” that provide disconnection between an arithmetic function f and a nilsequence $F(g(n)\Gamma)$ assuming that f is of⁹ “type I”, “type II”, or “type I_2 ”, and the nilsequence is “minor arc” in a suitable sense. To make this precise, we give some definitions:

⁹Informally, we use type I_k to refer to expressions resembling $\alpha * d_k$ for some arithmetic function α supported on a relatively short range, with the classical type I sums corresponding to the case $k = 1$, and type II sums to refer to convolutions $\alpha * \beta$ where both α and β are supported away from 1.

Definition 4.1 (Type I , II , I_2 sums). Let $0 < \delta < 1$ and $A_I, A_{II}^-, A_{II}^+, A_{I_2} \geq 1$.

(i) (Type I sum) A (δ, A_I) type I sum is an arithmetic function of the form $f = \alpha * \beta$, where α is supported in $[1, A_I]$, and one has the bounds

$$\sum_{n \leq A} |\alpha(n)|^2 \leq \frac{1}{\delta} A \quad (4.1)$$

and

$$\|\beta\|_{\text{TV}(\mathbb{N}; q)} \leq \frac{1}{\delta} \quad (4.2)$$

for all $A \geq 1$ and some $1 \leq q \leq \frac{1}{\delta}$.

(ii) (Type II sum) A $(\delta, A_{II}^-, A_{II}^+)$ type II sum is an arithmetic function of the form $f = \alpha * \beta$, where α is supported on $[A_{II}^-, A_{II}^+]$, and one has the bound (4.1) and the bounds

$$\sum_{n \leq B} |\beta(n)|^2 \leq \frac{1}{\delta} B \quad \text{and} \quad \sum_{n \leq B} |\beta(n)|^4 \leq \frac{1}{\delta^2} B \quad (4.3)$$

for all $A, B \geq 1$. (The type II sums become vacuous if $A_{II}^- > A_{II}^+$.)

(iii) (Type I_2 sum) A (δ, A_{I_2}) type I_2 sum is an arithmetic function of the form $f = \alpha * \beta_1 * \beta_2$, where α is supported on $[1, A_{I_2}]$ and obeys the bound (4.1) for all $A \geq 1$, and β_1, β_2 obey the bound (4.2) for some $1 \leq q \leq \frac{1}{\delta}$.

We now state the inverse theorems we will establish here.

Theorem 4.2 (Inverse theorems). Let $d, D \geq 1$, $2 \leq H \leq X$, $0 < \delta < \frac{1}{\log X}$, let G/Γ be a filtered nilmanifold of degree at most d , dimension at most D , and complexity at most $1/\delta$. Let $F: G/\Gamma \rightarrow \mathbb{C}$ be Lipschitz of norm at most $1/\delta$ and mean zero. Let $f: \mathbb{N} \rightarrow \mathbb{C}$ be an arithmetic function such that

$$\left| \sum_{X < n \leq X+H} f(n) F(g(n)\Gamma) \right|^* \geq \delta H. \quad (4.4)$$

for some polynomial map $g: \mathbb{Z} \rightarrow G$.

(i) (Type I inverse theorem) If f is a (δ, A_I) type I sum for some $A_I \geq 1$, then either

$$H \ll_{d,D} \delta^{-O_{d,D}(1)} A_I$$

or else there exists a non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm $O_{d,D}(\delta^{-O_{d,D}(1)})$ such that

$$\|\eta \circ g\|_{C^\infty(X, X+H]} \ll_{d,D} \delta^{-O_{d,D}(1)}.$$

(ii) (Type II inverse theorem, non-abelian case) If f is a $(\delta, A_{II}^-, A_{II}^+)$ type II sum for some $A_{II}^+ \geq A_{II}^- \geq 1$, G is non-abelian with one-dimensional center, and F oscillates with a non-zero central frequency ξ of Lipschitz norm at most $1/\delta$, then either

$$H \ll_{d,D} \delta^{-O_{d,D}(1)} \max(A_{II}^+, X/A_{II}^-)$$

or else there exists a non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm $O_{d,D}(\delta^{-O_{d,D}(1)})$ such that

$$\|\eta \circ g\|_{C^\infty(X, X+H]} \ll_{d,D} \delta^{-O_{d,D}(1)}. \quad (4.5)$$

(iii) (Type II inverse theorem, abelian case) If f is a $(\delta, A_{II}^-, A_{II}^+)$ type II sum for some $A_{II}^+ \geq A_{II}^- \geq 1$ and $F(g(n)\Gamma) = e(P(n))$ for some polynomial $P: \mathbb{Z} \rightarrow \mathbb{R}$ of degree at most d , then either

$$H \ll_d \delta^{-O_d(1)} \max(A_{II}^+, X/A_{II}^-)$$

or else there exists a real number $T \ll_d \delta^{-O_d(1)}(X/H)^{d+1}$ such that

$$\|e(P(n))n^{-iT}\|_{\text{TV}((X, X+H] \cap \mathbb{Z}; q)} \ll_d \delta^{-O_d(1)}$$

for some $1 \leq q \ll_d \delta^{-O_d(1)}$.

(iv) (Type I₂ inverse theorem) If f is a (δ, A_{I_2}) type I₂ sum for some $A_{I_2} \geq 1$, then either

$$H \ll_{d,D} \delta^{-O_{d,D}(1)} X^{1/3} A_{I_2}^{2/3} \quad (4.6)$$

or else there exists a non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm $O_{d,D}(\delta^{-O_{d,D}(1)})$ such that

$$\|\eta \circ g\|_{C^\infty((X, X+H] \cap \mathbb{Z})} \ll_{d,D} \delta^{-O_{d,D}(1)}.$$

In this section we show how Theorem 4.2, when combined with the major arc estimates in Theorem 3.1, gives Theorem 1.1.

4.1. Combinatorial decompositions

We start by describing the combinatorial decompositions that allow us to reduce sums involving μ , Λ , d_k to type I, type II, and type I₂ sums. Lemma 4.3 will be used to prove (1.5) and (1.6), Lemma 4.4 will be used to prove (1.7), and Lemma 4.5 will be used to prove (1.8) and (1.9).

The model function Λ^\sharp is not quite a type I sum, but we can approximate it well by the type I sum¹⁰

$$\Lambda_I^\sharp(n) := \frac{P(R)}{\varphi(P(R))} \sum_{\substack{d \leq X^{\theta/2} \\ d \mid (n, P(R))}} \mu(d). \quad (4.7)$$

Indeed by (1.1), Möbius inversion and Lemma 2.18 we have

$$\sum_{X < n \leq X+H} |\Lambda_I^\sharp(n) - \Lambda^\sharp(n)| \leq \frac{P(R)}{\varphi(P(R))} \sum_{\substack{X < dn \leq X+H \\ d > X^{\theta/2} \\ d \mid P(R)}} 1 \ll H \exp(-(\log X)^{1/20}). \quad (4.8)$$

In practice, this bound allows us to substitute Λ^\sharp with the type I sum Λ_I^\sharp with negligible cost.

Lemma 4.3 (Combinatorial decompositions of μ , Λ , and Λ_I^\sharp). *Let $X^{\theta+\varepsilon} \leq H \leq X$ for $\theta = 5/8$ and some fixed $\varepsilon > 0$. For each $g \in \{\mu, \Lambda, \Lambda_I^\sharp\}$, we may find a collection \mathcal{F} of $O((\log X)^{O(1)})$ functions $f: \mathbb{N} \rightarrow \mathbb{R}$ such that*

$$g(n) = \sum_{f \in \mathcal{F}} f(n)$$

for each $X/2 \leq n \leq 4X$, and each component $f \in \mathcal{F}$ satisfies one of the following:

(i) f is a $(\log^{-O(1)} X, O(X^\theta))$ type I sum;

¹⁰One could alternatively use a type I approximant coming from the β -sieve, using the fundamental lemma of the sieve (see e.g. [37, Lemma 6.3]) but the simpler approximant Λ_I^\sharp is sufficient for us.

(ii) f is a $(\log^{-O(1)} X, O(X^{(3\theta-1)/2}))$ type I_2 sum.
 (iii) f is a $(\log^{-O(1)} X, A_{II}^-, A_{II}^+)$ type II sum for some $X^{1-\theta} \ll A_{II}^- \leq A_{II}^+ \ll X^\theta$, and it obeys the bound

$$\sup_{(X/H)(\log X)^{50A} \leq |T| \leq X^A} \left| \sum_{X < n \leq X+H} f(n) n^{iT} \right|^* \ll_A H \log^{-A} X \quad (4.9)$$

for all sufficiently large $A \geq 1$.

Lemma 4.4 (Combinatorial decompositions of d_k and d_k^\sharp). *Let $k \geq 2$. Let $X^{\theta+\varepsilon} \leq H \leq X$ for $\theta = \theta_k$ and some fixed $\varepsilon > 0$, where $\theta_2 = 1/3$, $\theta_3 = 5/9$, and $\theta_k = 5/8$ for $k \geq 4$. For each $g \in \{d_k, d_k^\sharp\}$, we may find a collection \mathcal{F} of $O((\log X)^{O(1)})$ functions $f: \mathbb{N} \rightarrow \mathbb{R}$ such that*

$$g(n) = \sum_{f \in \mathcal{F}} f(n)$$

for each $X/2 \leq n \leq 4X$, and each component $f \in \mathcal{F}$ satisfies one of the following:

(i) f is a $(\log^{-O(1)} X, O(X^\theta))$ type I sum;
 (ii) f is a $(\log^{-O(1)} X, O(X^{(3\theta-1)/2}))$ type I_2 sum.
 (iii) f is a $(\log^{-O(1)} X, A_{II}^-, A_{II}^+)$ type II sum for some $X^{1-\theta} \ll A_{II}^- \leq A_{II}^+ \ll X^\theta$ and it obeys the bound

$$\sup_{(X/H)X^{2c} \leq |T| \leq X^A} \left| \sum_{X < n \leq X+H} f(n) n^{iT} \right|^* \ll_{A,k} H X^{-c} \quad (4.10)$$

for all $A > 0$, where $c = c_{k,A} > 0$ is a sufficiently small constant.

Lemma 4.5 (Flexible combinatorial decompositions of μ , d_k , and d_k^\sharp). *Let $X^{3/5+\varepsilon} \leq H \leq X$ for some fixed $\varepsilon > 0$, let $\exp((\log x)^{2/3+\varepsilon}) \leq P \leq Q \leq X^{1/(\log \log X)^2}$, and write $\mathcal{P}(P, Q) = \prod_{P < p \leq Q} p$. We can find a collection \mathcal{F} of functions, where $|\mathcal{F}| = O((\log X)^{O(1)})$, such that for any sequence $\{\omega_n\}$ with $|\omega_n| \leq 1$,*

$$\sum_{X < n \leq X+H} 1_{(n, \mathcal{P}(P, Q)) > 1} \mu(n) \omega_n = \sum_{f \in \mathcal{F}} \sum_{X < n \leq X+H} f(n) \omega_n + O\left(\frac{H \log^4 X}{P} + \frac{H}{\exp((\log \log X)^2)}\right).$$

Moreover, each component $f \in \mathcal{F}$ satisfies one of the following:

(i) f is a $(\log^{-O(1)} X, X^{3/5+\varepsilon/10})$ type I sum;
 (ii) f is a $(\log^{-O(1)} X, X^{2/5+\varepsilon/10})$ type I_2 sum.
 (iii) f is a $(\log^{-O(1)} X, X^{2/5-\varepsilon/10}, X^{3/5+\varepsilon/10})$ type II sum and it obeys the bound

$$\sup_{(X/H)(\log X)^{20A} \leq |T| \leq X^A} \left| \sum_{X < n \leq X+H} f(n) n^{iT} \right|^* \ll_A H \log^{-A} X \quad (4.11)$$

for all sufficiently large $A > 0$.

Similarly, for fixed $k \geq 2$ we can find a collection \mathcal{F} of functions, where $|\mathcal{F}| = O((\log X)^{O(1)})$, such that for any sequence $\{\omega_n\}$ with $|\omega_n| \leq 1$,

$$\sum_{X < n \leq X+H} d_k(n) \omega_n 1_{(n, \mathcal{P}(P, Q)) > 1} = \sum_{f \in \mathcal{F}} \sum_{X < n \leq X+H} f(n) \omega_n + O\left(\frac{H \log^{4k} X}{P} + \frac{H}{\exp((\log \log X)^2)}\right).$$

Moreover, each component $f \in \mathcal{F}$ is one of (i), (ii), or (iii) above, and a similar decomposition holds also with d_k^\sharp in place of d_k .

We will prove Lemmas 4.3, 4.4 and 4.5 by first decomposing the relevant functions into certain Dirichlet convolutions (using Lemma 2.16 in the proof of Lemma 4.3 and Lemma 2.19 in the proof of Lemma 4.5). We then use Lemma 2.20 to arrange each convolution into either type I, type II, or type I_2 sums. In the case of type II sums, Lemma 2.20 also allows us to arrange them into a triple convolution for which Lemma 3.5 is applicable.

Remark 4.6. Let us briefly discuss the type II conditions such as (4.9), concentrating on the case of the von Mangoldt function.

One may observe from the proof of Theorem 1.1(ii) below that if our major arc estimate (Theorem 3.1(i)) held, for any $T \leq X^{O(1)}$, with $(\Lambda(n) - \Lambda^\sharp(n))n^{iT}$ in place of $\Lambda(n) - \Lambda^\sharp(n)$, we could prove Theorem 1.1(ii) without the need to impose in Lemma 4.3 the condition (4.9) concerning type II sums.

Unfortunately, with current knowledge, one cannot obtain such a twisted version of Theorem 3.1, at least not in the whole range $X^{7/12+\varepsilon} \leq H \leq X^{1-\varepsilon}$. However, inserting special cases of our type I and type I_2 estimates into Section 3, it would be possible to obtain such a twisted variant in the relevant range $X^{5/8+\varepsilon} \leq H \leq X^{1-\varepsilon}$. If we did this, we would not need to impose the condition (4.9). However, we found it more natural to work out the major arc estimates first using existing methods without needing to appeal to the more involved I_2 case.

Proof of Lemma 4.3. The function Λ_I^\sharp is clearly a $(\log^{-O(1)} X, O(X^\theta))$ type I sum by definition (4.7). For Λ and μ , we apply Lemma 2.16 with $L = 10$. Each component $f \in \mathcal{F}$ takes the form

$$f = a^{(1)} * \dots * a^{(\ell)} \quad (4.12)$$

for some $\ell \leq 20$, where each $a^{(i)}$ is supported on $(N_i, 2N_i]$ for some $N_i \geq 1/2$, and each $a^{(i)}(n)$ is either $1_{(N_i, 2N_i]}(n)$, $(\log n)1_{(N_i, 2N_i]}(n)$, or $\mu(n)1_{(N_i, 2N_i]}(n)$. Moreover, $N_1 N_2 \dots N_\ell \asymp X$, and $N_i \leq X^{1/10}$ for each i with $a^{(i)}(n) = \mu(n)1_{(N_i, 2N_i]}(n)$.

We can find $\alpha_1, \dots, \alpha_\ell \in [0, 1]$ with $\sum_{i=1}^\ell \alpha_i = 1$, such that $N_i \asymp X^{\alpha_i}$ for each i . If $\alpha_i > 1/10$ for some i , then $a^{(i)}(n)$ is either $1_{(N_i, 2N_i]}(n)$ or $(\log n)1_{(N_i, 2N_i]}(n)$, and hence $\|a^{(i)}\|_{\text{TV}(\mathbb{N})} \ll \log X$.

Since $\theta = 5/8 \geq 3/5$, we may apply Lemma 2.20(i), (ii) to conclude that either (I) holds, or (I₂) holds, or both (II^{min}) and (II^{maj}) hold.

First consider the case (I) holds, i.e. $\alpha_i \geq 1 - \theta$ for some i . Since $\alpha_i > 1/10$, $\|a^{(i)}\|_{\text{TV}(\mathbb{N})} \ll \log X$, and (4.12) is a $(\log^{-O(1)} X, O(X^\theta))$ type I sum of the form $\alpha * \beta$ with $\beta = a^{(i)}$ and $\alpha = a^{(1)} * \dots * a^{(i-1)} * a^{(i+1)} * \dots * a^{(k)}$.

Henceforth we may assume that $\alpha_i < 1 - \theta$ for each i . Next consider the case (I₂) holds. Then $\alpha_i + \alpha_j \geq \frac{3}{2}(1 - \theta)$ for some $i < j$. Since $\alpha_i, \alpha_j \leq 1 - \theta$, this implies that $\alpha_i, \alpha_j > 1/10$ and thus $\|a^{(i)}\|_{\text{TV}(\mathbb{N})}, \|a^{(j)}\|_{\text{TV}(\mathbb{N})} \ll \log X$. Hence (4.12) is a $(\log^{-O(1)} X, O(X^{(3\theta-1)/2}))$ type I_2 sum of the form $f = \alpha * \beta_1 * \beta_2$, with $\beta_1 = a^{(i)}$, $\beta_2 = a^{(j)}$.

Finally consider the case when both (II^{min}) and (II^{maj}) hold. Let $\{1, \dots, \ell\} = J \uplus J'$ be the partition from (II^{min}), so that $\alpha_J, \alpha_{J'} \in [1 - \theta, \theta]$. Then (4.12) is a $(\log^{-O(1)} X, A_{II}^-, A_{II}^+)$ type II sum of the form $f = \alpha * \beta$, where α (resp. β) is the convolution of those $a^{(i)}$ with $i \in J$ (resp. $i \in J'$), and $X^{1-\theta} \ll A_{II}^- \leq A_{II}^+ \ll X^\theta$.

It remains to establish the bound (4.9). For any subinterval $(X_1, X_1 + H_1] \subset (X, X + H]$, any residue class $a \pmod{q}$, any fixed $A > 0$, and any $(X/H)(\log X)^{50A} \leq |T| \leq X^A$, we need to show that

$$\left| \sum_{\substack{X_1 < n \leq X_1 + H_1 \\ n \equiv a \pmod{q}}} f(n) n^{iT} \right| \ll_A H \log^{-A} X.$$

We may assume that A is sufficiently large, $H_1 \geq H(\log X)^{-2A}$ and $q \leq (\log X)^{2A}$. Let now $\{1, \dots, \ell\} = I \uplus J \uplus J'$ be the partition from (II^{maj}) , so that

$$2\theta - 1 \leq \alpha_I \leq 4\theta - 2, \quad |\alpha_J - \alpha_{J'}| \leq 2\theta - 1.$$

Let $\{a'_{m_1}\}, \{b'_{m_2}\}, \{v'_\ell\}$ be the convolution of those $a^{(i)}$ with $i \in J, i \in J', i \in I$, respectively. Note that they are supported on $m_1 \asymp X_1^{\alpha_J}, m_2 \asymp X_1^{\alpha_{J'}}, \ell \asymp X_1^{\alpha_I}$, respectively. Thus, after dyadic division of the ranges of m_1, m_2, ℓ , we need to show that

$$\left| \sum_{\substack{X_1 < m_1 m_2 \ell < X_1 + H_1 \\ m_1 \sim M_1, m_2 \sim M_2, \ell \sim L \\ m_1 m_2 \ell \equiv a \pmod{q}}} a'_{m_1} m_1^{iT} b'_{m_2} m_2^{iT} v'_\ell \ell^{iT} \right| \ll_A H \log^{-A} X$$

for $M_1 \asymp X_1^{\alpha_J}, M_2 \asymp X_1^{\alpha_{J'}}, L \asymp X_1^{\alpha_I}$. In view of Lemma 3.5(ii) applied with $W = (\log X)^{10A}$ and $v_\ell = v'_\ell \ell^{iT}$, it suffices to verify the hypothesis (3.8). There exists $i_0 \in I$ such that $\alpha_{i_0} \geq (2\theta - 1)/20 = 1/80$. Now (3.8) follows if we show that

$$\max_{r \mid (a, q)} \max_{\chi \pmod{\frac{q}{(a, q)}}} \sup_{|t| \leq \frac{X_1 (\log X)^{40A}}{H_1}} \left| \sum_{m \asymp X^{i_0}/r} \frac{a^{(i_0)}(mr)\chi(m)}{m^{1/2+i(t-T)}} \right| \ll_A \frac{(X^{i_0}/r)^{1/2}}{(\log X)^{10A}}.$$

Since $a^{(i_0)}$ is either 1, \log , or μ on its support, this follows from Lemma 3.9 applied with $T_0 = (\log X)^{45A}$. \square

Proof of Lemma 4.4. The function d_k^\sharp is clearly a $(\log^{-O(1)} X, O(X^\theta))$ type I sum by definition (1.2). On the other hand d_k can be decomposed into a sum of $\log^k X$ terms, each of which takes the form

$$f = 1_{(N_1, 2N_1]} * \dots * 1_{(N_k, 2N_k]}$$

for some $N_i \geq 1/2$ with $N_1 N_2 \dots N_k \asymp X$. The $k \geq 4$ case of the lemma then follows in a similar way as Lemma 4.3, with the only difference being that Lemma 3.5 is now applied with $W = X^{10c}$ instead of a power of $\log X$.

In the case $k = 2$ and $\theta = 1/3$, f is clearly a $(\log^{-O(1)} X, 1)$ type I_2 sum. In the case $k = 3$ and $\theta = 5/9$, at least one of the N_i 's (say N_3) is $\ll X^{1/3}$. Hence f is a $(\log^{-O(1)} X, O(X^{1/3}))$ type I_2 sum of the form $f = \alpha * \beta_1 * \beta_2$, with $\alpha = 1_{(N_3, 2N_3]}$ and $\beta_j = 1_{(N_j, 2N_j]}(n)$ for $j = 1, 2$. \square

Proof of Lemma 4.5. Let us first outline the proof for μ . We first apply Lemma 2.19 and then Heath-Brown's identity (Lemma 2.16) with $L = 10$ to $\mu(n)$ on the right-hand side; note that we now have extra flexibility with the p variable. We obtain a collection of functions \mathcal{F} , where each $f \in \mathcal{F}$ takes the form

$$f = a^{(0)} * a^{(1)} * \dots * a^{(\ell)}$$

for some $\ell \leq 21$, where each $a^{(i)}$ is supported on $(N_i, 2N_i]$ for some $N_i \geq 1/2$, with

$$P/2 \leq N_0 \leq Q, \quad N_1 \leq X^{\varepsilon/30}, \quad N_0 N_1 \dots N_\ell \asymp X.$$

(Here $a^{(0)}$ comes from the p variable, $a^{(1)}$ comes from the r variable, and $a^{(2)} * \dots * a^{(\ell)}$ comes from applying Heath-Brown's identity to $\mu(n)$.) Moreover, $a^{(0)}(n) = 1_{n \text{ prime}} 1_{(N_0, 2N_0]}(n)$, $a^{(1)}$ is divisor-bounded, and for each $i \geq 2$, $a^{(i)}(n)$ is either $1_{(N_i, 2N_i]}(n)$ or $\mu(n) 1_{(N_i, 2N_i]}(n)$, and $N_i \leq X^{1/10}$ for each i with $a^{(i)} = \mu(n) 1_{(N_i, 2N_i]}(n)$.

We can find $\alpha_1, \dots, \alpha_\ell \in [0, 1]$ with $\sum_{i=1}^\ell \alpha_i = 1$, such that $X^{\alpha_i - \varepsilon/20} \leq N_i \ll X^{\alpha_i}$ for each $1 \leq i \leq \ell$. We may apply Lemma 2.20(ii) to conclude that either (I) holds, or (I₂) holds, or (II^{min}) holds.

As in the proof of Lemma 4.3, if (I) holds then f is a desired type I sum, if (I_2) holds then f is a desired type I_2 sum, and if (II^{\min}) holds then f is a desired type II sum. It remains to establish the bound (4.11) in the type II case. Let $\{1, \dots, \ell\} = J \uplus J'$ be the partition from (II^{\min}) , so that $|\alpha_J - \alpha_{J'}| \leq 1/5$. In view of Lemma 3.5(ii) with $W = (\log X)^{4A}$, it suffices to verify the hypothesis (3.8) for the sequence

$$v_\ell = a_\ell^{(0)} \ell^{iT} = 1_{\ell \text{ prime}} \ell^{iT}.$$

Since $N_0 \gg P$, Lemma 3.9 implies that hypothesis (3.8) is satisfied when $(\log X)^{20A} X/H \leq |T| \leq X^A$ as required.

The claim for d_k follows similarly.

In case d_k^\sharp we use Möbius inversion to write

$$\begin{aligned} \sum_{X < n \leq X+H} d_k^\sharp(n) \omega_n 1_{(n, \mathcal{P}(P, Q)) > 1} &= \sum_{X < n \leq X+H} d_k^\sharp(n) \omega_n - \sum_{X < n \leq X+H} d_k^\sharp(n) \omega_n 1_{(n, \mathcal{P}(P, Q)) = 1} \\ &= \sum_{X < n \leq X+H} d_k^\sharp(n) \omega_n - \sum_{\substack{X < dn \leq X+H \\ d \mid \mathcal{P}(P, Q)}} \mu(d) d_k^\sharp(dn) \omega_{dn}. \end{aligned}$$

Now $d_k^\sharp(n)$ is immediately a $(\log^{-O(1)} X, X^{3/5})$ type I sum by the definition (1.2). Using Lemma 2.18 we can truncate the last sum above to $d \leq X^{\varepsilon/10}$ with an admissible error $O(H/\exp((\log \log X)^2/20))$ and it remains to show that

$$f(n) = \sum_{\substack{d \mid (n, \mathcal{P}(P, Q)) \\ d \leq X^{\varepsilon/10}}} \mu(d) d_k^\sharp(dn)$$

is also a $(\log^{-O(1)} X, X^{3/5})$ type I sum. But this follows easily from the definition (1.2) of d_k^\sharp . \square

4.2. Deduction of Theorem 1.1

In this subsection we deduce Theorem 1.1 from Theorem 4.2. We focus on establishing (1.6). The other estimates in Theorem 1.1 are established similarly and we mention the small differences at the end of the section. In this section we allow all implied constants to depend on d, D .

We induct on the dimension D of G/Γ . In view of the major arc estimates (Theorem 3.1), we may assume that F has mean zero (after replacing F by $F - \int F$). In view of Proposition 2.9 with $\delta = \log^{-A} X$, we may assume that F oscillates with a central frequency $\xi: Z(G) \rightarrow \mathbb{R}$. If the center $Z(G)$ has dimension larger than 1, or ξ vanishes, then $\ker \xi$ has positive dimension and the conclusion follows from induction hypothesis applied to $G/\ker \xi$ (via Lemma 2.8). Henceforth we assume that G has one-dimensional center and that ξ is non-zero. (A zero-dimensional center is not possible since G is nilpotent and non-trivial.)

Let $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$ for $\theta = 5/8$ and $\varepsilon > 0$. Redefining δ , we see that, to prove (1.6), it suffices to show the following claim: There exists a small $c > 0$ such that for any large A and $\delta = \log^{-A} X$, if G/Γ has complexity at most δ^{-c} and F has Lipschitz norm at most δ^{-c} , then we have

$$\left| \sum_{X < n \leq X+H} (\Lambda(n) - \Lambda^\sharp(n)) \overline{F}(g(n)\Gamma) \right|^* \leq \delta H. \quad (4.13)$$

Suppose that (4.13) fails, i.e.

$$\left| \sum_{X < n \leq X+H} (\Lambda(n) - \Lambda^\sharp(n)) \overline{F}(g(n)\Gamma) \right|^* > \delta H. \quad (4.14)$$

By (4.8) and the triangle inequality, we then have

$$\left| \sum_{X < n \leq X+H} (\Lambda(n) - \Lambda_I^\sharp(n)) \bar{F}(g(n)\Gamma) \right|^* \gg \delta H. \quad (4.15)$$

By Lemma 4.3, for some component $f \in \mathcal{F}$ as in that lemma, one has the bound

$$\left| \sum_{X < n \leq X+H} f(n) \bar{F}(g(n)\Gamma) \right|^* \gg \delta^{O(1)} H. \quad (4.16)$$

Consider first the case when f is a $(\log^{-O(1)} X, A_{II}^-, A_{II}^+)$ type II sum with $X^{1-\theta} \ll A_{II}^- \leq A_{II}^+ \ll X^\theta$ obeying (4.9), and G is abelian, hence one-dimensional since $G = Z(G)$. Then we may identify G/Γ with the standard circle \mathbb{R}/\mathbb{Z} (increasing the Lipschitz constants for F , ξ by $O(\delta^{-O(1)})$ if necessary) and ξ with an element of \mathbb{Z} of magnitude $O(\delta^{O(1)})$, and we can write

$$F(x) = b e(\xi x)$$

for some $b = O(\delta^{-O(1)})$ and all $x \in \mathbb{R}/\mathbb{Z}$. We can write $\xi \cdot g(n)\Gamma = P(n) \bmod 1$ for some polynomial $P: \mathbb{Z} \rightarrow \mathbb{R}$ of degree at most d , thus by (4.14), (4.16) we have

$$\left| \sum_{X < n \leq X+H} f(n) e(-P(n)) \right|^* \geq \delta^{O(1)} H \quad (4.17)$$

and

$$\left| \sum_{X < n \leq X+H} (\Lambda(n) - \Lambda_I^\sharp(n)) e(-P(n)) \right|^* \geq \delta^{O(1)} H. \quad (4.18)$$

Theorem 4.2(iii) implies that there exists a real number $T \ll \delta^{-O(1)} (X/H)^{d+1}$ such that

$$\|e(P(n))n^{-iT}\|_{\text{TV}((X, X+H] \cap \mathbb{Z}; q)} \ll \delta^{-O(1)} \quad (4.19)$$

for some $1 \leq q \leq \delta^{-O(1)}$. By Lemma 2.2(iii), we thus obtain

$$\left| \sum_{X < n \leq X+H} f(n) n^{-iT} \right|^* \gg \delta^{O(1)} H. \quad (4.20)$$

By (4.9), we must have $|T| \leq \delta^{-O(1)} X/H$, and thus by (2.1) we have

$$\|n^{iT}\|_{\text{TV}((X, X+H] \cap \mathbb{Z}; q)} \ll \delta^{-O(1)}.$$

Hence by (4.19) and (2.2) we have

$$\|e(P(n))\|_{\text{TV}((X, X+H] \cap \mathbb{Z}; q)} \ll \delta^{-O(1)}.$$

From (4.18) and Lemma 2.2(iii), we conclude that

$$\left| \sum_{X < n \leq X+H} \Lambda(n) - \Lambda_I^\sharp(n) \right|^* \gg \delta^{O(1)} H.$$

But this contradicts the major arc estimates (Theorem 3.1(i)).

Hence in case f is a type II sum we can assume that G is non-abelian with one-dimensional center. We claim that in all the remaining cases arising from Lemma 4.3, Theorem 4.2 implies that there exists

a non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm $\delta^{-O(1)}$ such that

$$\|\eta \circ g\|_{C^\infty(X, X+H]} \gg \delta^{-O(1)}. \quad (4.21)$$

Indeed, in the case when f is a $(\log^{-O(1)} X, A_I)$ type I sum for some $A_I = O(X^\theta)$, the bound $H \ll (\log X)^{O(1)} A_I$ fails since $H \geq X^{\theta+\varepsilon}$. Hence (4.21) follows from Theorem 4.2(i).

In the case when f is a $(\log^{-O(1)} X, A_{I_2})$ type I_2 sum for some $A_{I_2} = O(X^{(3\theta-1)/2})$, the bound $H \ll (\log X)^{O(1)} X^{1/3} A_{I_2}^{2/3}$ fails since $H \geq X^{\theta+\varepsilon}$ and $X^{1/3} A_{I_2}^{2/3} = O(X^\theta)$. Hence (4.21) follows from Theorem 4.2(iv).

In the case when f is a $(\log^{-O(1)} X, A_{II}^-, A_{II}^+)$ type II sum for some $X^{1-\theta} \ll A_{II}^- \ll A_{II}^+ \ll X^\theta$, we can assume that G is non-abelian with one-dimensional center as discussed above to meet the assumption in Theorem 4.2(ii). The bound $H \ll (\log X)^{O(1)} \max(A_{II}^+, X/A_{II}^-)$ fails since $H \geq X^{\theta+\varepsilon}$ and $\max(A_{II}^+, X/A_{II}^-) \ll X^\theta$, and thus (4.21) follows from Theorem 4.2(ii).

Now that we have (4.21), we can reduce the dimension (by passing to a proper subnilmanifold) and apply the induction hypothesis to conclude the proof. By (4.21) and Lemma 2.11, we have a decomposition $g = \varepsilon g' \gamma$ for some $\varepsilon, g', \gamma \in \text{Poly}(\mathbb{Z} \rightarrow G)$ such that

- (i) ε is $(\delta^{-O(1)}, (X, X+H])$ -smooth;
- (ii) There is a $\delta^{-O(1)}$ -rational proper subnilmanifold G'/Γ' of G/Γ such that g' takes values in G' (in fact $G' = \ker \eta$); and
- (iii) γ is $\delta^{-O(1)}$ -rational.

Let $q \leq \delta^{-O(1)}$ be the period of $\gamma\Gamma$. Form a partition $(X, X+H] = P_1 \cup \dots \cup P_r$ for some $r \leq \delta^{-O(1)}$, where each P_i is an arithmetic progression of modulus q and $d_G(\varepsilon(n), \varepsilon(n')) \leq \delta^4$ whenever $n, n' \in P_i$ (which can be ensured by the smoothness of ε as long as $|P_i| \leq \delta^C H$ for some sufficiently large constant C). By the triangle inequality in Lemma 2.2(i), we have

$$\left| \sum_{X < n \leq X+H} (\Lambda - \Lambda^\#)(n) F(g(n)\Gamma) \right|^* \leq \sum_{i=1}^r \left| \sum_{n \in P_i} (\Lambda - \Lambda^\#)(n) F(g(n)\Gamma) \right|^*.$$

For each i , fix any $n_i \in P_i$, and write $\gamma(n_i)\Gamma = \gamma_i\Gamma$ for some $\gamma_i \in G$ which is rational of height $O(\delta^{-O(1)})$. Let $g_i \in \text{Poly}(\mathbb{Z} \rightarrow G)$ be the polynomial sequence defined by

$$g_i(n) = \gamma_i^{-1} g'(n) \gamma_i,$$

which takes values in $\gamma_i^{-1} G' \gamma_i$. Let $F_i: G/\Gamma \rightarrow \mathbb{C}$ be the function defined by

$$F_i(x\Gamma) = F(\varepsilon(n_i) \gamma_i x \Gamma).$$

For each $n \in P_i$ we have

$$\begin{aligned} |F(g(n)\Gamma) - F_i(g_i(n)\Gamma)| &= |F(g(n)\Gamma) - F(\varepsilon(n_i) g'(n) \gamma_i \Gamma)| \\ &\leq \|F\|_{\text{Lip}} \cdot d_G(\varepsilon(n) g'(n) \gamma_i, \varepsilon(n_i) g'(n) \gamma_i) \\ &= \|F\|_{\text{Lip}} \cdot d_G(\varepsilon(n), \varepsilon(n_i)) \leq \delta^3. \end{aligned}$$

It follows that

$$\left| \sum_{X < n \leq X+H} (\Lambda - \Lambda^\#)(n) F(g(n)\Gamma) \right|^* \leq \sum_{i=1}^r \left| \sum_{n \in P_i} (\Lambda - \Lambda^\#)(n) F_i(g_i(n)\Gamma) \right|^* + O(\delta^2 H) \quad (4.22)$$

By Lemma 2.2(i) and the induction hypothesis, we have, for each $i = 1, \dots, r$,

$$\left| \sum_{n \in P_i} (\Lambda - \Lambda^\#)(n) F_i(g_i(n)\Gamma) \right|^* \leq \left| \sum_{X < n \leq X+H} (\Lambda - \Lambda^\#)(n) F_i(g_i(n)\Gamma) \right|^* \ll \delta^C H \quad (4.23)$$

for any sufficiently large constant C . Combining this with (4.22) we obtain

$$\left| \sum_{X < n \leq X+H} (\Lambda - \Lambda^\#)(n) F(g(n)\Gamma) \right|^* \ll \delta^2 H,$$

contradicting our assumption (4.14). This completes the proof of (1.6).

The proof of (1.5) is completely similar (with the role of $\Lambda^\#$ and $\Lambda_I^\#$ both replaced by $\mu^\# = 0$). For the estimate (1.7) involving d_k , one runs the argument above with $\delta = X^{-c\epsilon}$ for some sufficiently small constant $c > 0$, using Lemma 4.4, and with the role of $\Lambda^\#$ and $\Lambda_I^\#$ both replaced by $d_k^\#$.

Let us now turn to the estimate (1.8). We choose

$$P = \exp((\log x)^{2/3+\epsilon}) \quad \text{and} \quad Q = x^{1/(\log \log x)^2} \quad (4.24)$$

and write $\mathcal{P}(P, Q) = \prod_{P < p \leq Q} p$. We first use Shiu's bound (Lemma 2.17) to note that

$$\sum_{X < n \leq X+H} \mu(n) \bar{F}(g(n)\Gamma) = \sum_{X < n \leq X+H} 1_{(n, \mathcal{P}(P, Q)) > 1} \mu(n) \bar{F}(g(n)\Gamma) + O\left(H \frac{\log P}{\log Q}\right).$$

Now one can repeat the previous arguments with $\delta = \log^{-A} X$ and $1_{(n, \mathcal{P}(P, Q)) > 1} \mu(n)$ in place of Λ and 0 in place of $\Lambda^\#$ and $\Lambda_I^\#$ — this time we use Lemma 4.5 to replace $1_{(n, \mathcal{P}(P, Q)) > 1} \mu(n)$ by the approximant $\sum_{f \in \mathcal{F}} f(n)$ and Corollary 3.10 gives the required major arc estimate for $1_{(n, \mathcal{P}(P, Q)) > 1} \mu(n)$.

The estimate (1.9) follows similarly, noting first that, with P, Q as in (4.24) we have by Shiu's bound (Lemma 2.17)

$$\begin{aligned} \sum_{X < n \leq X+H} d_k(n) \bar{F}(g(n)\Gamma) &= \sum_{X < n \leq X+H} 1_{(n, \mathcal{P}(P, Q)) > 1} d_k(n) \bar{F}(g(n)\Gamma) \\ &\quad + O\left(H(\log X)^{k-1} \left(\frac{\log P}{\log Q}\right)^k\right) \end{aligned}$$

and then arguing as for (1.8).

5. The type I case

In this section we establish the type I case (i) of Theorem 4.2, basically following the arguments in [18]. In this section we allow all implied constants to depend on d, D .

Writing $f = \alpha * \beta$, we see from Lemma 2.2(i) that

$$\left| \sum_{X < n \leq X+H} f(n) F(g(n)\Gamma) \right|^* \leq \sum_{a \leq A_I} |\alpha(a)| \left| \sum_{X/a < b \leq X/a+H/a} \beta(b) F(g(ab)\Gamma) \right|^*.$$

By the pigeonhole principle (and the hypothesis $\delta \leq \frac{1}{\log X}$), we can thus find a scale $1 \leq A \leq A_I$ such that

$$\sum_{A < a \leq 2A} |\alpha(a)| \left| \sum_{X/a < b \leq X/a+H/a} \beta(b) F(g(ab)\Gamma) \right|^* \gg \delta^{O(1)} H$$

and hence by (4.1) and the Cauchy–Schwarz inequality

$$\sum_{A < a \leq 2A} \left(\left| \sum_{X/a < b \leq X/a+H/a} \beta(b) F(g(ab)\Gamma) \right|^* \right)^2 \gg \delta^{O(1)} H^2 / A.$$

From Lemma 2.2(iii) and (4.2) we conclude that

$$\sum_{A < a \leq 2A} \left(\left| \sum_{X/a < b \leq X/a+H/a} F(g(ab)\Gamma) \right|^* \right)^2 \gg \delta^{O(1)} H^2 / A. \quad (5.1)$$

We may assume that $H \geq C\delta^{-C}A$ for some large constant C depending on d, D , since otherwise we have $H \leq \delta^{-O(1)}A_I$ and can conclude. Trivially

$$\left| \sum_{X/a < b \leq X/a+H/a} F(g(ab)\Gamma) \right|^* \ll \delta^{-1} H / A$$

for all $A < a \leq 2A$, and hence by (5.1) we must have

$$\left| \sum_{X/a < b \leq X/a+H/a} F(g(ab)\Gamma) \right|^* \gg \delta^{O(1)} H / A$$

for $\gg \delta^{O(1)}A$ choices of $a \in (A, 2A]$. For each such a , we apply Theorem 2.7 to find a non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm $O(\delta^{-O(1)})$ such that

$$\|\eta \circ g(a \cdot)\|_{C^\infty(X/a, X/a+H/a)} \ll \delta^{-O(1)}. \quad (5.2)$$

This character η could initially depend on a , but the number of possible choices for η is $O(\delta^{-O(1)})$, hence by the pigeonhole principle we may refine the set of a under consideration to make η independent of a . The function $\eta \circ g: \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ is a polynomial of degree at most d , hence by Corollary 2.4 (and the assumption $H \geq C\delta^{-C}A$) we have

$$\|q\eta \circ g\|_{C^\infty(X, X+H)} \ll \delta^{-O(1)}$$

for some $1 \leq q \ll \delta^{-O(1)}$. Replacing η by $q\eta$, we obtain Theorem 4.2(i) as required.

Remark 5.1. It should also be possible to establish Theorem 4.2(i) using the variant of Theorem 2.12 given in [26, Theorem 3.6].

6. The non-abelian type II case

In this section we establish the non-abelian type II case (ii) of Theorem 4.2. Let $d, D, H, X, \delta, G/\Gamma, F, f, A_{II}^-, A_{II}^+$ be as in that theorem. For the rest of this section we allow all constants

to depend on d, D . We will need several constants

$$1 < C_1 < C_2 < C_3 < C_4$$

depending on d, D , with each C_i assumed to be sufficiently large depending on the preceding constants.

We first eliminate the role of α by a standard Cauchy–Schwarz argument. By Definition 4.1(ii), we can write $f = \alpha * \beta$, where α is supported on $[A_{II}^-, A_{II}^+]$, and one has the bounds (4.1), (4.3) for all $A, B \geq 1$. From (4.4) we have

$$\left| \sum_{n \in P} \alpha * \beta(n) F(g(n)\Gamma) \right| \geq \delta H$$

for some arithmetic progression $P \subset (X, X + H]$. By the triangle inequality, we have

$$\left| \sum_{n \in P} (\alpha * \beta)(n) F(g(n)\Gamma) \right| \leq \sum_{A_{II}^- \leq a \leq A_{II}^+} |\alpha(a)| \left| \sum_{b: ab \in P} \beta(b) F(g(ab)\Gamma) \right|.$$

By the pigeonhole principle and the hypothesis $\delta \leq \frac{1}{\log X}$, one can thus find $A_{II}^- \leq A \leq A_{II}^+$ such that

$$\sum_{A < a \leq 2A} |\alpha(a)| \left| \sum_{b: ab \in P} \beta(b) F(g(ab)\Gamma) \right| \gg \delta^{O(1)} H. \quad (6.1)$$

We may assume that

$$\delta^{-C_4} \frac{X}{H} \leq A \leq \delta^{C_4} H, \quad (6.2)$$

since otherwise the first conclusion of Theorem 4.2(ii) holds. Now by (6.1), the Cauchy–Schwarz inequality, and (4.1)

$$\sum_{A < a \leq 2A} \left| \sum_{b: ab \in P} \beta(b) F(g(ab)\Gamma) \right|^2 \gg \delta^{O(1)} \frac{H^2}{A}. \quad (6.3)$$

Next, we dispose of the large values of β . Namely, we now show that the contribution of those b for which $|\beta(b)| > \delta^{-C_2}$ to the left-hand side is negligible. They contribute

$$\begin{aligned} & \ll \delta^{-2} \sum_{A < a \leq 2A} \left(\sum_{b: ab \in P} 1_{|\beta(b)| > \delta^{-C_2}} |\beta(b)| \right)^2 \ll \delta^{2C_2-2} \sum_{A < a \leq 2A} \left(\sum_{b: ab \in P} |\beta(b)|^2 \right)^2 \\ & \ll \delta^{2C_2-2} \sum_{b_1, b_2} |\beta(b_1)|^2 |\beta(b_2)|^2 \sum_{\substack{A < a \leq 2A \\ ab_1, ab_2 \in P}} 1 \end{aligned} \quad (6.4)$$

Since $P \subseteq (X, X + H]$, the inner sum can be non-empty only if $b_j \asymp X/A$ and $|b_1 - b_2| \leq H/A$ and in this case it has size $\ll H/(X/A) = AH/X$. Using also the inequality $|xy|^2 \leq |x|^4 + |y|^4$ and (4.3), we see that (6.4) is

$$\ll \delta^{2C_2-2} \sum_{b_1 \asymp X/A} |\beta(b_1)|^4 \sum_{\substack{b_2 \\ |b_1 - b_2| \leq H/A}} \frac{AH}{X} \ll \delta^{2C_2-4} \frac{H^2}{A}.$$

From now on in this section we allow all implied constants to depend on C_2 . Write

$$\tilde{\beta}(b) := \beta(b) 1_{|\beta(b)| \leq \delta^{-C_2}} = O(\delta^{-O(1)}).$$

By above and the triangle inequality, (6.3) holds with $\tilde{\beta}(b)$ in place of $\beta(b)$. Hence, by Markov's inequality, we see that, for C_2 large enough, we have

$$\left| \sum_{X/a < b \leq (X+H)/a} \tilde{\beta}(b) F(g(ab)\Gamma) \right|^* \gg \delta^{O(1)} H/A \quad (6.5)$$

for $\gg \delta^{O(1)} A$ choices of $a \in (A, 2A]$. We cover $(A, 2A]$ by $O(X/H)$ boundedly overlapping intervals of the form $I_{A'} := (A', (1 + \frac{H}{X})A']$ with $A \leq A' \leq 2A$. Note that these intervals are non-empty by the lower bound on A in (6.2). By the pigeonhole principle, we see that for $\gg \delta^{O(1)} X/H$ of these intervals, (6.5) holds for $\gg \delta^{O(1)} \frac{H}{X} A$ choices of $a \in I_{A'}$. For all such A' and a , the interval $(X/a, (X+H)/a]$ is contained in

$$J_{A'} := \left(\left(1 - \frac{10H}{X} \right) \frac{X}{A'}, \left(1 + \frac{10H}{X} \right) \frac{X}{A'} \right], \quad (6.6)$$

hence

$$\left| \sum_{b \in J_{A'}} \tilde{\beta}(b) F(g(ab)\Gamma) \right|^* \gg \delta^{O(1)} H/A$$

for $\gg \delta^{O(1)} \frac{H}{X} A$ choices of $a \in I_{A'}$. We can now apply Proposition 2.15 and the pigeonhole principle to reach one of two conclusions for $\gg \delta^{O(1)} X/H$ of the intervals $I_{A'}$:

- (i) There exists a non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm $O(\delta^{-O(1)})$ such that $\|\eta \circ g(a \cdot)\|_{C^\infty(J_{A'})} \ll \delta^{-O(1)}$ for $\gg \delta^{O(1)} |I_{A'}|$ values of $a \in I_{A'}$.
- (ii) For $\gg \delta^{O(1)} |I_{A'}|^2$ pairs $(a, a') \in I_{A'}^2$, there exists a factorization

$$g(a' \cdot) = \varepsilon_{aa'} g(a \cdot) \gamma_{aa'} \quad (6.7)$$

where $\varepsilon_{aa'}$ is $(O(\delta^{-O(1)}), J_{A'})$ -smooth and $\gamma_{aa'}$ is $O(\delta^{-O(1)})$ -rational.

Suppose first that conclusion (i) holds for $\gg \delta^{O(1)} X/H$ of the intervals $I_{A'}$. By pigeonholing we may make η independent of A' , and then by collecting all the a we see that

$$\|\eta \circ g(a \cdot)\|_{C^\infty((X/a, (X+H)/a])} \ll \delta^{-O(1)}$$

for $\gg \delta^{O(1)} A$ values of a with $a \asymp A$. Applying Corollary 2.4, we see that either $H \ll \delta^{-O(1)} A$, or else there is another non-trivial horizontal character $\eta': G \rightarrow \mathbb{R}/\mathbb{Z}$ of Lipschitz norm $O(\delta^{-O(1)})$ such that

$$\|\eta' \circ g\|_{C^\infty((X, X+H])} \ll \delta^{-O(1)}.$$

In either case the conclusion of Theorem 4.2(ii) is satisfied.

Now suppose that conclusion (ii) holds for some A' which we now fix (discarding the information collected for all other choices of A'). We will formalize the argument that follows as a proposition, as we will need this precise proposition also in our followup work [46].

Proposition 6.1 (Abstract non-abelian Type II inverse theorem). *Let $C \geq 1$, $d, D \geq 1$, $2 \leq H, A \leq X$, $0 < \delta < \frac{1}{\log X}$, and let G/Γ be a filtered nilmanifold of degree at most d , dimension at most D , and complexity at most $1/\delta$, with G non-abelian. Let $g: \mathbb{Z} \rightarrow G$ be a polynomial map. Cover $(A, 2A]$ by at most CX/H intervals $I_{A'} = (A', (1 + \frac{H}{X})A')$ with $A \leq A' \leq 2A$, with each point belonging to at most C of these intervals. Suppose that for at least $\frac{1}{C} \delta^C X/H$ of the intervals $I_{A'}$, there exist at least $\frac{1}{C} \delta^C |I_{A'}|^2$ pairs $(a, a') \in I_{A'}^2$ for which there exists a factorization*

$$g(a' \cdot) = \varepsilon_{aa'} g(a \cdot) \gamma_{aa'}$$

where $\varepsilon_{aa'}$ is $(C\delta^{-C}, J_{A'})$ -smooth and $\gamma_{aa'}$ is $C\delta^{-C}$ -rational, with $J_{A'}$ defined by (6.6).

Then either

$$H \ll_{d,D,C} \delta^{-O_{d,D,C}(1)} \max(A, X/A) \quad (6.8)$$

or there exists a non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}/\mathbb{Z}$ having Lipschitz norm $O_{d,D,C}(\delta^{-O_{d,D,C}(1)})$ such that

$$\|\eta \circ g\|_{C^\infty(X, X+H]} \ll_{d,D,C} \delta^{-O_{d,D,C}(1)}.$$

Indeed, applying this proposition (with a suitable choice of $C = O(1)$, and the other parameters given their obvious values), the conclusion (6.8) is not compatible with (6.2) for C_4 large enough, so we obtain the desired conclusion (4.5).

It remains to establish the proposition. We allow all implied constants to depend on d, D, C . We will now proceed by analyzing the equidistribution properties of the four-parameter polynomial map

$$(a, b, a', b') \mapsto (g(ab), g(ab'), g(a'b), g(a'b')).$$

The one-parameter equidistribution theorem in Theorem 2.12 is not directly applicable for this purpose. Fortunately, we may apply the multi-parameter equidistribution theory in Theorem 2.13 instead. We conclude that either

$$\min(|I_{A'}|, |J_{A'}|) \ll_{C_3} \delta^{-O_{C_3}(1)}, \quad (6.9)$$

or else there exists

$$\delta^{-C_3} \leq M \ll \delta^{-O_{C_3}(1)} \quad (6.10)$$

and a factorization

$$(g(ab), g(ab'), g(a'b), g(a'b')) = \varepsilon(a, a', b, b')g'(a, a', b, b')\gamma(a, a', b, b') \quad (6.11)$$

where $\varepsilon, \tilde{g}, \gamma \in \text{Poly}(\mathbb{Z}^4 \rightarrow G^4)$ are such that

(i) (ε smooth) For all $(a, a', b, b') \in I_{A'} \times I_{A'} \times J_{A'} \times J_{A'}$, we have the smoothness estimates

$$\begin{aligned} d_G(\varepsilon(a, a', b, b'), 1) &\leq M \\ d_G(\varepsilon(a+1, a', b, b'), \varepsilon(a, a', b, b')) &\leq M/|I_{A'}| \\ d_G(\varepsilon(a, a'+1, b, b'), \varepsilon(a, a', b, b')) &\leq M/|I_{A'}| \\ d_G(\varepsilon(a, a', b+1, b'), \varepsilon(a, a', b, b')) &\leq M/|J_{A'}| \\ d_G(\varepsilon(a, a', b, b'+1), \varepsilon(a, a', b, b')) &\leq M/|J_{A'}|. \end{aligned}$$

(ii) (g' equidistributed) There is an M -rational subnilmanifold G'/Γ' of G^4/Γ^4 such that g' takes values in G' and one has the total equidistribution property

$$\left| \sum_{(a, a', b, b') \in P_1 \times P_2 \times P_3 \times P_4} F(g'(a, a', b, b')\Gamma'') \right| \leq \frac{|I_{A'}|^2 |J_{A'}|^2}{M^{C_3^2}} \|F\|_{\text{Lip}}$$

for any arithmetic progressions $P_1, P_2 \subset I_{A'}, P_3, P_4 \subset J_{A'}$, any finite index subgroup Γ'' of Γ' of index at most $M^{C_3^2}$, and any Lipschitz function $F: G'/\Gamma'' \rightarrow \mathbb{C}$ of mean zero.

(iii) (γ rational) There exists $1 \leq r \leq M$ such that $\gamma^r(a, a', b, b') \in \Gamma^4$ for all $a, a', b, b' \in \mathbb{Z}$.

The alternative (6.9) of course implies (6.8), so we may assume we are in the opposite alternative. Thus we may assume that we have a scale M and a factorization (6.11) with the claimed properties.

We know that (6.7) holds for $\gg M^{-O(1)}|I_{A'}|^2$ pairs $(a, a') \in I_{A'}^2$. By pigeonholing we may assume there is a fixed $1 \leq r \ll M^{O(1)}$ such that $\gamma_{aa'}(b)^r \in \Gamma$ for all such pairs (a, a') and all b , and also such that $\gamma^r(a, a', b', b') \in \tilde{\Gamma}^4$. This implies that there is some lattice $\tilde{\Gamma}$ independent of a, a' that contains Γ as an index $O(\delta^{-O(1)})$ subgroup, such that $\gamma_{aa'}(b) \in \tilde{\Gamma}$ for all such pairs (a, a') , and $\gamma(a, a', b, b') \in \tilde{\Gamma}^4$; indeed, by [19, Lemma A.8(i), Lemma A.11(iii)], we could take $\tilde{\Gamma}$ to be generated by $\exp(\frac{1}{Q'}X_i)$ for the Mal'cev basis X_1, \dots, X_D of G/Γ , and some $Q' \ll M^{O(1)}$. From (6.7) we then have

$$g(a'b)\tilde{\Gamma} = \varepsilon_{aa'}(b)g(ab)\tilde{\Gamma}$$

for all such pairs (a, a') and all $b \in \mathbb{Z}$. If we introduce the subinterval

$$J'_{A'} := \left(\frac{X}{A'}, \left(1 + \frac{1}{M^{C_3}} \frac{H}{X} \right) \frac{X}{A'} \right]$$

of $J_{A'}$, then from the smoothness of $\varepsilon_{aa'}$ we have

$$\varepsilon_{aa'}(b') = O_G(M^{-C_3+O(1)})\varepsilon_{aa'}(b) = O_G(M^{O(1)})$$

whenever $b, b' \in J'_{A'}$, where $O_G(r)$ denotes an element of G at a distance $O(r)$ from the identity. This implies that

$$(g(ab)\tilde{\Gamma}, g(ab')\tilde{\Gamma}, g(a'b)\tilde{\Gamma}, g(a'b')\tilde{\Gamma}) \in \Omega$$

where $\Omega \subset (G/\tilde{\Gamma})^4$ consists of all quadruples of the form

$$(x, y, \varepsilon x, \kappa \varepsilon y) \tag{6.12}$$

for some $x, y \in G/\Gamma$ and $\varepsilon, \kappa \in G$ with $\varepsilon = O_G(M^{O(1)})$ and $\kappa = O_G(M^{-C_3+O(1)})$ (with appropriate choices of implied constants). We conclude that

$$\sum_{a, a' \in I_{A'}; b, b' \in J'_{A'}} 1_{\Omega}(g(ab)\tilde{\Gamma}, g(ab')\tilde{\Gamma}, g(a'b)\tilde{\Gamma}, g(a'b')\tilde{\Gamma}) \gg M^{-O(1)}|I_{A'}|^2|J'_{A'}|^2.$$

Applying (6.11), we conclude that

$$\sum_{a, a' \in I_{A'}; b, b' \in J'_{A'}} 1_{\Omega}(\varepsilon(a, a', b, b')g'(a, a', b, b')\tilde{\Gamma}^4) \gg M^{-O(1)}|I_{A'}|^2|J'_{A'}|^2.$$

By the pigeonhole principle, we can find intervals $I'_{A'}, I''_{A'}$ in $I_{A'}$ of length $M^{-C_3}I_{A'}$ such that

$$\sum_{a \in I'_{A'}, a' \in I''_{A'}; b, b' \in J'_{A'}} 1_{\Omega}(\varepsilon(a, a', b, b')g'(a, a', b, b')\tilde{\Gamma}^4) \gg M^{-O(1)}|I'_{A'}||I''_{A'}||J'_{A'}|^2.$$

By the smoothness of ε we have

$$\varepsilon(a, a', b, b') = O_G(M^{-C_3+O(1)})\varepsilon(a_0, a'_0, b_0, b_0) = O_G(M^{O(1)})$$

where a_0, a'_0, b_0 are the left endpoints of $I'_{A'}, I''_{A'}, J'_{A'}$ respectively. Let φ be a bump function¹¹ supported on $\tilde{\Omega}$ that equals 1 on Ω , with Lipschitz norm $O(M^{O(C_3)})$, where $\tilde{\Omega}$ is defined similarly to Ω in (6.12) but with slightly larger choices of implied constants $O(1)$ in the definition of ε, κ . This implies that

$$1_{\Omega}(\varepsilon(a, a', b, b')g'(a, a', b, b')\tilde{\Gamma}^4) \leq \varphi(\varepsilon(a_0, a'_0, b_0, b_0))g'(a, a', b, b')\tilde{\Gamma}^4$$

¹¹Indeed, one could set $\varphi(x) = \max(1 - K\text{dist}(x, \Omega), 0)$ for some $K = O(M^{O(C_3)})$.

whenever $a \in I'_{A'}$, $a' \in I''_{A'}$; $b, b' \in J'_{A'}$. Abbreviating $\varepsilon_0 := \varepsilon(a_0, a'_0, b_0, b_0) = O_G(M^{O(1)})$, we conclude that

$$\sum_{a \in I'_{A'}, a' \in I''_{A'}; b, b' \in J'_{A'}} \varphi(\varepsilon_0 g'(a, a', b, b') \tilde{\Gamma}^4) \gg M^{-O(1)} |I'_{A'}| |I''_{A'}| |J'_{A'}|^2.$$

Using the equidistribution properties of g' , we conclude that

$$\int_{G'/(G' \cap \tilde{\Gamma}^4)} \varphi(\varepsilon_0 x) d\mu_{G'/(G' \cap \tilde{\Gamma}^4)} \gg M^{-O(1)}. \quad (6.13)$$

We now use this bound to obtain control on the group G' . Let us introduce the slice

$$L := \{g \in G : (1, 1, 1, g) \in G'\}. \quad (6.14)$$

This is a $O(M^{O(1)})$ -rational subgroup of G . Suppose first that this group is non-trivial, then $L \cap \Gamma'$ contains a non-trivial element $\gamma = O_G(M^{O(1)})$. For $0 \leq t \leq 1$, the group element $\gamma^t := \exp(t \log \gamma) = O_G(M^{O(1)})$ is such that $(1, 1, 1, \gamma^t)$ lies in G' , and hence from (6.13) and invariance of Haar measure we have

$$\int_{G'/(G' \cap \tilde{\Gamma}^4)} \varphi(\varepsilon_0(1, 1, 1, \gamma^t)x) d\mu_{G'/(G' \cap \tilde{\Gamma}^4)} \gg M^{-O(1)}.$$

Integrating this and using the Fubini–Tonelli theorem, we have

$$\int_{G'/(G' \cap \tilde{\Gamma}^4)} \int_0^1 \varphi(\varepsilon_0(1, 1, 1, \gamma^t)x) dt d\mu_{G'/(G' \cap \tilde{\Gamma}^4)} \gg M^{-O(1)}.$$

and thus by the pigeonhole principle there exists $x \in (G/\Gamma)^4$ such that

$$\int_0^1 \varphi(\varepsilon_0(1, 1, 1, \gamma^t)x) dt \gg M^{-O(1)}.$$

In particular, we have

$$\varepsilon_0(1, 1, 1, \gamma^t)x \in \tilde{\Omega} \subset (G/\Gamma)^4 \quad (6.15)$$

for a set of $t \in [0, 1]$ of measure $\gg M^{-O(1)}$. But if we let x_1, x_2, x_3 be the first three components of $\varepsilon_0 x$, we see from (6.12) that in order for (6.15) to hold, the fourth coordinate of $\varepsilon_0(1, 1, 1, \gamma^t)x$ must take the form $\kappa \varepsilon x_2$, where $\varepsilon = O(M^{O(1)})$ is such that $x_3 = \varepsilon x_1$. Since the equation $x_3 = \varepsilon x_1$ fixes ε to a double coset of $\tilde{\Gamma}$, there are at most $O(M^{O(1)})$ choices for ε , and for each such choice, $\kappa \varepsilon x_2$ is confined to a ball of radius $O(M^{-C_3+O(1)})$; thus the fourth coordinate of $\varepsilon_0(1, 1, 1, \gamma^t)x$ is confined to the union of $O(M^{O(1)})$ balls of radius $O(M^{-C_3+O(1)})$. Since γ is non-trivial, $t \in [0, 1]$ is thus confined to the union of $O(M^{O(1)})$ intervals of radius $O(M^{-C_3+O(1)})$. Thus the set of $t \in [0, 1]$ obeying (6.15) has measure at most $O(M^{-C_3+O(1)})$, leading to a contradiction for C_3 large enough. Thus L must be trivial.

Now we apply a ‘‘Furstenberg–Weiss’’ argument [15] (see also the argument attributed to Serre in [57, Lemma 3.3]). Consider the groups

$$\begin{aligned} L_1 &:= \{g \in G : (1, g', 1, g) \in G' \text{ for some } g' \in G\} \\ L_2 &:= \{g \in G : (1, 1, g', g) \in G' \text{ for some } g' \in G\}. \end{aligned}$$

Taking logarithms, we have

$$\begin{aligned} \log L_1 &:= \{X \in \log G : (X, X', 0, X) \in \log G' \text{ for some } X' \in \log G\} \\ \log L_2 &:= \{X \in \log G : (0, 0, X', X) \in \log G' \text{ for some } X' \in \log G\}, \end{aligned}$$

thus $\log L_1, \log L_2$ are projections of certain slices of $\log G'$. Since G' was a $O(M^{O(1)})$ -rational subgroup of G^4 , we conclude from linear algebra that L_1, L_2 are $O(M^{O(1)})$ -rational subgroups of G ; comparing with (6.14), we also see that $[L_1, L_2] \subset L$; since L is trivial, $[L_1, L_2]$ is trivial. Since G is non-abelian by hypothesis, $[G, G]$ is non-trivial; thus at least one of L_1, L_2 must be a proper subgroup of G . For sake of discussion let us assume that L_1 is a proper subgroup, as the other case is similar. Then there exists a non-trivial horizontal character $\eta_4: G \rightarrow \mathbb{R}/\mathbb{Z}$ on $G/\tilde{\Gamma}$ of Lipschitz norm $O(M^{O(1)})$ that annihilates L_1 , that is to say $\eta_4(g) = 0$ whenever $(1, g', 1, g) \in G'$ for some $g' \in G$. Thus, the homomorphism $(1, g', 1, g) \mapsto \eta_4(g)$ on $1 \times G \times 1 \times G$ annihilates the restriction of G' to this group, as well as $1 \times G \times 1 \times 1$. Taking logarithms, we obtain a linear functional on the Lie algebra $0 \times \log G \times 0 \times \log G$ (with all coefficients $O(M^{O(1)})$ in the Mal'cev basis) that annihilates the restriction of $\log G'$ to this Lie algebra, as well as to $0 \times \log G \times 0 \times 0$; by composing with a suitable linear projection we can then extend this linear functional to a linear functional on all of $(\log G)^4$ that annihilates all of $\log G'$, again with all coefficients $O(M^{O(1)})$. Undoing the logarithm, we may find (possibly trivial) additional horizontal characters $\eta_1, \eta_3: G \rightarrow \mathbb{R}/\mathbb{Z}$ on $G/\tilde{\Gamma}$ of Lipschitz norm $O(M^{O(1)})$ such that

$$\eta_1(g_1) + \eta_3(g_3) + \eta_4(g_4) = 0$$

for all $(g_1, g_2, g_3, g_4) \in G'$. In particular, writing $g' = (g'_1, g'_2, g'_3, g'_4)$, we have

$$\eta_1(g'_1(a, a', b, b')) + \eta_3(g'_3(a, a', b, b')) + \eta_4(g'_4(a, a', b, b')) = 0$$

for all $a, a', b, b' \in \mathbb{Z}$. Applying the factorization (6.11), and noting that the horizontal characters η_1, η_3, η_4 annihilate the components of γ , we conclude that

$$\eta_1(g(ab)) + \eta_3(g(a'b)) + \eta_4(g(a'b')) = \tilde{\varepsilon}(a, a', b, b') \quad (6.16)$$

for all $a, a', b, b' \in \mathbb{Z}$, where

$$\tilde{\varepsilon}(a, a', b, b') := \eta_1(\varepsilon_1(a, a', b, b')) + \eta_3(\varepsilon_3(a, a', b, b')) + \eta_4(\varepsilon_4(a, a', b, b'))$$

and $\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4$ are the components of ε . From the smoothness properties of ε , we see in particular that

$$\|\tilde{\varepsilon}(a, a', b, b' + 1) - \tilde{\varepsilon}(a, a', b, b')\|_{\mathbb{R}/\mathbb{Z}} \ll M^{O(1)} / |J_{A'}|$$

for $a, a' \in I_{A'}$, $b, b' \in J_{A'}$, and hence from (6.16)

$$\|\eta_4(g(a'(b' + 1))) - \eta_4(g(a'b'))\|_{\mathbb{R}/\mathbb{Z}} \ll M^{O(1)} / |J_{A'}|$$

whenever $a' \in I_{A'}$, $b' \in J_{A'}$. For any $a' \in I_{A'}$, the map $b' \mapsto \eta_4(g(a'b'))$ is a polynomial of degree at most d , so by Vinogradov's lemma (Lemma 2.3), for each such a' , we either have

$$|J_{A'}| \ll M^{O(1)},$$

or else there exists $1 \leq q \ll M^{O(1)}$ such that

$$\|q\eta_4(g(a' \cdot))\|_{C^\infty(J_{A'})} \ll M^{O(1)}. \quad (6.17)$$

The former possibility is not compatible with (6.2) if C_4 is large enough, so we may assume the latter possibility (6.17) holds for all $a' \in I_{A'}$. Currently the quantity q may depend on a' , but by the pigeonhole principle we may fix a q so that (6.17) holds for $\gg M^{-O(1)} |I_{A'}|$ choices of $a' \in I_{A'}$. Applying Corollary 2.4, we conclude that either

$$|I_{A'}| \ll M^{O(1)},$$

or else there exists $1 \leq q' \ll M^{O(1)}$ such that

$$\|q' \eta_4 \circ g\|_{C^\infty([X, X+H])} \ll M^{O(1)}.$$

In either case we obtain one of the conclusions of Proposition 6.1. The proof of Theorem 4.2(ii) is now complete.

7. The abelian type II case

In this section we establish the abelian Type II case (iii) of Theorem 4.2 using arguments from [49]. We shall need the following variant of [49, Proposition 2.2].

Proposition 7.1. *Let $\delta \in (0, 1/2)$, $M \geq 2$ and $L = X/M$. Assume that $H \geq \delta^{-C} \max(L, M)$ for some sufficiently large constant $C = C(k) > 0$. Let $\alpha(\ell), \beta(m) \in \mathbb{C}$. Let $k \in \mathbb{N}$ and let*

$$g(n) = \sum_{j=1}^k \nu_j (n - X)^j$$

be a polynomial of degree k with real coefficients ν_j . If

$$\left| \sum_{\substack{\ell, m \\ m \sim M \\ X < \ell m \leq X+H}} \alpha(\ell) \beta(m) e(g(\ell m)) \right| \geq \delta H \left(\frac{1}{L} \sum_{L/2 < \ell \leq 2L} |\alpha(\ell)|^2 \right)^{1/2} \left(\frac{1}{M} \sum_{m \sim M} |\beta(m)|^4 \right)^{1/4},$$

then there exists a positive integer $q \leq \delta^{-O_k(1)}$ such that

$$\|q(j\nu_j + (j+1)X\nu_{j+1})\|_{\mathbb{R}/\mathbb{Z}} \leq \delta^{-O_k(1)} \frac{X}{H^{j+1}}$$

for all $1 \leq j \leq k$, with the convention that $\nu_{k+1} = 0$.

Proof. This follows from the same argument as [49, Proposition 2.2]. The only difference is that we do not assume that the coefficients $\alpha(\ell)$ and $\beta(m)$ are divisor bounded and due to this in the beginning of the proof we do not estimate the sums $\sum_{L/2 < \ell \leq 2L} |\alpha(\ell)|^2$ and $\sum_{m \sim M} |\beta(m)|^4$ with bounds for averages of divisor functions but keep them as they are. \square

Let us get back to the proof of Theorem 4.2(iii). We can assume that

$$\max\{A_{II}^+, X/A_{II}^-\} \ll \delta^{O_d(1)} H$$

since otherwise the claim is immediate. Note that in particular $H \geq \delta^{-O_d(1)} X^{1/2}$. By assumption and dyadic splitting (noting that $\delta < 1/\log X$)

$$\left| \sum_{\substack{x < \ell m \leq x+h \\ m \sim M \\ \ell m \equiv u \pmod{v}}} \alpha(\ell) \beta(m) e(P(\ell m)) \right| \geq \delta^2 H \tag{7.1}$$

for some $(x, x+h] \subseteq (X, X+H]$, some $M \in [X/A_{II}^+, X/A_{II}^-]$, some polynomial $P(x)$ of degree at most d and some $u, v \in \mathbb{N}$ with $u \leq v$. Before applying Proposition 7.1 we will show that (7.1) can hold only if $v \ll \delta^{-8}$ and $h \gg \delta^8 H$. In order to show this, we give an upper bound for the left-hand side using the

Cauchy–Schwarz inequality. Using also (4.1) and denoting $L = X/M$, we obtain, using the inequality $|xy| \leq |x|^2 + |y|^2$

$$\begin{aligned} \delta^4 H^2 &\leq \left| \sum_{\substack{x < \ell m \leq x+h \\ m \sim M \\ \ell m \equiv u \pmod{v}}} \alpha(\ell) \beta(m) e(P(\ell m)) \right|^2 \\ &\ll \sum_{L/2 < \ell \leq 2L} |\alpha(\ell)|^2 \cdot \sum_{L/2 < \ell \leq 2L} \left(\sum_{\substack{m \sim M \\ x < \ell m \leq x+h \\ \ell m \equiv u \pmod{v}}} |\beta(m)| \right)^2 \\ &\ll \frac{L}{\delta} \sum_{\substack{m_1, m_2 \sim M \\ |m_1 - m_2| \leq 2h/L \\ (m_j, v) \mid u}} |\beta(m_1) \beta(m_2)| \sum_{\substack{L/2 < \ell \leq 2L \\ x < \ell m_1, \ell m_2 < x+h \\ \ell m_j \equiv u \pmod{v}}} 1 \\ &\ll \frac{L}{\delta} \sum_{\substack{m_1, m_2 \sim M \\ |m_1 - m_2| \leq 2h/L \\ (m_2, v) \mid u}} |\beta(m_1)|^2 \left(1 + \frac{h(m_2, v)}{Mv} \right). \end{aligned}$$

Writing $d = (m_2, v)$ and $m'_2 = m_2/d$ and using (4.3) we obtain

$$\begin{aligned} \delta^4 H^2 &\ll \frac{L}{\delta} \sum_{m_1 \sim M} |\beta(m_1)|^2 \left(\frac{h}{L} + 1 + \sum_{d \mid u} \sum_{\substack{m'_2 \\ |m_1 - dm'_2| \leq 2h/L}} \frac{hd}{Mv} \right) \\ &\ll \frac{hM}{\delta^2} + \frac{LM}{\delta^2} + \frac{LM}{\delta^2} \sum_{d \mid u} \frac{hd}{Mv} \left(\frac{h}{Ld} + 1 \right) \\ &\ll \frac{hM}{\delta^2} + \frac{LM}{\delta^2} + \frac{h^2 d_2(u)}{v\delta^2} + \frac{hL}{\delta^2 v} \cdot \frac{u^2}{\varphi(u)}. \end{aligned}$$

Since $L, M \ll \delta^{O(1)} H$ and $LM \ll \delta^{O(1)} H^2$, this is a contradiction unless $v \ll \delta^{-8}$ and $h \gg \delta^8 H$.

From (7.1) together with (4.1) and (4.3) we have

$$\left| \sum_{\substack{x < \ell m \leq x+h \\ m \sim M \\ \ell m \equiv u \pmod{v}}} \alpha(\ell) \beta(m) e(P(\ell m)) \right| \geq \delta^9 h \left(\frac{1}{L} \sum_{L/2 < \ell \leq 2L} |\alpha(\ell)|^2 \right)^{1/2} \left(\frac{1}{M} \sum_{m \sim M} |\beta(m)|^4 \right)^{1/4}.$$

We can write, for some $\nu_j \in \mathbb{R}$,

$$P(n) = \sum_{j=0}^d \nu_j (n - X)^j.$$

We can assume that $\nu_0 = 0$. Furthermore we can spot the condition $\ell m = u \pmod{v}$ using additive characters, so that, for some $r \pmod{v}$ we have

$$\left| \sum_{\substack{x < \ell m \leq x+h \\ m \sim M}} \alpha(\ell) \beta(m) e \left(P(\ell m) + \frac{r\ell m}{v} \right) \right| \geq \delta^9 h \left(\frac{1}{L} \sum_{L/2 < \ell \leq 2L} |\alpha(\ell)|^2 \right)^{1/2} \left(\frac{1}{M} \sum_{m \sim M} |\beta(m)|^4 \right)^{1/4}.$$

Now we are in the position to apply Proposition 7.1 to the polynomial $P(n) + rn/v$. By multiplying the resulting q by v we see that the conclusion of the proposition holds also for the coefficients of $P(n)$, ignoring rn/v . Hence we get that there exists a positive integer $q' \leq \delta^{-O_d(1)}$ such that

$$\|q'(j\nu_j + (j+1)X\nu_{j+1})\|_{\mathbb{R}/\mathbb{Z}} \leq \delta^{-O_d(1)} \frac{X}{H^{j+1}}$$

for all $1 \leq j \leq d$, with the convention that $\nu_{d+1} = 0$.

Next we use a variant of the argument in the treatment of type II sums in [49, Proof of Theorem 1.3 in Section 4]. We start by shifting each ν_j by $(q'j)^{-1}a_j$ for an appropriate $a_j \in \mathbb{Z}$ to get ν'_j such that

$$|q'(j\nu'_j + (j+1)X\nu'_{j+1})| \leq \delta^{-O_d(1)} \frac{X}{H^{j+1}} \quad (7.2)$$

for all $1 \leq j \leq d$. Let

$$P_1(n) = \sum_{j=1}^d \nu'_j (n-X)^j,$$

so that

$$e(P(n)) = e(P_1(n))e\left(-\sum_{j=1}^d \frac{a_j}{q'j} (n-X)^j\right).$$

Choosing $q = q'd!$, we see that $e(P(n) - P_1(n))$ is constant in any arithmetic progression $(\bmod q)$ and thus

$$\|e(P(n) - P_1(n))\|_{\text{TV}([X, X+H] \cap \mathbb{Z}; q)} \leq q \ll \delta^{-O_d(1)} \quad (7.3)$$

By induction one can deduce from (7.2) that

$$\left| \nu'_j - \frac{(-1)^{j-1}}{jX^{j-1}} \nu'_1 \right| \leq \delta^{-O_d(1)} \frac{1}{H^j} \quad (7.4)$$

for all $1 \leq j \leq d+1$. In particular when $j = d+1$ this gives

$$|\nu'_1| \leq \delta^{-O_d(1)} \frac{X^d}{H^{d+1}}.$$

We set $T = 2\pi X\nu'_1$, so that

$$|T| \leq \delta^{-O_d(1)} \left(\frac{X}{H} \right)^{d+1}. \quad (7.5)$$

We write also

$$P_2(n) = \sum_{j=1}^d \frac{(-1)^{j-1}}{jX^{j-1}} \nu'_1 (n-X)^j = \frac{T}{2\pi} \sum_{j=1}^d \frac{(-1)^{j-1}}{j} \left(\frac{n-X}{X} \right)^j.$$

By (7.4) we have that

$$\|e(P_1(n) - P_2(n))\|_{\text{TV}([X, X+H] \cap \mathbb{Z}; q)} \leq q \delta^{-O_d(1)} \ll \delta^{-O_d(1)}. \quad (7.6)$$

By Taylor expansion, for any $k \geq 0$ and $n \in (X, X+H]$,

$$\log \frac{n}{X} = \log \left(1 + \frac{n-X}{X} \right) = \sum_{j=1}^{d+k} \frac{(-1)^{j-1}}{j} \left(\frac{n-X}{X} \right)^j + O \left(\left(\frac{H}{X} \right)^{d+k+1} \right),$$

so that, using (7.5),

$$P_2(n) = \frac{T}{2\pi} \log \frac{n}{X} - \frac{T}{2\pi} \sum_{j=d+1}^{d+k} \frac{(-1)^{j-1}}{j} \left(\frac{n-X}{X} \right)^j + O \left(\delta^{-O_d(1)} \left(\frac{H}{X} \right)^k \right).$$

Hence

$$e(P_2(n))n^{-iT} = X^{-iT} e \left(-\frac{T}{2\pi} \sum_{j=d+1}^{d+k} \frac{(-1)^{j-1}}{j} \left(\frac{n-X}{X} \right)^j \right) + O \left(\delta^{-O_d(1)} \left(\frac{H}{X} \right)^k \right).$$

Taking k large enough in terms of θ , this implies that

$$\|e(P_2(n))n^{-iT}\|_{\text{TV}([X, X+H] \cap \mathbb{Z}; q)} \ll \delta^{-O_d(1)}. \quad (7.7)$$

Now the claim follows by combining (7.3), (7.6), and (7.7) utilizing (2.2).

8. The type I_2 case

In this section we establish the type I_2 case (iv) of Theorem 4.2. Our main tool will be the following elementary partition¹² of the hyperbolic neighborhood $\{(m, n) \in \mathbb{Z}^2 : m \in J; X < nm \leq X + H\}$ into arithmetic progressions, which is non-trivial when H is much larger than $X^{1/3}$.

Theorem 8.1 (Partition of hyperbolic neighborhood). *Let $X, H, M \geq 1$ be such that*

$$X^{1/3} \leq H \leq X \quad \text{and} \quad M \ll X^{1/2},$$

and let J be a subinterval of $(M, 2M]$. Then the set

$$\{(m, n) \in \mathbb{Z}^2 : m \in J; X < nm \leq X + H\} \quad (8.1)$$

can be partitioned for any integer Q obeying

$$\frac{M}{H} \leq Q \leq \frac{M}{(HX)^{1/4}} \quad (8.2)$$

as

$$\bigcup_{q=1}^Q \bigcup_{\substack{a \asymp \frac{X}{M^2}q \\ (a, q)=1}} \bigcup_{P \in \mathcal{P}_{a, q}} P$$

where for each pair a, q of coprime integers with $1 \leq q \leq Q$ and $a \asymp \frac{X}{M^2}q$, $\mathcal{P}_{a, q}$ is a family of $O(\frac{M^3}{XQ^2q})$ arithmetic progressions P in (8.1), each of spacing $(q, -a)$ and length at most $\frac{HQ}{M}$.

In particular, the cardinality of the set (8.1) does not exceed

$$\ll \sum_{1 \leq q \leq Q} \sum_{\substack{a \asymp \frac{X}{M^2}q \\ (a, q)=1}} \frac{M^3}{XQ^2q} \frac{HQ}{M} \ll H. \quad (8.3)$$

¹²In this section only, (m, n) will denote the element of the lattice \mathbb{Z}^2 with coordinates m, n , rather than the greatest common divisor of m and n . We hope that this collision of notation will not cause confusion.

Proof of Theorem 8.1. For future reference, we note from (8.2) and $X^{1/3} \leq H \leq X$ that

$$Q \leq \frac{M}{(HX)^{1/4}} \leq \frac{M}{X^{1/3}} \leq \frac{MH^{1/2}}{X^{1/2}} \leq M. \quad (8.4)$$

Note that if (m, n) lies in (8.1) then $m \asymp M$ and $nm \asymp X$, thus $\frac{n}{m} \asymp \frac{X}{M^2}$. By the Dirichlet approximation theorem, we then have

$$\frac{n}{m} \in \left[\frac{a}{q} - \frac{1}{Qq}, \frac{a}{q} + \frac{1}{Qq} \right]$$

for some $1 \leq q \leq Q$ and some $a \asymp \frac{X}{M^2}q$ coprime to q . If for any such a, q we define $I_{a,q}$ to be the portion of the interval $[\frac{a}{q} - \frac{1}{Qq}, \frac{a}{q} + \frac{1}{Qq}]$ that is not contained in any other such interval $I_{a',q'}$ with $q' < q$, we see that the $I_{a,q}$ are disjoint intervals, and that we can partition (8.1) into sets

$$\{(m, n) \in \mathbb{Z}^2 : m \in J; \frac{n}{m} \in I_{a,q}; \quad X < nm \leq X + H\} \quad (8.5)$$

where a, q range over those coprime integers with

$$1 \leq q \leq Q; \quad \frac{a}{q} \asymp \frac{X}{M^2}. \quad (8.6)$$

It then suffices to show that each such set (8.5) can be partitioned into $O(\frac{M^3}{XQ^2q})$ arithmetic progressions P in \mathbb{Z}^2 , each of spacing $(q, -a)$ and length at most $\frac{HQ}{M}$.

Fix a, q , and write $I = I_{a,q}$. It in fact suffices to show that the set (8.5) can be partitioned into $O(\frac{M^3}{XQ^2q})$ arithmetic progressions P of spacing $(q, -a)$ and arbitrary length, so long as we also show that the total cardinality of (8.5) is $O(\frac{HM^2}{XQq})$. This is because any such progression P can be partitioned into $O(\frac{M}{HQ} \#P + 1)$ subprogressions of the same spacing $(q, -a)$ and length at most $\frac{HQ}{M}$, and

$$\sum_P \left(\frac{M}{HQ} \#P + 1 \right) \ll \frac{M}{HQ} \frac{HM^2}{XQq} + \frac{M^3}{XQ^2q} \ll \frac{M^3}{XQ^2q}.$$

It remains to obtain such a partition. From Bezout's theorem we see that for any integer c , the set $\{(m, n) \in \mathbb{Z}^2 : qn + am = c\}$ is an infinite arithmetic progression of spacing $(q, -a)$. The intersection of (8.5) with this set is

$$E_c := \left\{ \left(m, \frac{c - am}{q} \right) : m, \frac{c - am}{q} \in \mathbb{Z}; m \in J; \frac{c}{mq} - \frac{a}{q} \in I; X < \frac{(c - am)m}{q} \leq X + H \right\}. \quad (8.7)$$

The constraints

$$m \in J; \frac{c}{mq} - \frac{a}{q} \in I; X < \frac{(c - am)m}{q} \leq X + H$$

confine m to the union of at most two intervals in the real line, and hence the set E_c is the union of at most two arithmetic progressions in \mathbb{Z}^2 of spacing $(q, -a)$. It thus suffices to show that E_c is non-empty for at most $O(\frac{M^3}{XQ^2q})$ choices of c , and that

$$\sum_c \#E_c \ll \frac{HM^2}{XQq}. \quad (8.8)$$

We begin with the first claim. If $(m, n) \in E_c$ then $c = qn + am$ and $nm = X + O(H)$ and hence

$$c^2 - (qn - am)^2 = (qn + am)^2 - (qn - am)^2 = 4aqnm = 4aqX + O(aqH). \quad (8.9)$$

On the other hand, we have

$$qn - am = mq \left(\frac{n}{m} - \frac{a}{q} \right) \ll \frac{mq}{qQ} \ll \frac{M}{Q}. \quad (8.10)$$

We thus have

$$c^2 = 4aqX + O(aqH) + O\left(\frac{M^2}{Q^2}\right).$$

From (8.6), (8.2) we have

$$aqH \ll \frac{X}{M^2} q^2 H \ll \frac{M^2}{Q^2} \frac{XHQ^4}{M^4} \ll \frac{M^2}{Q^2}$$

and thus

$$c^2 = 4aqX + O\left(\frac{M^2}{Q^2}\right).$$

Also $\frac{M^2}{Q^2} \leq M^2 \ll X \leq aqX$. Thus on taking square roots we have

$$c = \sqrt{4aqX} + O\left(\frac{1}{\sqrt{aqX}} \frac{M^2}{Q^2}\right)$$

and hence by (8.6)

$$c = \sqrt{4aqX} + O\left(\frac{M^3}{XQ^2q}\right)$$

giving the first claim.

It remains to prove (8.8). We first consider the contribution of those c for which

$$c = \sqrt{4aqX} + O\left(\frac{1}{\sqrt{aqX}} aqH + 1\right),$$

so the total number of possible c here is $O(\frac{1}{\sqrt{aqX}} aqH + 1)$. For a fixed such c , we then have from (8.9) that

$$qn - am = O(\sqrt{aqH}).$$

But once one fixes $c = qn + am$, the residue class of $qn - am$ modulo q and modulo a are both fixed, thus by the Chinese remainder theorem $qn - am$ is restricted to a single residue class modulo aq . Thus the number of possible values of $qn - am$ is $O(\frac{\sqrt{aqH}}{aq} + 1)$. The net contribution of this case to (8.8) is then

$$\ll \left(\frac{1}{\sqrt{aqX}} aqH + 1 \right) \left(\frac{\sqrt{aqH}}{aq} + 1 \right)$$

which expands out to

$$\ll \frac{H^{3/2}}{X^{1/2}} + \frac{a^{1/2}q^{1/2}H}{X^{1/2}} + \frac{H^{1/2}}{a^{1/2}q^{1/2}} + 1.$$

Using (8.6), this becomes

$$\ll \frac{H^{3/2}}{X^{1/2}} + \frac{qH}{M} + \frac{H^{1/2}M}{qX^{1/2}} + 1.$$

Thus we need to show that

$$\frac{H^{3/2}}{X^{1/2}}, \frac{qH}{M}, \frac{H^{1/2}M}{qX^{1/2}}, 1 \ll \frac{HM^2}{XQq}$$

which on using $1 \leq q \leq Q$ rearranges to

$$Q \ll \frac{M}{H^{1/4}X^{1/4}}, \frac{M}{X^{1/3}}, \frac{H^{1/2}M}{X^{1/2}}, \frac{H^{1/2}M}{X^{1/2}}$$

and the claim now follows from (8.4).

Now we consider the contribution of the opposite case, in which $|c - \sqrt{4aqX}|$ exceeds a large multiple of $\frac{1}{\sqrt{aqX}}aqH + 1$. Then $|c^2 - 4aqX|$ exceeds a large multiple of aqH , so from (8.9) we have

$$c^2 = 4aqX + O((qn - am)^2)$$

and thus if we restrict to a dyadic range $qn - am \in \pm[A, 2A]$ for some $1 \leq A \ll \frac{M}{Q}$ that is a power of two (the upper bound coming from (8.10)) we have

$$c = \sqrt{4aqX} + O\left(\frac{1}{\sqrt{aqX}}A^2\right).$$

Thus for a fixed A , the total number of possible c here is $O(\frac{1}{\sqrt{aqX}}A^2)$ (note that we have already excluded those c that lie within $O(1)$ of $\sqrt{4aqX}$). On the other hand, once c is fixed, we see from (8.9) that $(qn - am)^2$ is constrained to an interval of length $O(aqH)$. The quantity $qn - am$ is also constrained to lie in $\pm[A, 2A]$ and to a single residue class modulo aq , so the squares $(qn - am)^2$ are separated by $\gg Aaq$ when $qn - am$ is positive, and similarly when $qn - am$ is negative. Thus the total number of possible values of $qn - am$ available is $O(\frac{aqH}{Aaq} + 1) = O(\frac{H}{A})$, since from (8.2) one has $\frac{H}{A} \gg \frac{H}{M/Q} \geq 1$. Thus the total contribution of this case to (8.8) is

$$\ll \sum_{\substack{1 \leq A \ll \frac{M}{Q} \\ A=2^j}} \frac{A^2}{\sqrt{aqX}} \cdot \frac{H}{A} \ll \frac{1}{\sqrt{aqX}} H \frac{M}{Q}$$

which after applying (8.6) gives $O(\frac{HM^2}{XQq})$ as required. \square

Combining this with the pigeonhole principle we obtain

Corollary 8.2 (Pigeonholing on a hyperbola neighborhood). *Let $X, H, M, Q \geq 1$ be such that*

$$X^{1/3} \leq H \leq X, \quad M \ll X^{1/2}, \quad \text{and} \quad \frac{M}{H} \leq Q \leq \frac{M}{(HX)^{1/4}},$$

and let J be a subinterval of $[M, 2M]$.

Let P_0 be an arithmetic progression in $(X, X + H]$, and let $\beta_1, \beta_2: \mathbb{N} \rightarrow \mathbb{C}$ be functions obeying the bounds

$$\|\beta_1\|_{\text{TV}(\mathbb{N}; q_0)}, \|\beta_2\|_{\text{TV}(\mathbb{N}; q_0)} \leq 1/\delta$$

for some $1 \leq q_0 \leq 1/\delta$ and some¹³ $0 < \delta < 1/(\log X)$. Let $f : \mathbb{Z}^2 \rightarrow \mathbb{C}$ be a 1-bounded function such that

$$\left| \sum_{m \in J} \sum_{\substack{n \\ X < nm \leq X+H}} \beta_1(m) \beta_2(n) 1_{P_0}(nm) f(n, m) \right| \geq \delta H. \quad (8.11)$$

Then for $\gg \delta^{O(1)} \frac{XQ^2}{M^2}$ pairs of coprime integers q, a with $\delta^{O(1)} Q \ll q \leq Q$ and $a \asymp \frac{X}{M^2} q$, one can find an arithmetic progression P in (8.1) of spacing $(q, -a)$ and length at most $\frac{HQ}{M}$ such that

$$\left| \sum_{(m, n) \in P} f(n, m) \right|^* \gg \delta^{O(1)} \frac{HQ}{M}.$$

Here we extend the maximal sum notation (1.4) to sums over arithmetic progressions in \mathbb{Z}^2 in the obvious fashion.

Proof. Let q'_0 be the spacing of P_0 . We first claim that $q'_0 \ll \delta^{-10}$. Indeed, by Shiu's bound (Lemma 2.17) we have

$$\sum_{m \in J} \sum_{\substack{X < nm \leq X+H \\ nm \equiv b(q'_0)}} 1 \leq \sum_{\substack{X < n \leq X+H \\ n \equiv b(q'_0)}} d_2(n) \ll_{\varepsilon} d_2(q'_0) \left((\log X) \frac{H}{q'_0} + X^{\varepsilon} \right),$$

and if $q'_0 \gg \delta^{-10}$ then this together with the triangle inequality contradicts our assumption (8.11). Now we may assume that $q'_0 \ll \delta^{-10}$.

By Lemma 2.2(iii), the left-hand side of (8.11) is bounded by

$$\frac{1}{\delta} \left| \sum_{m \in J} \left(\sum_{\substack{n \\ X < nm \leq X+H}} \beta_2(n) 1_{P_0}(nm) f(n, m) \right) \right|^*$$

which by definition is equal to

$$\frac{1}{\delta} \left| \sum_{m \in J} \sum_{\substack{n \\ X < nm \leq X+H}} 1_{P_1}(m) \beta_2(n) 1_{P_0}(nm) f(n, m) \right|$$

for some arithmetic progression $P_1 \subset J$. Interchanging the n and m sums and using Lemma 2.2(iii) again, we can bound this in turn by

$$\frac{1}{\delta^2} \left| \sum_{m \in J} \sum_{\substack{n \\ X < nm \leq X+H}} 1_{P_1}(m) 1_{P_2}(n) 1_{P_0}(nm) f(n, m) \right|$$

¹³It is likely that with more effort the restriction on δ can be increased up to 1, but that we will not need to do so here.

for some arithmetic progression P_2 . From Theorem 8.1 and the triangle inequality, we have

$$\begin{aligned} & \sum_{m \in J} \sum_{\substack{n \\ X < nm \leq X+H}} 1_{P_1}(m) 1_{P_2}(n) 1_{P_0}(nm) f(n, m) \\ & \ll \sum_{q=1}^Q \sum_{\substack{a \asymp \frac{X}{M^2}q \\ (a, q)=1}} \frac{M^3}{XQ^2q} \sup_{P \in \mathcal{P}_{a,q}} \left| \sum_{(m,n) \in P} 1_{P_1}(m) 1_{P_2}(n) 1_{P_0}(nm) f(n, m) \right| \end{aligned}$$

and since the set $\{(m, n) \in P : m \in P_1, n \in P_2, nm \in P_0\}$ is the union of at most $O(\delta^{-O(1)})$ arithmetic progressions in P (recalling that $q'_0 \ll \delta^{-O(1)}$), we have

$$\left| \sum_{(m,n) \in P} 1_{P_1}(m) 1_{P_2}(n) 1_{P_0}(nm) f(n, m) \right| \ll \delta^{-O(1)} \left| \sum_{(m,n) \in P} f(n, m) \right|^*.$$

We conclude that

$$\sum_{q=1}^Q \sum_{\substack{a \asymp \frac{X}{M^2}q \\ (a, q)=1}} \frac{M^3}{XQ^2q} \sup_{P \in \mathcal{P}_{a,q}} \left| \sum_{(m,n) \in P} f(n, m) \right|^* \gg \delta^{O(1)} H. \quad (8.12)$$

As f is 1-bounded, we have here

$$\frac{M^3}{XQ^2q} \sup_{P \in \mathcal{P}_{a,q}} \left| \sum_{(m,n) \in P} f(n, m) \right|^* \leq \frac{M^3}{XQ^2q} \frac{HQ}{M} = \frac{M^2H}{XQq}, \quad (8.13)$$

since the number of a associated to a fixed q is $O(Xq/M^2)$, we conclude that, for any $q \leq Q$,

$$\sum_{\substack{a \asymp \frac{X}{M^2}q \\ (a, q)=1}} \frac{M^3}{XQ^2q} \sup_{P \in \mathcal{P}_{a,q}} \left| \sum_{(m,n) \in P} f(n, m) \right|^* \ll \frac{H}{Q}.$$

Comparing this with (8.12), we conclude that

$$\sum_{\substack{a \asymp \frac{X}{M^2}q \\ (a, q)=1}} \frac{M^3}{XQ^2q} \sup_{P \in \mathcal{P}_{a,q}} \left| \sum_{(m,n) \in P} f(n, m) \right|^* \gg \delta^{O(1)} \frac{H}{Q} \quad (8.14)$$

for $\gg \delta^{O(1)} Q$ choices of $1 \leq q \leq Q$. By dropping small values of q , we may restrict attention to those q with $\delta^{O(1)} Q \ll q \ll Q$. For each such q , we combine (8.13) with (8.14) to conclude that

$$\frac{M^3}{XQ^2q} \sup_{P \in \mathcal{P}_{a,q}} \left| \sum_{(m,n) \in P} f(n, m) \right|^* \gg \frac{M^2}{Xq} \delta^{O(1)} \frac{H}{Q}$$

for $\gg \delta^{O(1)} \frac{Xq}{M^2} \gg \delta^{O(1)} \frac{XQ}{M^2}$ choices of a , and the claim follows. \square

We can now obtain a preliminary version of Theorem 4.2(iv) (which basically corresponds to the case $A_{I_2} = 1$, after some dyadic decomposition):

Proposition 8.3 (Preliminary type I_2 inverse theorem). *Let $X, H, M \geq 1$ be such that*

$$X^{1/3} \leq H \leq X \quad \text{and} \quad M \ll X^{1/2},$$

and let J be a subinterval of $(M, 2M]$. Let $0 < \delta < 1/(\log X)$, let P_0 be an arithmetic progression in $(X, X+H]$, and let $\beta_1, \beta_2 : \mathbb{N} \rightarrow \mathbb{C}$ be functions obeying the bounds

$$\|\beta_1\|_{\text{TV}(\mathbb{N}; q_0)}, \|\beta_2\|_{\text{TV}(\mathbb{N}; q_0)} \leq 1/\delta$$

for some $1 \leq q_0 \leq 1/\delta$.

Let G/Γ be a filtered nilmanifold of degree d , dimension D , and complexity at most $1/\delta$ for some $d, D \geq 1$, and let $F : G/\Gamma \rightarrow \mathbb{C}$ be a Lipschitz function of norm $1/\delta$ and mean zero, and $g : \mathbb{Z} \rightarrow G$ a polynomial map. Suppose that

$$\left| \sum_{m \in J} \sum_{\substack{n \\ X < nm \leq X+H}} \beta_1(m) \beta_2(n) 1_{P_0}(nm) F(g(nm)\Gamma) \right| \geq \delta H.$$

Then either

$$H \ll_{d,D} \delta^{-O_{d,D}(1)} X^{1/3} \tag{8.15}$$

or else there exists non-trivial horizontal character $\eta : G \rightarrow \mathbb{R}$ of Lipschitz norm $O_{d,D}(\delta^{-O_{d,D}(1)})$ such that

$$\|\eta \circ g\|_{C^\infty(X, X+H]} \ll_{d,D} \delta^{-O_{d,D}(1)}.$$

Proof. We allow all implied constants to depend on d, D . We apply Corollary 8.2 with

$$Q := \left\lfloor \frac{M}{(HX)^{1/4}} \right\rfloor.$$

This gives that for $\gg \delta^{O(1)} XQ^2/M^2$ pairs a, q with $q = O(Q)$ and $a = O(XQ/M^2)$, we have

$$\left| \sum_{k=1}^K F(g((n_0 - ka)(m_0 + kq))\Gamma) \right|^* \gg \delta^{O(1)} \frac{HQ}{M}$$

for some integers n_0, m_0 and some $1 \leq K \leq \frac{HQ}{M}$.

Applying the quantitative Leibman equidistribution theorem (Theorem 2.7), we can find a non-trivial horizontal character $\eta : G \rightarrow \mathbb{R}$ of Lipschitz norm $O(\delta^{-O(1)})$ such that

$$\|\eta \circ g((n_0 - \cdot a)(m_0 + \cdot q))\|_{C^\infty([HQ/M])} \ll \delta^{-O(1)}. \tag{8.16}$$

By pigeonholing we can make η independent of a, q , so that (8.16) holds for $\gg \delta^{O(1)} XQ^2/M^2$ pairs a, q with $q = O(Q)$ and $a = O(XQ/M^2)$. Fix this choice of η . The map $P = \eta \circ g : \mathbb{Z} \rightarrow \mathbb{R}$ is a polynomial of degree at most d ; say

$$P(n) = \eta \circ g(n) = \sum_{0 \leq j \leq d} \alpha_j (n - X)^j.$$

Now suppose that (8.15) fails. We will show that

$$\|q_0 \alpha_j\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)} H^{-j} \tag{8.17}$$

for some $1 \leq q_0 \ll \delta^{-O(1)}$ and all $1 \leq j \leq d$.

We use downward induction on j . Extracting out the top degree coefficient α_d of P , we see that

$$\|\alpha_d(qa)^d\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)} (HQ/M)^{-2d}.$$

We apply the polynomial Vinogradov lemma (Lemma 2.3) twice. Since $HQ/M \ll \delta^{-O(1)}$ implies (8.15), we must have

$$\|q_0\alpha_d\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)} (HQ/M)^{-2d} Q^{-d} (XQ/M^2)^{-d} = \delta^{-O(1)} H^{-2d} X^{-d} Q^{-4d} M^{4d} = \delta^{-O(1)} H^{-d}$$

for some $1 \leq q_0 \ll \delta^{-O(1)}$ by choice of Q . This proves (8.17) for $j = d$.

For the induction step, let $1 \leq j_0 < d$, and assume that (8.17) has already been proved for $j \in \{j_0 + 1, \dots, d\}$. Then the polynomials $n \mapsto q_0\alpha_j(n - X)^j$ has $C^\infty((X, X + H])$ -norm $\ll \delta^{-O(1)}$ for $j \in \{j_0 + 1, \dots, d\}$, and thus the polynomial Q defined by

$$Q(n) = q_0 \left(P(n) - \sum_{j=j_0+1}^d \alpha_j(n - X)^j \right) = q_0 \sum_{0 \leq j \leq j_0} \alpha_j(n - X)^j$$

also satisfies the bound (8.16). By repeating the analysis above with inspecting the top degree coefficient $q_0\alpha_{j_0}$ of Q and applying twice the polynomial Vinogradov lemma, we deduce that

$$\|q_1 \cdot q_0\alpha_{j_0}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)} H^{-j_0}$$

for some $1 \leq q_1 \ll \delta^{-O(1)}$. This completes the induction step after replacing q_0 by q_0q_1 .

Now that we have (8.17), it follows that q_0P has $C^\infty((X, X + H])$ -norm $\ll \delta^{-O(1)}$, and the claim follows after replacing η by $q_0\eta$. \square

Now we are ready to establish Theorem 4.2(iv) in full generality, using an argument similar to that employed in Section 5. Let $d, D, H, X, \delta, G/\Gamma, F, f, A_{I_2}$ be as in Theorem 4.2(iv). Henceforth we allow implied constants to depend on d, D . By Definition 4.1 we can write $f = \alpha * \beta_1 * \beta_2$ where α is supported on $[1, A_{I_2}]$ and obeys (4.1) for all A , and β_1, β_2 obey (4.2). From (4.4) we have

$$\left| \sum_{1 \leq a \leq A_{I_2}} \alpha(a) \sum_m \sum_{\substack{n \\ X/a < nm \leq X/a + H/a}} \beta_1(m) \beta_2(n) 1_{P_0}(anm) F(g(anm)) \Gamma \right| \geq \delta H$$

for some arithmetic progression $P_0 \subset (X, X + H]$. Applying a dyadic decomposition in the a, m, n variables, we may assume that α, β_1, β_2 are supported in $(A, 2A], (M, 2M], (N, 2N]$ for some $1 \leq A \leq A_{I_2}$ and $M, N \geq 1/2$, at the cost of worsening the above bound to

$$\left| \sum_{a \in (A, 2A]} \alpha(a) \sum_{m \in (M, 2M]} \sum_{\substack{N < n \leq 2N \\ X/a < nm \leq X/a + H/a}} \beta_1(m) \beta_2(n) 1_{P_0}(anm) F(g(anm)) \Gamma \right| \geq \delta^{O(1)} H \quad (8.18)$$

(here we use the hypothesis $\delta \leq \frac{1}{\log X}$). By symmetry we may assume that $M \leq N$. We may also assume that $AMN \asymp X$ since the sum is empty otherwise; this implies in particular that $M \ll (X/A)^{1/2}$. We may also assume that

$$H/A \geq \delta^{-C} (X/A)^{1/3} \quad (8.19)$$

for some large constant C (depending only on d, D), since otherwise we have (4.6) after some algebra. By (8.18), Cauchy–Schwarz, and the bound (4.1) we obtain

$$\sum_{a \in (A, 2A]} \left| \sum_{m \in (M, 2M]} \sum_{\substack{N < n \leq 2N \\ X/a < nm \leq X/a + H/a}} \beta_1(m) \beta_2(n) 1_{P_0}(anm) F(g(anm) \Gamma) \right|^2 \geq \delta^{O(1)} H^2 / A. \quad (8.20)$$

For each $a \in (A, 2A]$, we see from the triangle inequality and (4.2) that

$$\begin{aligned} & \sum_{m \in (M, 2M]} \sum_{\substack{N < n \leq 2N \\ X/a < nm \leq X/a + H/a}} \beta_1(m) \beta_2(n) 1_{P_0}(anm) F(g(anm) \Gamma) \\ & \ll \delta^{-O(1)} \sum_{m \in (M, 2M]} \sum_{\substack{n \\ X/a < nm \leq X/a + H/a}} 1 \end{aligned}$$

and hence by the bound (8.3)

$$\sum_{m \in (M, 2M]} \sum_{\substack{n \in (N, 2N) \\ X/a < nm \leq X/a + H/a}} \beta_1(m) \beta_2(n) 1_{P_0}(anm) F(g(anm) \Gamma) \ll \delta^{-O(1)} H / A.$$

Combining this with (8.20) implies that

$$\left| \sum_{m \in (M, 2M]} \sum_{\substack{n \in (N, 2N) \\ X/a < nm \leq X/a + H/a}} \beta_1(m) \beta_2(n) 1_{P_0}(anm) F(g(anm) \Gamma) \right| \gg \delta^{O(1)} H / A$$

for $\gg \delta^{O(1)} A$ values of $a \in (A, 2A]$. Applying Proposition 8.3 (and (8.19)), we conclude that for each such a there exists a non-trivial horizontal character $\eta: G \rightarrow \mathbb{R}$ of Lipschitz norm $O(\delta^{-O(1)})$ such that

$$\|\eta \circ g(a \cdot)\|_{C^\infty(X/a, X/a + H/a)} \ll \delta^{-O(1)}.$$

This η currently is permitted to vary in a , but there are only $O(\delta^{-O(1)})$ choices for η , so by the pigeonhole principle we may assume without loss of generality that η is independent of a . Applying Corollary 2.4 (and (8.19)), we conclude that there exists $1 \leq q \ll \delta^{-O(1)}$ such that

$$\|q\eta \circ g\|_{C^\infty(X, X + H)} \ll \delta^{-O(1)}$$

and the claim follows.

At this point we have proved all cases of Theorem 4.2 which are necessary for our main Theorem (Theorem 1.1).

9. Controlling the Gowers uniformity norms

In order to deduce our Gowers uniformity result in short intervals (Theorem 1.5) from Theorem 1.1, we wish to apply the inverse theorem for the Gowers norms to $\Lambda - \Lambda^\sharp, d_k - d_k^\sharp, \mu$. However, before we can apply the inverse theorem, we need to show that the functions $\Lambda - \Lambda^\sharp, d_k - d_k^\sharp$ possess pseudorandom majorants even when localized to *short* intervals. In the case of long intervals, the existence of pseudorandom majorants for these functions follows from existing works [17], [52], and the main purpose of this section is to show that these long interval majorants also work over short intervals $(X, X + X^\theta]$.

We begin by defining what we mean by pseudorandomness localized to a short range¹⁴.

Definition 9.1 (Pseudorandomness over short intervals). Let $x, H \geq 1$. Let $D \in \mathbb{N}$ and $0 < \eta < 1$. We say that a function $v : \mathbb{Z} \rightarrow \mathbb{R}_{\geq 0}$ is (D, η) -pseudorandom at location x and scale H if the function $v_x(n) := v(x+n)$ satisfies the following. Let ψ_1, \dots, ψ_t be affine-linear forms, where each $\psi_i : \mathbb{Z}^d \rightarrow \mathbb{Z}$ has the form $\psi_i(\mathbf{x}) = \dot{\psi}_i \cdot \mathbf{x} + \psi_i(0)$, with $\dot{\psi}_i \in \mathbb{Z}^d$ and $\psi_i(0) \in \mathbb{Z}$ satisfying $d, t \leq D$, $|\dot{\psi}_i| \leq D$ and $|\psi_i(0)| \leq DH$, and with ψ_i and ψ_j linearly independent whenever $i \neq j$. Then, for any convex body $K \subset [-H, H]^d$,

$$\left| \sum_{\mathbf{n} \in K} v_x(\psi_1(\mathbf{n})) \cdots v_x(\psi_t(\mathbf{n})) - \text{vol}(K) \right| \leq \eta H^d.$$

Remark 9.2. We note that the (D, η) -pseudorandomness of v at location x and scale H directly implies the short interval Gowers uniformity bound $\|v - 1\|_{U^D(x, x+H]} \ll_D \eta^{1/2^D}$, just by the definition of the Gowers norm as a correlation along linear forms.

Our notion of pseudorandomness in the “long interval” case $x = 0$ differs from that of Green–Tao [17, Section 6] in two ways. Firstly, we do not need to impose the *correlation condition* [17, Definition 6.3] (making use of the later work of Dodos and Kanellopoulos [8]). Secondly, we work with pseudorandom functions defined on the integers, as opposed to those defined on cyclic groups. The latter is only a minor technical convenience, as then we do not need to extend majorants defined on the integers into a cyclic group. The next lemma shows that the notion of pseudorandomness over the integers is very closely related to pseudorandomness over a cyclic group.

Lemma 9.3. Let $x, H \geq 1$, $D \in \mathbb{N}$, and $0 < \eta < 1$. Suppose that $v : \mathbb{Z} \rightarrow \mathbb{R}_{\geq 0}$ is (D, η) -pseudorandom at location x and scale H . Then there exists a prime $H < H' \ll_D H$ and a function $\tilde{v} : \mathbb{Z}/H'\mathbb{Z} \rightarrow \mathbb{R}_{\geq 0}$ such that $v(x+n) \leq 2\tilde{v}(n)$ for all $n \in [0, H]$ (where $[0, H]$ is embedded into $\mathbb{Z}/H'\mathbb{Z}$ in the natural way) and such that \tilde{v} satisfies the following. Let ψ_1, \dots, ψ_t be affine-linear forms, where each $\psi_i : \mathbb{Z}^d \rightarrow \mathbb{Z}$ has the form $\psi_i(\mathbf{x}) = \dot{\psi}_i \cdot \mathbf{x} + \psi_i(0)$, with $\dot{\psi}_i \in \mathbb{Z}^d$ and $\psi_i(0) \in \mathbb{Z}$ satisfying $t \leq D$, $|\dot{\psi}_i| \leq D$. Then

$$\sum_{\mathbf{n} \in (\mathbb{Z}/H'\mathbb{Z})^d} \tilde{v}(\psi_1(\mathbf{n})) \cdots \tilde{v}(\psi_t(\mathbf{n})) = (1 + O_D(\eta))(H')^d, \quad (9.1)$$

where the affine-linear forms $\psi_j : (\mathbb{Z}/H'\mathbb{Z})^d \rightarrow \mathbb{Z}/H'\mathbb{Z}$ are induced from their global counterparts in the obvious way.

Proof. Let $H' \in [C_D H, 2C_D H]$ be a prime for large enough $C_D \geq 1$. Take $\tilde{v}(n) = (\frac{1}{2} + \frac{1}{2}v(x+n))1_{n \in [0, H]} + 1_{(H, H')}(n)$, extended to an H' -periodic function. Then the claim (9.1) follows from the (D, η) -pseudorandomness of v at location x and scale H by splitting \tilde{v} into its components. \square

We then state the inverse theorem for unbounded functions that we are going to use.

Proposition 9.4 (An inverse theorem for pseudorandomly bounded functions). Let $s \in \mathbb{N}$ and $0 < \eta < 1$. Let I be an interval of length ≥ 2 . Let $f : I \rightarrow \mathbb{C}$ be a function, and suppose that the following hold.

- There exists a function $v : I \rightarrow \mathbb{R}_{\geq 0}$ such that $\|v - 1\|_{U^{2s}(I)} \leq \eta$ and $|f(n)| \leq v(n)$.
- For any filtered $(s-1)$ -step nilmanifold G/Γ and any Lipschitz function $F : G/\Gamma \rightarrow \mathbb{C}$, we have

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \frac{1}{|I|} \sum_{n \in I} f(n) \overline{F}(g(n)\Gamma) \right| \ll_{\|F\|_{\text{Lip}}, G/\Gamma} \eta.$$

¹⁴Strictly speaking, H does not need to be small in terms of x in Definition 9.1, but that is the regime we are most interested in.

Then we have the Gowers uniformity estimate

$$\|f\|_{U^s(I)} = o_{s;\eta \rightarrow 0}(1).$$

Proof. Let $I = (X, X + H]$, where without loss of generality X and H are integers. The desired result follows from the work of Dodos and Kanellopoulos [8, Theorem 5.1] (which gives the inverse theorem of [17, Proposition 10.1] under weaker hypotheses). Indeed, we can apply [8, Theorem 5.1] to the function $n \mapsto f(X+n)$ on $[1, H]$, noting that the interval Gowers norm estimate $\|\nu - 1\|_{U^{2s}(I)} = o_{\eta \rightarrow 0}(1)$ is equivalent to the cyclic group Gowers norm estimate $\|\tilde{\nu} - 1\|_{U^{2s}(\mathbb{Z}/N'\mathbb{Z})} = o_{\eta \rightarrow 0}(1)$ for all primes $N' \in [100sH, 200sH]$, where $\tilde{\nu}(n)$ is defined as $\nu(X+n)1_{n \in [1, H]}$ for $0 \leq n < N'$ and extended periodically to $\mathbb{Z}/N'\mathbb{Z}$. \square

The following lemma tells us that if a function has a pseudorandom majorant over a long interval, and if the majorant is given by a type I sum, then it in fact has a pseudorandom majorant over short intervals as well. This allows us to conveniently reduce the concept of pseudorandom majorants over short intervals to that over long intervals.

Lemma 9.5 (Pseudorandomness over long intervals implies pseudorandomness over short intervals). *Let $\varepsilon \in (0, 1)$, $D, k \in \mathbb{N}$ be fixed. Let $C \geq 1$ be large enough in terms of k and D . Let $H \in [X^\varepsilon, X/2]$ and $\eta \in ((\log X)^{-C}, 1/2)$, with $X \geq 3$ large enough. Let $\nu : \mathbb{Z} \rightarrow \mathbb{R}_{\geq 0}$ be (D, η) -pseudorandom at location 0 and scale H . Also let $1 \leq A, B \leq \log X$ be integers.*

Suppose that there is an exceptional set $\mathcal{S} \subset \mathbb{Z}$ and a sequence λ_n such that

$$\begin{aligned} \nu(n) &= \sum_{\substack{d|An+B \\ d \leq X^{\varepsilon/(2D)}}} \lambda_d \quad \text{for } n \notin \mathcal{S}, \\ |\lambda_n| &\leq (\log X)^k d(n)^k \quad \text{for all } n, \\ |\nu(n)| &\leq (\log X)^k d(An+B)^k \quad \text{for } n \in \mathcal{S}. \end{aligned} \tag{9.2}$$

Also suppose that \mathcal{S} is small in the sense that

$$|\mathcal{S} \cap [y - 2DH, y + 2DH]| \ll H/(\log X)^{4C} \text{ for } y \in \{0, X\} \tag{9.3}$$

Then ν is $(D, 2\eta)$ -pseudorandom at location X and scale H .

Proof. By (9.2), we can write

$$\begin{aligned} \nu(n) &= 1_{n \notin \mathcal{S}} \sum_{\substack{d|An+B \\ d \leq X^{\varepsilon/(2D)}}} \lambda_d + O((\log X)^k d(An+B)^k 1_{n \in \mathcal{S}}) \\ &= \sum_{\substack{d|An+B \\ d \leq X^{\varepsilon/(2D)}}} \lambda_d + O((\log X)^k d(An+B)^{k+1} 1_{n \in \mathcal{S}}). \end{aligned}$$

Hence, for any convex body $K \subset [-H, H]^d$ and for $x \in \{0, X\}$, we can split the sum

$$\sum_{\mathbf{n} \in K} \prod_{i=1}^t \nu_x(\psi_i(\mathbf{n}))$$

(where $\nu_x(n) := \nu(x+n)$) as the main term

$$\sum_{e_1, \dots, e_t \leq X^{\varepsilon/(2D)}} \lambda_{e_1} \cdots \lambda_{e_t} \sum_{\mathbf{n} \in K} \prod_{i=1}^t 1_{e_i|A(x+\psi_i(\mathbf{n}))+B}. \tag{9.4}$$

and $2^t - 1$ error terms whose contribution is for some $j \leq t$ bounded using (9.2) by

$$\ll (\log X)^{kt} \sum_{\mathbf{n} \in K} \prod_{i=1}^t d(A(x + \psi_i(\mathbf{n})) + B)^{k+1} 1_{x+\psi_j(\mathbf{n}) \in \mathcal{S}}. \quad (9.5)$$

Now, using Cauchy–Schwarz, the inequality $\prod_{i=1}^t x_i \leq \sum_{i=1}^t x_i^t$, (9.3), and Shiu’s bound (Lemma 2.17), (9.5) is

$$\begin{aligned} &\ll (\log X)^{kt} \left(\sum_{\mathbf{n} \in K} 1_{x+\psi_j(\mathbf{n}) \in \mathcal{S}} \right)^{1/2} \left(\sum_{\mathbf{n} \in K} \prod_{i=1}^t d(A(x + \psi_i(\mathbf{n})) + B)^{2(k+1)} \right)^{1/2} \\ &\ll (\log X)^{kt} \left(\sum_{\mathbf{n} \in K} 1_{x+\psi_j(\mathbf{n}) \in \mathcal{S}} \right)^{1/2} \left(\sum_{\mathbf{n} \in K} \sum_{i=1}^t d(A(x + \psi_i(\mathbf{n})) + B)^{2(k+1)t} \right)^{1/2} \\ &\ll H^d (\log X)^{kt-2C} (\log X)^{M_{D,k}} \end{aligned}$$

for some constant $M_{D,k} \geq 1$. If C is large enough in terms of D and k , this is $\ll H^d (\log X)^{-3C/2}$.

We lastly estimate the main term in (9.4). A lattice point counting argument as in [17, Appendix A] gives us

$$\sum_{\mathbf{n} \in K} \prod_{i=1}^t 1_{e_i | A(x + \psi_i(\mathbf{n})) + B} = \alpha_{A,B}(e_1, \dots, e_t) \text{vol}(K) + O(H^{d-1})$$

for some $\alpha_{A,B}(e_1, \dots, e_t) \in [0, 1]$ independent of x and H (since the left-hand side is counting elements of K in some shifted lattice $\mathbf{q}\mathbb{Z} + \mathbf{a}$). Combining this with the estimates $e_1 \cdots e_t \leq X^{\varepsilon/2} \leq H^{1/2}$ and $|\lambda_d| \ll X^{o(1)}$, we see that

$$\sum_{\mathbf{n} \in K} \prod_{i=1}^t \nu_x(\psi_i(\mathbf{n})) = \sum_{e_1, \dots, e_t \leq X^{\varepsilon/(2D)}} \lambda_{e_1} \cdots \lambda_{e_t} \alpha_{A,B}(e_1, \dots, e_t) \text{vol}(K) + O(H^d (\log X)^{-3C/2}). \quad (9.6)$$

Since the main term on the right-hand side of (9.6) is independent of $x \in \{0, X\}$, we see that

$$\sum_{\mathbf{n} \in K} \prod_{i=1}^t \nu_X(\psi_i(\mathbf{n})) = \sum_{\mathbf{n} \in K} \prod_{i=1}^t \nu_0(\psi_i(\mathbf{n})) + O(H^d (\log X)^{-3C/2}).$$

Hence, using the assumption that ν is (D, η) -pseudorandom at location 0 and scale H , ν must also be $(D, 2\eta)$ -pseudorandom at location X and scale H . \square

Lemma 9.5 leads to the existence pseudorandom majorants over short intervals for W -tricked versions of our functions of interest. Let us recall that, for any $w \geq 2$,

$$\Lambda_w(n) := \frac{W}{\varphi(W)} 1_{(n,W)=1},$$

where $W = \prod_{p \leq w} p$. We note for later use that in this notation our model function Λ^\sharp equals to Λ_R , where $R = \exp((\log X)^{1/10})$.

Lemma 9.6 (Pseudorandom majorants over short intervals for $\Lambda - \Lambda_w$, $d_k - d_k^\sharp$). *Let $\varepsilon > 0$ and $D, k \in \mathbb{N}$ be fixed. Let $X \geq H \geq X^\varepsilon \geq 2$. Let $2 \leq w \leq w(X)$, where $w(X)$ is a slowly growing function of X , and denote $W = \prod_{p \leq w} p$. Also let $w \leq \tilde{w} \leq \exp((\log X)^{1/10})$.*

1. There exists a constant $C_0 \geq 1$ such that each of the functions

$$\frac{\varphi(W)}{W} \Lambda(Wn + b)/C_0, \quad \frac{\varphi(W)}{W} \Lambda_{\bar{w}}(Wn + b) \quad (9.7)$$

for $1 \leq b \leq W$ with $(b, W) = 1$, is majorized on $(X, X + H]$ by a (D, η) -pseudorandom function at location X and scale H for some $\eta = o_{w \rightarrow \infty}(1)$. In fact, the latter of the two functions is (D, η) -pseudorandom at location X and scale H .

2. Let W' be such that $W \mid W' \mid W^{\lfloor w \rfloor}$. Suppose that $H \geq X^{1/5+\varepsilon}$. There exists a constant $C_k \geq 1$ such that each of the functions

$$\begin{aligned} & (\log X) \frac{\varphi(W)}{W} \prod_{w \leq p \leq X} \left(1 + \frac{k}{p}\right)^{-1} d_k(W'n + b)/C_k, \\ & (\log X) \frac{\varphi(W)}{W} \prod_{w \leq p \leq X} \left(1 + \frac{k}{p}\right)^{-1} d_k^{\sharp}(W'n + b)/C_k \end{aligned} \quad (9.8)$$

for $1 \leq b \leq W'$ with $(b, W') = 1$, is majorized on $(X, X + H]$ by a (D, η) -pseudorandom function at location X and scale H for some $\eta = o_{w \rightarrow \infty}(1)$.

Remark 9.7. Note that if $\|v_1 - 1\|_{UD(x, x+H]} \leq \eta$ and $\|v_2 - 1\|_{UD(x, x+H]} \leq \eta$, then by the triangle inequality for the Gowers norms also $\|(v_1 + v_2)/2 - 1\|_{UD(x, x+H]} \leq \eta$. Hence, by Remark 9.2, Lemma 9.5 in particular provides us a majorant ν for the difference of the two functions in (9.7) or (9.8) satisfying $\|\nu - 1\|_{UD(x, x+H]} = o_{w \rightarrow \infty}(1)$, allowing us to apply the inverse theorem (Proposition 9.4).

Proof. (1) Let us first consider the function $\frac{\varphi(W)}{W} \Lambda(Wn + b)/C_0$. Let $R' = X^\gamma$ with $\gamma > 0$ small enough in terms of ε, D . Let ψ be a smooth function supported on $[-2, 2]$ with $\psi(0) = -1$ and $\int_0^\infty |\psi'(y)|^2 dy = 1$. Define

$$\Lambda_{R', \psi}(n) := -(\log R') \sum_{d|n} \mu(d) \psi\left(\frac{\log d}{\log R'}\right).$$

Put

$$\nu_b(n) := \frac{\varphi(W)}{W} (\log R')^{-1} \Lambda_{R', \psi}(Wn + b)^2 + 2(\log X) 1_{Wn+b \in S},$$

where S is the set of perfect powers. Then

$$\frac{\varphi(W)}{W} \Lambda(Wn + b) \leq 2\gamma^{-1} \nu_b(n)$$

for $X/2 \leq n \leq X$, since $Wn + b$ being prime implies that $Wn + b$ has no divisors $1 < d \leq X^{2\gamma}$.

From [17, Theorem D.3] we see that ν_b is $(D, o_{w \rightarrow \infty}(1))$ -pseudorandom at location 0 and scale H (since the term $2(\log X) 1_{Wn+b \in S}$ has negligible contribution to the correlations that arise in the definition of pseudorandomness). Moreover, $\nu_b(n)$ can be expanded out as

$$\sum_{\substack{d|Wn+b \\ d \leq X^{4\gamma}}} \lambda_d + 2(\log X) 1_{Wn+b \in S}$$

for some

$$|\lambda_n| \ll (\log X) \sum_{\substack{d_1, d_2 \geq 1 \\ n=[d_1, d_2]}} 1 \ll (\log X) d(n)^2.$$

Hence, by Lemma 9.5, ν_b is $(D, o_{w \rightarrow \infty}(1))$ -pseudorandom also at location X and scale H (since the set $\mathcal{S} := \{n : Wn + b \in \mathcal{S}\}$ certainly obeys (9.3)).

For the case of $\frac{\varphi(W)}{W} \Lambda_{\tilde{w}}(Wn + b)$, we can apply [61, Proposition 5.2] to directly deduce that this function is $(D, o_{w \rightarrow \infty}(1))$ -pseudorandom at location 0 and scale X . To prove the $(D, o_{w \rightarrow \infty}(1))$ -pseudorandomness of this function also at location X and scale H , we show that it is well-approximated by a type I sum. By Möbius inversion,

$$\frac{\varphi(W)}{W} \Lambda_{\tilde{w}}(Wn + b) = \frac{\varphi(W)}{W} \prod_{p \leq \tilde{w}} \left(1 - \frac{1}{p}\right)^{-1} \sum_{\substack{d|Wn+b \\ d|P(\tilde{w})}} \mu(d),$$

and by Lemma 2.18 we have

$$\sum_{X < n \leq X+H} \left| \sum_{\substack{d|Wn+b \\ d|P(\tilde{w}) \\ d \geq X^{\varepsilon/(2D)}}} \mu(d) \right| \ll H \frac{(\log X)^{2\varepsilon}}{\exp(\frac{\varepsilon}{2D} \frac{\log X}{\log \tilde{w}})} \ll H \exp(-(\log X)^{4/5}),$$

say. Hence $\frac{\varphi(W)}{W} \Lambda_{\tilde{w}}(Wn + b) = \nu(n) + \eta(n)$, where ν is of the form of Lemma 9.5 and $\sum_{X < n \leq X+H} |\eta(n)| \ll H \exp(-(\log X)^{3/5})$, say. It suffices to show that ν is $(D, o_{w \rightarrow \infty}(1))$ -pseudorandom at location X and scale H , and this follows from Lemma 9.5.

(2) Note that by (3.14) we have $d_k^\sharp(n) \ll_k d_k(n)$ for all $n \geq 1$, so by Lemma 9.5 it suffices to show that the function

$$h(n) := (\log X) \frac{\varphi(W)}{W} \prod_{w \leq p \leq X} \left(1 + \frac{k}{p}\right)^{-1} d_k(W'n + b) / C'_k$$

is for some $C'_k \geq 1$ majorized by a $(D, o_{w \rightarrow \infty}(1))$ -pseudorandom function at location 0 and scale H , which is of the form (9.2) outside an exceptional set \mathcal{S} satisfying (9.3).

By [52, Proposition 9.4], for any $X \geq 2$ and $1 \leq n \leq 2DX$, we have

$$h(n) \ll \nu(n) + h(n)1_{n \in \mathcal{S}},$$

where ν is a certain $(D, o_{X \rightarrow \infty}(1))$ -pseudorandom function at location 0 and scale X , and \mathcal{S} is defined in [52, Section 7] as

$$\begin{aligned} \mathcal{S} &= \mathcal{S}_1 \cup \mathcal{S}_2, \\ \mathcal{S}_1 &:= \left\{ n \leq 2DX : \exists p : \nu_p(n) \geq \max \left\{ 2, C_1 \frac{\log \log X}{\log p} \right\} \right\}. \\ \mathcal{S}_2 &:= \left\{ n \leq 2DX : \prod_{p \leq X^{1/(\log \log X)^3}} p^{\nu_p(n)} \geq X^{\gamma/\log \log X} \right\} \end{aligned}$$

Here C_1 can be taken arbitrarily large, so we may assume that $C_1 > 8C$ for any given constant C . To show that \mathcal{S} satisfies (9.3), it suffices to show that for $j \in \{1, 2\}$ we have

$$|\mathcal{S}_j \cap [X - 2DH, X + 2DH]| \ll H/(\log X)^{4C}, \quad (9.9)$$

$$|\mathcal{S}_j \cap [-2DH, 2DH]| \ll H/(\log X)^{4C}. \quad (9.10)$$

Let us prove (9.9), the proof of (9.10) is similar but easier.

We first prove (9.9) for $j = 1$. By splitting into shorter intervals if necessary, we may assume that $H \leq X^{1/3}$, say. Note that the number of $n \in (X-2DH, X+2DH]$ satisfying $v_p(n) \geq \max\{2, C_1 \frac{\log \log X}{\log p}\}$ for some p is

$$\begin{aligned} &\ll \sum_{p < (\log X)^{4C}} H \exp(-C_1(\log \log X)) + \sum_{(\log X)^{4C} \leq p \leq (4DH)^{1/2}} \frac{H}{p^2} \\ &+ \sum_{(4DH)^{1/2} < p \leq (2X)^{1/2}} \left(\left\lfloor \frac{X+2DH}{p^2} \right\rfloor - \left\lfloor \frac{X-2DH}{p^2} \right\rfloor \right) \\ &\ll H(\log X)^{-4C} + \sum_{(4DH)^{1/2} < p \leq (2X)^{1/2}} \left(\left\lfloor \frac{X+2DH}{p^2} \right\rfloor - \left\lfloor \frac{X-2DH}{p^2} \right\rfloor \right), \end{aligned}$$

since $C_1 > 8C$.

We can trivially bound

$$\begin{aligned} \sum_{(4DH)^{1/2} < p \leq H(\log X)^{-4C}} \left(\left\lfloor \frac{X+2DH}{p^2} \right\rfloor - \left\lfloor \frac{X-2DH}{p^2} \right\rfloor \right) &\ll \sum_{(4DH)^{1/2} < p \leq H(\log X)^{-4C}} 1 \\ &\ll H(\log X)^{-4C}. \end{aligned}$$

Next, we bound

$$\sum_{H(\log X)^{4C} < p \leq (4DH)^{1/2}} \left(\left\lfloor \frac{X+2DH}{p^2} \right\rfloor - \left\lfloor \frac{X-2DH}{p^2} \right\rfloor \right). \quad (9.11)$$

Note that for any $p \geq H(\log X)^{4C}$ there is at most one multiple of p^2 in $(X-2DH, X+2DH]$, so (9.11) is at most $|S(H(\log X)^{4C}, (4DH)^{1/2})|$, where

$$S(t_1, t_2) := \{d \in (t_1, t_2] : md^2 \in [X-2DH, X+2DH] \text{ for some } m \in \mathbb{N}\}$$

In [11, p. 221], it is proven for $H \geq X^{1/5+\varepsilon}$ that

$$|S(H \log X, 2\sqrt{X})| \ll X^{1/5} \log X,$$

so (9.11) is $\ll H(\log X)^{-4C}$.

Finally, we bound

$$\begin{aligned} &\sum_{H(\log X)^{-4C} \leq p \leq H(\log X)^{4C}} \left(\left\lfloor \frac{X+2DH}{p^2} \right\rfloor - \left\lfloor \frac{X-2DH}{p^2} \right\rfloor \right) \\ &= \sum_{H(\log X)^{-4C} \leq p \leq H(\log X)^{4C}} \left(\frac{4DH}{p^2} - \left\{ \frac{X+2DH}{p^2} \right\} + \left\{ \frac{X-2DH}{p^2} \right\} \right). \quad (9.12) \end{aligned}$$

The first term in the sum gives a negligible contribution of $\ll (\log X)^{4C}$. Pick two 1-periodic smooth functions W^-, W^+ such that $W^-(t) \leq \{t\} \leq W^+(t)$ for all $t \in \mathbb{R}$ and such that $W^\pm(t)$ differs from $\{t\}$ only in the region where $\|t\|_{\mathbb{R}/\mathbb{Z}} \leq (\log X)^{-8C}$, and W^\pm satisfy the derivative bounds $\sup_t |(W^\pm)^{(\ell)}(t)| \ll (\log X)^{8C\ell}$ for $1 \leq \ell \leq 3$. Then (9.12) is

$$\leq O\left((\log X)^{4C}\right) + \sum_{H(\log X)^{-4C} \leq p \leq H(\log X)^{4C}} \left(-W^-\left(\frac{X+2DH}{p^2}\right) + W^+\left(\frac{X-2DH}{p^2}\right) \right).$$

By [45, Proposition 1.12(ii)] and the fact that for any $u, h \geq 0$ we have $\{u + h\} - \{u\} = h$ unless $\|u\|_{\mathbb{R}/\mathbb{Z}} \leq h$, the main term here is

$$\begin{aligned} & \int_{H(\log X)^{-4C}}^{H(\log X)^{4C}} \left(W^+ \left(\frac{X - 2DH}{t^2} \right) - W^- \left(\frac{X + 2DH}{t^2} \right) \right) \frac{dt}{\log t} + O(H(\log X)^{-4C}) \\ & \ll \max_{\sigma \in \{-1, +1\}} \int_{H(\log X)^{-4C}}^{H(\log X)^{4C}} \left(\frac{4DH}{t^2} + 1_{\|(X+2DH\sigma)/t^2\|_{\mathbb{R}/\mathbb{Z}} \leq (\log X)^{-8C}} \right) \frac{dt}{\log t} + H(\log X)^{-4C} \\ & \ll H(\log X)^{-4C}, \end{aligned}$$

since the condition $\|(X+2DH\sigma)/t^2\|_{\mathbb{R}/\mathbb{Z}} \leq (\log X)^{-8C}$ for $t \in [H(\log X)^{-4C}, H(\log X)^{4C}]$ holds in a union of intervals of total measure $\ll H(\log X)^{-4C}$.

Putting the above estimates together, we obtain (9.9) for $j = 1$.

Let us then prove (9.9) for $j = 2$. We thus bound the number of integers $n \in I := (X-2DH, X+2DH]$ that satisfy $\prod_{p \leq X^{1/(\log \log X)^3}} p^{v_p(n)} \geq X^{\gamma/\log \log X}$. Writing $v = X^{1/(\log \log X)^3}$, the number of such $n \in I$ is

$$\ll \sum_{\substack{ab \in I \\ p|a \implies p > v \\ p|b \implies p \leq v \\ b \geq X^{\gamma/\log \log X}}} 1 \leq \sum_{\substack{ab \in I \\ p|a \implies p > v \\ p|b \implies p \leq v}} \left(\frac{b}{X^{\gamma/\log \log X}} \right)^{\frac{10C(\log \log X)^2}{\gamma \log X}} \ll \frac{1}{(\log X)^{10C}} \sum_{n \in I} g(n), \quad (9.13)$$

where g is the completely multiplicative function for which

$$g(p) = \begin{cases} 1 & \text{if } p > v; \\ p^{\frac{10C(\log \log X)^2}{\gamma \log X}} & \text{if } p \leq v. \end{cases}$$

Then Shiu's bound (Lemma 2.17) implies that (9.13) is $\ll H/(\log X)^{4C}$. This proves (9.9) for $j = 2$.

Hence $|\mathcal{S}| \ll H/(\log X)^{4C}$, and in particular arguing as in the beginning of the proof of Lemma 9.5 we see that the fact that v is a $(D, o_{X \rightarrow \infty}(1))$ -pseudorandom function at location 0 and scale X implies that so is $v(n) + h(n)1_{n \in \mathcal{S}}$.

Hence it suffices to show that $v(n)$ is of the form (9.2). The majorant $v(n)$ is defined in [52, Section 7], for some $\gamma > 0$ small enough in terms of D, k , as

$$v(n) := \sum_{u|n} d_k(u) \sum_{\kappa=4/\gamma}^{\lfloor (\log \log X)^3 \rfloor} \sum_{\lambda=\lceil \log(\kappa)/\log 2 - 2 \rceil}^{\lfloor \log((\log \log X)^3)/\log 2 \rfloor} 2^{k\kappa} 1_{u \in U(\lambda, \kappa)} h_\gamma \left(\frac{n}{\prod_{p|u} p^{v_p(n)}} \right), \quad (9.14)$$

where

- $U(\lambda, \kappa)$, defined in [52, Section 7], is a set contained in $[1, X^{10\gamma^{1/2}}]$ and satisfying

$$\begin{aligned} u \in U(\lambda, \kappa), u > 1 & \implies \omega(u) \geq \frac{\gamma\kappa(\lambda + 3 - (\log \kappa)/(\log 2))}{200} \\ 1 \in U(\lambda, \kappa) & \implies \kappa = 4/\gamma; \end{aligned}$$

- $h_\gamma(n) = \sum_{\ell|n} (d_k * \mu)(\ell) \chi \left(\frac{\log \ell}{\log X^\gamma} \right)$, where $\chi : \mathbb{R} \rightarrow [0, 1]$ is some smooth function supported in $[-1, 1]$.

Therefore, in particular, in (9.14) we have

$$\kappa \leq (200/\gamma)(\omega(u) + 1),$$

so that

$$2^{k\kappa} \ll d(u)^M$$

for some constant $M = M_{k,\gamma} \geq 1$. Inserting the definition of h_γ into the definition of ν , and setting $T = X^{10\gamma^{1/2}}$, we see that for some $|\lambda_u| \ll d(u)^{k+M} (\log \log X)^{O_{D,k}(1)}$ we have

$$\nu(n) = \sum_{\substack{u|n \\ u \leq T}} \lambda_u \sum_{\substack{\ell|n \\ \ell \leq X^\gamma}} (d_k * \mu)(\ell) 1_{(\ell,u)=1} \chi\left(\frac{\log \ell}{\log X^\gamma}\right).$$

Writing $e = \ell u$, we see that for some $|\lambda'_e| \ll (\log \log X)^{O_{D,k}(1)} d(e)^{k+M+1} d_{k+1}(e)$ the function ν is of the form

$$\nu(n) = \sum_{\substack{e|n \\ e \leq X^{10\gamma^{1/2}+\gamma}}} \lambda'_e.$$

Taking γ small enough in terms of D, k , this is of the form required in Lemma 9.5, so appealing to that lemma we conclude that ν is $(D, o_{w \rightarrow \infty}(1))$ -pseudorandom at location X and scale H . \square

We need two more lemmas before proving Theorem 1.5.

Lemma 9.8. *Let $D \in \mathbb{N}$ be fixed. Let $1 \leq q \leq H^{1/4}$ be an integer. Let $X \geq H \geq 2$, and let $f : (X, X+H] \rightarrow \mathbb{C}$ be a function with $|f(n)| \ll H^{1/2^{D+2}}$. Then we have*

$$\|f\|_{U^D(X, X+H]} \leq \frac{1}{q} \sum_{1 \leq a \leq q} \|f_{q,a}\|_{U^D(X/q, (X+H)/q]} + O(H^{-1/2}),$$

where $f_{q,a}(n) := f(qn + a)$.

Proof. Denote by $1_{a(q)}$ the indicator of the arithmetic progression $a \pmod{q}$. Then, by the triangle inequality for the Gowers norms, we have

$$\|f\|_{U^D(X, X+H]} \leq \sum_{1 \leq a \leq q} \|f 1_{a(q)}\|_{U^D(X, X+H]}.$$

The claim now follows by making a linear change of variables $(n, \mathbf{h}) = (qn' + a, q\mathbf{h}')$ in the definition of $\|f 1_{a(q)}\|_{U^D(X, X+H)}$. \square

Lemma 9.9. *Let $D, k \in \mathbb{N}$ and $\varepsilon > 0$ be fixed, with $\varepsilon > 0$ small enough. Let $X \geq H \geq X^\varepsilon$, and let $1 \leq q \leq X^{\varepsilon^2}$ be an integer. Let $f(n) = (\log X)^{1-k} d_k(n)$. Then for $1 \leq a \leq q$ with $(a, q) = 1$ we have*

$$\|f_{q,a}\|_{U^D(X, X+H]} \ll \left(\frac{\varphi(q)}{q}\right)^{k-1},$$

where $f_{q,a}(n) := f(qn + a)$.

Proof. Let $g_{q,a}(n) := d_k(qn + a)$. By the definition of the interval Gowers norms and the fact that $\|1_{(X,X+H]}\|_{U^D(\mathbb{Z})}^{2^D} \asymp H^{D+1}$, we have

$$\begin{aligned} \|g_{q,a}\|_{U^D(X,X+H]}^{2^D} &\asymp \frac{1}{H^{D+1}} \sum_n \sum_{h_1, \dots, h_D} \prod_{\omega \in \{0,1\}^D} d_k(q(n + \omega \cdot \mathbf{h}) + a) 1_{(X,X+H]}(n + \omega \cdot \mathbf{h}) \\ &\ll \frac{1}{H^{D+1}} \sum_{X < n \leq X+H} \sum_{\substack{|h_1|, \dots, |h_D| \leq 2H \\ h_i \text{ distinct}}} \prod_{\omega \in \{0,1\}^D} d_k(q(n + \omega \cdot \mathbf{h}) + a) + H^{-1/2}. \end{aligned} \quad (9.15)$$

We can upper bound the correlation of these multiplicative functions using Henriot's bound [30, Theorem 3] (taking $x \rightarrow X$, $y \rightarrow H$, $\delta \rightarrow 2^{-D-2}$, $Q(n) \rightarrow \prod_{\omega \in \{0,1\}^D} (q(n + \omega \cdot \mathbf{h}) + a)$ there), obtaining

$$\begin{aligned} &\frac{1}{H} \sum_{X < n \leq X+H} \prod_{\omega \in \{0,1\}^D} d_k(q(n + \omega \cdot \mathbf{h}) + a) \\ &\ll \Delta_{\mathcal{D}} \prod_{p \leq X} \left(1 - \frac{\rho_Q(p)}{p}\right) \prod_{\omega \in \{0,1\}^D} \sum_{\substack{n \leq X \\ (n, \mathcal{D})=1}} \frac{d_k(n) \rho_{Q_\omega}(n)}{n}, \end{aligned} \quad (9.16)$$

where

$$Q_\omega(u) = q(u + \omega \cdot \mathbf{h}) + a, \quad Q = \prod_{\omega \in \{0,1\}^D} Q_\omega,$$

$$\rho_P(n) = |\{u \pmod n : P(u) \equiv 0 \pmod n\}|,$$

$$\mathcal{D} = \mathcal{D}(\mathbf{h}) = (-1)^{2^D(2^D-1)/2} q^{2^{2D}-2^D} \prod_{\omega \neq \omega'} ((\omega - \omega') \cdot \mathbf{h}) =: (-1)^{2^{D-1}} q^{2^{2D}-2^D} \mathcal{D}',$$

$$\begin{aligned} \Delta_{\mathcal{D}} &= \prod_{p \mid \mathcal{D}} \left(1 + \sum_{\substack{0 \leq \nu_1, \dots, \nu_{2D} \leq 1 \\ (\nu_1, \dots, \nu_{2D}) \neq (0, \dots, 0)}} d_k(p^{\nu_1}) \cdots d_k(p^{\nu_{2D}}) \frac{|\{n \pmod{p^2} : p^{\nu_j} \mid Q_{\omega_j}(n) \forall j\}|}{p^2} \right) \\ &\ll \prod_{p \mid \mathcal{D}'} \left(1 + \frac{O_{D,k}(1)}{p} \right), \end{aligned}$$

where $\omega_1, \dots, \omega_{2D}$ is any ordering of $\{0, 1\}^D$. In order to bound the various expressions above, note that

$$\prod_{p \leq X} \left(1 - \frac{\rho_Q(p)}{p} \right) \ll \prod_{\substack{p \leq X \\ p \nmid \mathcal{D}}} \left(1 - \frac{2^D}{p} \right) \ll (\log X)^{-2^D} \prod_{p \mid \mathcal{D}'} \left(1 + \frac{2^D}{p} \right) \cdot \left(\frac{q}{\varphi(q)} \right)^{2^D}$$

and

$$\sum_{\substack{n \leq X \\ (n, \mathcal{D})=1}} \frac{d_k(n) \rho_{Q_\omega}(n)}{n} \ll \prod_{\substack{p \leq X \\ p \nmid q}} \left(1 + \frac{k}{p} \right) \ll (\log X)^k \left(\frac{\varphi(q)}{q} \right)^k.$$

We now conclude that (9.16) is

$$\ll (\log X)^{(k-1) \cdot 2^D} \left(\frac{\varphi(q)}{q} \right)^{(k-1) \cdot 2^D} \prod_{p \mid \mathcal{D}'} \left(1 + \frac{O_{D,k}(1)}{p} \right)$$

By the inequality $\prod_{i=1}^k x_i \leq \sum_{i=1}^k x_i^k$ and an elementary upper bound for moments of $n/\varphi(n)$, we have

$$\sum_{\substack{|h_1|, \dots, |h_D| \leq 2H \\ h_i \text{ distinct}}} \prod_{p \mid \mathcal{D}'(\mathbf{h})} \left(1 + \frac{O_{D,k}(1)}{p}\right) \ll \sum_{\substack{|h_1|, \dots, |h_D| \leq 2H \\ h_i \text{ distinct}}} \sum_{\omega \in \{-1, 0, 1\}^D \setminus \{\mathbf{0}\}} \prod_{p \mid \omega \cdot \mathbf{h}} \left(1 + \frac{1}{p}\right)^{O_{D,k}(1)} \ll H^D.$$

The claim now follows by combining this with (9.15). \square

We are now ready to prove Theorem 1.5.

Proof of Theorem 1.5. (i) Let H be as in Theorem 1.5(i). By the triangle inequality for the Gowers norms, to prove (1.18) it suffices to show that

$$\|\Lambda^\sharp - \Lambda_w\|_{U^s(X, X+H]} = o_{w \rightarrow \infty}(1). \quad (9.17)$$

and

$$\|\Lambda - \Lambda^\sharp\|_{U^s(X, X+H]} = o_{X \rightarrow \infty}(1) \quad (9.18)$$

The first claim (9.17) follows directly from Lemma 9.6 and Remark 9.2.

We are then left with proving (9.18) and (1.19). Let $1 \leq b \leq W' \leq \log X$ be integers. For $f = \Lambda - \Lambda^\sharp$, by Theorem 1.1 for any $x \in [X/(\log X)^A, X(\log X)^A], H(\log X)^{-A} \leq H' \leq H$ and $G/\Gamma, F$ as in that theorem, we have

$$\begin{aligned} & \sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H'} f(W'n + b) \bar{F}(g(n)\Gamma) \right| \\ &= \sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{\substack{W'x+b < n \leq W'(x+H')+b \\ n \equiv b \pmod{W'}}} f(n) \bar{F}(g(\frac{n-b}{W'})\Gamma) \right| \\ &\ll_A H' / (\log X)^A, \end{aligned} \quad (9.19)$$

since there exists a polynomial sequence $\tilde{g} : \mathbb{Z} \rightarrow G$ such that $\tilde{g}(n) = g((n-b)/W')$ for all $n \equiv b \pmod{W'}$.

Now (1.19) follows by combining the inverse theorem (Proposition 9.4) with the estimate (9.19), Lemma 9.6, and Remark 9.7. Lastly, (1.18) follows from (1.19) and Lemma 9.8.

(ii) We then turn to the case $f = d_k - d_k^\sharp$. Again, Theorem 1.1 gives us the bound (9.19). Together with the inverse theorem (Proposition 9.4), Lemma 9.6 and Remark 9.7, this implies (1.21).

Let

$$h(n) := (\log X)^{1-k} (d_k(n) - d_k^\sharp(n)).$$

Then, to prove (1.20), we must show that

$$\|h\|_{U^D(X, X+H]} = o_{X \rightarrow \infty}(1).$$

Let $\tilde{W} := W^w$ with w an integer tending to infinity slowly¹⁵. By Lemma 9.8, we have

$$\begin{aligned}
 \|h\|_{U^s(X, X+H]} &\leq \frac{1}{\tilde{W}} \sum_{1 \leq a \leq \tilde{W}} \|h_{\tilde{W},a}\|_{U^s(X/\tilde{W}, (X+H)/\tilde{W})} + O(H^{-1/2}) \\
 &= \frac{1}{\tilde{W}} \sum_{\substack{1 \leq a \leq \tilde{W} \\ (a, \tilde{W}) \mid W^{w-1}}} \|h_{\tilde{W},a}\|_{U^s(X/\tilde{W}, (X+H)/\tilde{W})} \\
 &\quad + \frac{1}{\tilde{W}} \sum_{\substack{1 \leq a \leq \tilde{W} \\ (a, \tilde{W}) \nmid W^{w-1}}} \|h_{\tilde{W},a}\|_{U^s(X/\tilde{W}, (X+H)/\tilde{W})} + O(H^{-1/2}). \tag{9.20}
 \end{aligned}$$

The number of terms in the last sum is

$$\ll \sum_{p \leq w} \frac{\tilde{W}}{p^w} \ll \frac{\tilde{W}}{2^w},$$

so by Lemma 9.9 the contribution of this sum is $\ll 2^{-w/2}$, say. The first sum over a in (9.20) can further be written as

$$\sum_{\ell \mid W^{w-1}} d_k(\ell) \sum_{\substack{1 \leq a \leq \tilde{W} \\ (a, \tilde{W}) = \ell}} \left\| \frac{h_{\tilde{W},a}}{d_k(\ell)} \right\|_{U^s(X/\tilde{W}, (X+H)/\tilde{W})} \tag{9.21}$$

Since $d_k^\sharp(m) \ll d_k(m)$, for $(a, \tilde{W}) = \ell$ we have

$$\begin{aligned}
 \left(\frac{W}{\varphi(W)} \right)^{k-1} \frac{h_{\tilde{W},a}(n)}{d_k(\ell)} &\ll \left(\frac{W}{\varphi(W)} \right)^{k-1} (\log X)^{1-k} \frac{d_k(\tilde{W}n + a)}{d_k(\ell)} \\
 &= \left(\frac{W}{\varphi(W)} \right)^{k-1} (\log X)^{1-k} d_k \left(\frac{\tilde{W}}{\ell} n + \frac{a}{\ell} \right),
 \end{aligned}$$

and since $W \mid \frac{\tilde{W}}{\ell}$, by Lemma 9.6 and Mertens's theorem this function is pseudorandomly majorized by a $(D, o_{X \rightarrow \infty}(1))$ -pseudorandom function at location 0 and scale H/\tilde{W} . This combined with (9.19) (with \tilde{W}/ℓ in place of W') and Proposition 9.4 yields

$$\left\| \frac{h_{\tilde{W},a}}{d_k(\ell)} \right\|_{U^D(X/\tilde{W}, (X+H)/\tilde{W})} = o_{w \rightarrow \infty} \left(\left(\frac{\varphi(W)}{W} \right)^{k-1} \right), \tag{9.22}$$

uniformly in $1 \leq a \leq \tilde{W}$ with $(\tilde{W}, a) = \ell$.

¹⁵Let us explain why we perform the W -trick for the divisor function with the modulus $\tilde{W} := W^w$ rather than with the modulus W . In order to apply the inverse theorem, we wish to find a modulus W' such that $h(W'n + a)$ is pseudorandomly majorized for almost all $1 \leq a \leq W'$. Since $|h(Wn + a)| \ll d_k((W', a))d_k(\frac{W'}{(W', a)}n + \frac{a}{(W', a)})$, we want to show that this latter function is pseudorandomly majorized for almost all $1 \leq a \leq W'$. By Lemma 9.6, we thus want that $W \mid \frac{W'}{(W', a)}$ for almost all $1 \leq a \leq W$. This property fails if $W' = W$ but holds if $W' = W^w$ with $w \rightarrow \infty$.

Now the bound (1.20) follows from (9.21), (9.22), and the estimate

$$\begin{aligned} \sum_{\ell|W^{w-1}} d_k(\ell) \sum_{\substack{1 \leq a \leq \tilde{W} \\ (a, \tilde{W}) = \ell}} \left(\frac{\varphi(W)}{W} \right)^{k-1} &\ll \sum_{\ell|W^{w-1}} d_k(\ell) \frac{\tilde{W}}{\ell} \left(\frac{\varphi(W)}{W} \right)^k \\ &\ll \tilde{W} \prod_{p|w} \left(1 + \frac{k}{p} + O\left(\frac{1}{p^2}\right) \right) \left(\frac{\varphi(W)}{W} \right)^k \ll \tilde{W}. \end{aligned}$$

(iii) This case follows directly from the inverse theorem (Proposition 9.4 with $\nu = 1$) and Theorem 1.1(iv). \square

10. Applications

In this section, we shall prove the applications stated in Section 1.

Proof of Corollary 1.3. Parts (i) and (iii) follow immediately from Theorem 1.1, as polynomial phases are special cases of nilsequences. By Theorem 1.1 and the triangle inequality, the proof of part (ii) reduces to proving that

$$\left| \sum_{X < n \leq X+H} \Lambda^\sharp(n) e(P(n)) \right| \gg \frac{H}{(\log X)^A}$$

implies (1.10). Recalling from (4.8) that $\Lambda^\sharp(n) = \Lambda_I^\sharp(n) + E(n)$, where Λ_I^\sharp is a $((\log X)^{O(1)}, X^\varepsilon)$ type I sum and $\sum_{X < n \leq X+H} |E(n)| \ll_A H \log^{-A} X$, the claim follows from the type I estimate in [49, Proposition 2.1]. \square

Proof of Theorem 1.6. First note that, since $\log p = (1 + o(1)) \log N$ for $p \in (N, N + N^\kappa]$ and since the contribution of higher prime powers is negligible, we have

$$\mathbb{E}_{N < p \leq N + N^\kappa} f_1(T^{h_1 p} x) \cdots f_k(T^{h_k p} x) = \mathbb{E}_{N < n \leq N + N^\kappa} \Lambda(n) f_1(T^{h_1 n} x) \cdots f_k(T^{h_k n} x) + o_{N \rightarrow \infty}(1). \quad (10.1)$$

Hence, it suffices to show that the right-hand side of (10.1) converges in $L^2(\mu)$.

Let w be a large parameter (which we will eventually send to infinity), and let $W = \prod_{p \leq w} p$. Let

$$\epsilon(n) := \Lambda(n) - \Lambda_w(n);$$

this is a function that has small Gowers norms over short intervals by Theorem 1.5.

We first claim that

$$\int_X \left| \mathbb{E}_{N < n \leq N + N^\kappa} \epsilon(n) f_1(T^{h_1 n} x) \cdots f_k(T^{h_k n} x) \right|^2 d\mu(x) = o_{w \rightarrow \infty}(1). \quad (10.2)$$

Since the average over n in (10.2) is bounded, it is enough to show for all bounded $f_0 : X \rightarrow \mathbb{C}$ that

$$\int_X \mathbb{E}_{N < n \leq N + N^\kappa} \epsilon(n) f_0(x) f_1(T^{h_1 n} x) \cdots f_k(T^{h_k n} x) d\mu(x) = o_{w \rightarrow \infty}(\|f_0\|_{L^2(\mu)}). \quad (10.3)$$

To prove this, we first make the changes of variables $n' = n + N$, $x = T^m y$, with m arbitrary, and use the T -invariance of μ to rewrite the left-hand side of (10.3) as

$$\int_X \mathbb{E}_{m \leq N^\kappa} \mathbb{E}_{n' \leq N^\kappa} \epsilon_{N'}(n') f_0(T^m y) f(T^{m+h_1 n'} T^{h_1 N} y) \cdots f(T^{m+h_k n'} T^{h_k N} y) d\mu(y), \quad (10.4)$$

where $\epsilon_N(n') := \epsilon(n' + N)$. Since $f_i : X \rightarrow \mathbb{C}$ are bounded, we can appeal to the generalized von Neumann theorem in the form of [13, Lemma 2] (after embedding $[N^\kappa]$ to $\mathbb{Z}/M\mathbb{Z}$ for some $M \ll N^\kappa$) to bound (10.4) as

$$\ll \|\epsilon_N\|_{U^k([N^\kappa])} \|f_0\|_{L^2(\mu)} = o_{w \rightarrow \infty}(\|f_0\|_{L^2(\mu)}),$$

where for the second estimate we used Theorem 1.5. Now (10.2) has been proved. Then let $w' > w$. By an argument identical to the proof of (10.2), but using in the end the fact that $\|\Lambda_w - \Lambda_{w'}\|_{U^k[N, N+N^\kappa]} = o_{w \rightarrow \infty}(1)$ (which follows from Theorem 1.5 and the triangle inequality, but could also be proved more directly), we see that also

$$\int_X \left| \mathbb{E}_{N < n \leq N+N^\kappa} (\Lambda_w(n) - \Lambda_{w'}(n)) f_1(T^{h_1 n} x) \cdots f_k(T^{h_k n} x) \right|^2 d\mu(x) = o_{w \rightarrow \infty}(1). \quad (10.5)$$

Consider now

$$\mathbb{E}_{N < n \leq N+N^\kappa} \Lambda_w(n) f_1(T^{h_1 n} x) \cdots f_k(T^{h_k n} x).$$

This can be rewritten as

$$\frac{W}{\varphi(W)} \sum_{\substack{1 \leq b \leq W \\ (b, W) = 1}} \mathbb{E}_{N/W < n \leq (N+N^\kappa)/W} f_1(T^{h_1(Wn+b)} x) \cdots f_k(T^{h_k(Wn+b)} x) + o_{N \rightarrow \infty}(1).$$

Since the sequence $((N/W, (N+N^\kappa)/W))_N$ of intervals are translates of a Følner sequence, from [2, Theorem 1.1] it follows that there exists $\phi_{w,b} : X \rightarrow \mathbb{C}$ such that

$$\int_X \left| \mathbb{E}_{N/W < n \leq (N+N^\kappa)/W} f_1(T^{h_1(Wn+b)} x) \cdots f_k(T^{h_k(Wn+b)} x) - \phi_{w,b}(x) \right|^2 d\mu(x) = o_{N \rightarrow \infty, w}(1).$$

Hence there exists also $\phi_w : X \rightarrow \mathbb{C}$ such that

$$\int_X \left| \mathbb{E}_{N < n \leq N+N^\kappa} \Lambda_w(n) f_1(T^{h_1 n} x) \cdots f_k(T^{h_k n} x) - \phi_w(x) \right|^2 d\mu(x) = o_{N \rightarrow \infty, w}(1). \quad (10.6)$$

By (10.5), for $w' > w$ we have

$$\|\phi_w - \phi_{w'}\|_{L^2(\mu)} = o_{w \rightarrow \infty}(1),$$

so the sequence $(\phi_w)_w$ is Cauchy in $L^2(\mu)$. Let $\phi \in L^2(\mu)$ be its limit. Then, denoting

$$F(x) = \mathbb{E}_{N < n \leq N+N^\kappa} \Lambda(n) f_1(T^{h_1 n} x) \cdots f_k(T^{h_k n} x),$$

from the triangle inequality, (10.2) and (10.6), we have

$$\begin{aligned} \|F - \phi\|_{L^2(\mu)} &= \|\phi_w - \phi\|_{L^2(\mu)} + o_{w \rightarrow \infty}(1) + o_{N \rightarrow \infty, w}(1) \\ &= o_{w \rightarrow \infty}(1) + o_{N \rightarrow \infty, w}(1). \end{aligned}$$

By sending $N, w \rightarrow \infty$ with w tending to ∞ slowly enough, and recalling (10.1), this proves the claim of Theorem 1.6, with the limit being ϕ . \square

For proving Theorem 1.7, we need the generalized von Neumann theorem, so we state here a version of it that is suitable for us.

Lemma 10.1 (Generalized von Neumann theorem). *Let $s, d, t, L \geq 1$ be fixed, and let D be large enough in terms of s, d, t, L . Let v be $(D, o_{N \rightarrow \infty}(1))$ -pseudorandom at location 0 and scale N , and let*

$f_1, \dots, f_t : \mathbb{Z} \rightarrow \mathbb{R}$ satisfy $|f_i(x)| \leq \nu(x)$ for all $i \in [t]$ and $x \in [N]$. Let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms with integer coefficients in s -normal form such that all the linear coefficients of ψ_i are bounded by L in modulus and $|\psi_i(0)| \leq DN$. Let $K \subset [-N, N]^d$ be a convex body with $\Psi(K) \subset (0, N]^d$. Suppose that for some $\delta > 0$ we have

$$\min_{1 \leq i \leq t} \|f_i\|_{U^{s+1}[N]} \leq \delta.$$

Then we have

$$\sum_{\mathbf{n} \in K} \prod_{i=1}^t f_i(\psi_i(\mathbf{n})) = o_{\delta \rightarrow 0}(N^d).$$

Proof. Note that by Lemma 9.3 there exists a prime $N' \ll N$ such that we have a majorant for f_i on the cyclic group $\mathbb{Z}/N'\mathbb{Z}$ satisfying the (D, D, D) -linear forms condition of [17, Definition 6.2]. Then the claim follows from [17, Proposition 7.1], observing that its proof only used the (D, D, D) -linear forms condition of [17, Definition 6.2] and not the correlation condition. \square

Proof of Theorem 1.7. Let w be a sufficiently slowly growing function of X , and let $W = \prod_{p \leq w} p$. Let $\mathbf{N} = (X, \dots, X) \in \mathbb{R}^d$. We can write $K = \mathbf{N} + K'$, where $K' \subset (0, H]^d$ is a convex body. Now the sum (1.25) becomes

$$\sum_{\mathbf{n} \in K' \cap \mathbb{Z}^d} \prod_{i=1}^t \Lambda(\psi_i(\mathbf{n}) + \dot{\psi}_i \cdot \mathbf{N}). \quad (10.7)$$

Writing $\Lambda = \Lambda_w + (\Lambda - \Lambda_w)$, this splits as the main term

$$\sum_{\mathbf{n} \in K' \cap \mathbb{Z}^d} \prod_{i=1}^t \Lambda_w(\psi_i(\mathbf{n}) + \dot{\psi}_i \cdot \mathbf{N})$$

and $2^t - 1$ error terms

$$\sum_{\mathbf{n} \in K' \cap \mathbb{Z}^d} \prod_{i=1}^t \Lambda_i(\psi_i(\mathbf{n}) + \dot{\psi}_i \cdot \mathbf{N}) \quad (10.8)$$

where $\Lambda_i \in \{\Lambda_w, \Lambda - \Lambda_w\}$ and at least one Λ_i equals to $\Lambda - \Lambda_w$. Following [17, Section 5] verbatim, we see that the main term is

$$\text{vol}(K \cap \Psi^{-1}(\mathbb{R}_{>0}^t)) \prod_p \beta_p + o_{X \rightarrow \infty}(H^d).$$

Following [17, Section 4], we may assume that the system of linear forms involved in (10.8) is in s -normal form for some $s \ll_D 1$.

We make the change of variables $\mathbf{n} = W\mathbf{m} + \mathbf{b}$ with $\mathbf{b} \in [0, W]^d$ in (10.8) and abbreviate $M_{\mathbf{b},i} := \dot{\psi}_i \cdot \mathbf{b} + \psi_i(0)$ to rewrite that sum as

$$\begin{aligned}
& \sum_{\mathbf{b} \in [0, W]^d} \sum_{\substack{\mathbf{m} \in \mathbb{Z}^d \\ W\mathbf{m} + \mathbf{b} \in K'}} \prod_{i=1}^t \Lambda_i(\psi_i(W\mathbf{m} + \mathbf{b}) + \dot{\psi}_i \cdot \mathbf{N}) \\
&= \sum_{\mathbf{b} \in [0, W]^d} \sum_{\substack{\mathbf{m} \in \mathbb{Z}^d \\ W\mathbf{m} + \mathbf{b} \in K'}} \prod_{i=1}^t \Lambda_i(W\dot{\psi}_i \cdot \mathbf{m} + \dot{\psi}_i \cdot \mathbf{b} + \psi_i(0)). \\
&= \left(\frac{W}{\varphi(W)} \right)^t \sum_{\substack{\mathbf{b} \in [0, W]^d \\ (M_{\mathbf{b},i}, W) = 1 \forall i \leq t}} \sum_{\substack{\mathbf{m} \in \mathbb{Z}^d \\ \mathbf{m} \in (K' - \mathbf{b})/W}} \prod_{\substack{1 \leq i \leq t \\ \Lambda_i = \Lambda - \Lambda_w}} \left(\frac{\varphi(W)}{W} \Lambda(W\dot{\psi}_i \cdot \mathbf{m} + M_{\mathbf{b},i}) - 1 \right) \\
&\quad + o_{X \rightarrow \infty}(H^d), \tag{10.9}
\end{aligned}$$

where the error term comes from the contribution of integers in the support of Λ that are not w -rough.

By Theorem 1.5(i), uniformly for integers $1 \leq M \leq X$ with $(M, W) = 1$ we have

$$\max_{\substack{1 \leq a \leq W \\ (a, W) = 1}} \left\| \frac{\varphi(W)}{W} \Lambda(W \cdot + M) - 1 \right\|_{U^{s+1}[0, H/W]} = o_{X \rightarrow \infty; s}(1).$$

Moreover, by Lemma 9.6 the function $\frac{\varphi(W)}{W} \Lambda(W \cdot + M) - 1$ is majorized by a $(D, o_{X \rightarrow \infty}(1))$ -pseudorandom measure ν_M at location 0 and scale H/W for any fixed $D \geq 1$. Hence, applying the generalized von Neumann theorem (Lemma 10.1, with $\nu = \frac{1}{t} \sum_{i \leq t} \nu_{M_{\mathbf{b},i}}$), we conclude that (10.9) is

$$\ll \left(\frac{W}{\varphi(W)} \right)^t \cdot W^d \left(\frac{\varphi(W)}{W} \right)^t \cdot o_{X \rightarrow \infty} \left(\left(\frac{H}{W} \right)^d \right) = o_{X \rightarrow \infty}(H^d),$$

completing the proof. \square

Proof of Corollary 1.9. This follows directly from Theorem 1.7, since the assumptions imply that $\beta_p > 0$ for all p , and on the other hand $\beta_p = 1 + O_{t,d,L}(1/p^2)$ by [17, Lemmas 1.3 and 1.6], so we have $\prod_p \beta_p > 0$. \square

A. Variants of the main result

In this appendix discuss in more detail the variants of the main results described in Remark 1.4.

A.1. Results for the Liouville function

It is an easy matter to replace the Möbius function μ by the Liouville function λ in our main results:

Proposition A.1. *The results in Theorem 1.1(i), (iv) (and hence also Corollary 1.3(i), (iv)) continue to hold if μ is replaced by λ .*

Proof. We illustrate the argument for the estimate (1.5), as the other estimates are proven similarly. Under the hypotheses of Theorem 1.1(i), we wish to show that

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X < n \leq X+H} \lambda(n) \overline{F}(g(n)\Gamma) \right|^* \ll_{A, \varepsilon, d, D} \delta^{-O_{d,D}(1)} H \log^{-A} X.$$

Writing $\lambda(n) = \sum_{m \leq \sqrt{2X}: m^2 \mid n} \mu(n/m^2)$ for $n \leq 2X$ and using the triangle inequality, we can bound the left-hand side by

$$\sum_{m \leq \sqrt{2X}} \sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X/m^2 < n \leq X/m^2 + H/m^2} \mu(n) \overline{F}(g(m^2 n) \Gamma) \right|^*.$$

If $m \leq X^{\varepsilon/10}$ (say), then by Theorem 1.1(i) (with X, H, g replaced by $X/m^2, H/m^2, g(m^2 \cdot)$, and ε reduced slightly) we have

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X/m^2 < n \leq X/m^2 + H/m^2} \mu(n) \overline{F}(g(m^2 n) \Gamma) \right|^* \ll_{A, \varepsilon, d, D} m^{-2} \delta^{-O_{d, D}(1)} H \log^{-A} X.$$

For $X^{\varepsilon/10} < m \ll \sqrt{X}$, we simply use the triangle inequality and the trivial bound $|\overline{F}(g(n) \Gamma)| \leq 1/\delta$ to conclude

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X/m^2 < n \leq X/m^2 + H/m^2} \mu(n) \overline{F}(g(m^2 n) \Gamma) \right|^* \ll \frac{1}{\delta} \left(\frac{H}{m^2} + 1 \right).$$

Summing in m , we obtain the claim after a brief calculation (since H is significantly larger than $X^{1/2}$). \square

A.2. Results for the indicator function of the primes

It is also easy to replace the von Mangoldt function Λ with the indicator function $1_{\mathcal{P}}$ of the primes \mathcal{P} :

Proposition A.2. *The results in Theorem 1.1(ii) (and hence also Corollary 1.3(ii)) continue to hold if Λ is replaced by $1_{\mathcal{P}}$, and $\Lambda^{\sharp}(n)$ is replaced by $\frac{1}{\log n} \Lambda^{\sharp}(n)$.*

Proof. From (1.6) and Lemma 2.2(iii) we have

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X < n \leq X+H} \left(\frac{1}{\log n} \Lambda(n) - \frac{1}{\log n} \Lambda^{\sharp}(n) \right) \overline{F}(g(n) \Gamma) \right|^* \ll_{A, \varepsilon, d, D} \delta^{-O_{d, D}(1)} H \log^{-A} X$$

and so by the triangle inequality it will suffice to show that

$$\sum_{X < n \leq X+H} \left| 1_{\mathcal{P}}(n) - \frac{1}{\log n} \Lambda(n) \right| \ll_A H \log^{-A} X.$$

But the summand is supported on prime powers p^j with $2 \leq j \ll \log X$ and $p \ll \sqrt{X}$, so there are at most $O(\sqrt{X} \log X)$ terms, each of which gives a contribution of $O(1)$. Since H is significantly larger than $X^{1/2}$, the claim follows. \square

A.3. Results for the counting function of sums of two squares

It is a classical fact that the counting function

$$r_2(n) := \sum_{\substack{a, b \in \mathbb{Z} \\ a^2 + b^2 = n}} 1$$

can be factorized as $r_2(n) = 4(1 * \chi_4)(n)$, where χ_4 is the non-principal Dirichlet character of modulus 4. This is formally very similar to the divisor function $d_2(n) = (1 * 1)(n)$. In this paper we use the

Dirichlet hyperbola method to expand $d_2(n)$ for $X < n \leq X + H$ as

$$d_2(n) = \sum_{\substack{R_2 \leq n_1 \leq n/R_2 \\ n_1 | n}} 1 + \sum_{\substack{n_1 < R_2 \\ n_1 | n}} 2$$

with $R_2 := X^{1/20}$, and approximate this function by the Type I sum

$$d_2^\sharp(n) = \sum_{\substack{R_2 \leq n_1 < R_2 \\ n_1 | n}} \frac{\log n - \log R_2^2}{\log R_2} + \sum_{\substack{n_1 < R_2 \\ n_1 | n}} 2$$

(these are the $k = 2$ cases of (3.15), (1.2) respectively). In a similar vein, we can expand

$$r_2(n) = \sum_{\substack{R_2 \leq n_1 \leq n/R_2 \\ n_1 | n}} 4\chi_4(n_1) + \sum_{\substack{n_1 < R_2 \\ n_1 | n}} 4(\chi_4(n_1) + \chi_4(n/n_1))$$

and then introduce the twisted Type I approximant

$$r_2^\sharp(n) = \sum_{\substack{R_2 \leq n_1 < R_2 \\ n_1 | n}} 4\chi_4(n_1) \frac{\log n - \log R_2^2}{\log R_2} + \sum_{\substack{n_1 < R_2 \\ n_1 | n}} 4(\chi_4(n_1) + \chi_4(n/n_1)).$$

We then have

Proposition A.3. *The $k = 2$ results in Theorem 1.1(iii) continue to hold if d_2, d_2^\sharp are replaced by r_2, r_2^\sharp respectively.*

This proposition is established by repeating the arguments used to establish Theorem 1.1(iii), but by inserting “twists” by the character χ_4 at various junctures. However, such twists are quite harmless (for instance, since $\|\chi_4\|_{\text{TV}(P;4)} \ll 1$ for any arithmetic progression P , Proposition 2.2(iii) allows one to insert this character into maximal sum estimates without difficulty), and there is no difficulty in modifying the arguments to accommodate this twist.

A.4. Potential result for the indicator function of the sums of two squares

Let $S = \{n^2 + m^2 : n, m \in \mathbb{Z}\}$ be the set of numbers representable as sums of two squares. The Dirichlet series for S is equal to $\zeta(s)^{1/2} L(s, \chi_4)^{1/2}$ times a holomorphic function near $s = 1$, and in particular extends into the classical zero-free region after making a branch cut to the left of $s = 1$ on the real axis.

By a standard Perron formula calculation, one can then obtain asymptotics of the form

$$\sum_{n \leq x} 1_S(n) = x \sum_{j=0}^{A-1} B_j \log^{-j-1/2} x + O_A(x \log^{-A-1/2} x)$$

for any $A > 0$ and some real constants B_j which are in principle explicitly computable; see for instance [7, Theorem 1.1] for a recent treatment (in significantly greater generality) using the Selberg–Delange method. Similar calculations give asymptotics of the form

$$\sum_{\substack{n \leq x \\ n=a \pmod{q}}} 1_S(n) = x \sum_{j=0}^{A-1} B_{j,a,q} \log^{-j-1/2} x + O_A(x \log^{-A-1/2} x)$$

for any fixed residue class a (q) and some further real constants $B_{j,a,q}$. With further effort one can also localize such estimates to intervals $\{X < n \leq X + H\}$ with H not too small (e.g., $H = X^{5/8+\varepsilon}$ or $H = X^{7/12+\varepsilon}$).

This suggests the existence of an approximant $1_S^{\#,A}$ for any given accuracy $A > 0$ that is well approximated by Type I sums, and is such that one has the major arc estimate

$$\left| \sum_{X < n \leq X+H} 1_S(n) - 1_S^{\#,A}(n) \right|^* \ll_A H \log^{-A} x$$

(cf. Theorem 3.1). For small A , it seems likely that one could construct $1_S^{\#,A}$ by a variant of the Cramér–Granville construction used to form $\Lambda^\#$; but for large A it appears that the approximant is more difficult to construct (for instance one may have to use Fourier-analytic methods such as the delta method). However, once such an approximant is constructed, we conjecture that the methods of this paper will produce analogues of Theorem 1.1(ii) (and hence also of Corollary 1.3(ii)) if $\Lambda, \Lambda^\#$ are replaced by $1_S, 1_S^{\#,A'}$ respectively, with A' sufficiently large depending on A . The main point is that a satisfactory analogue of the Heath–Brown decompositions in Lemma 2.16 for 1_S is known; see [59, Lemma 7.2].

We do not foresee any significant technical issues with the remaining portions of the argument, though of course one would need to define the approximant $1_S^{\#,A}$ more precisely before one could say with certainty that the portions of the argument involving this approximant continue to be valid.

A.5. Potential result for the indicator function of smooth numbers

Let $0 < \eta < \frac{1}{2}$, let X be large, and let S_η denote the set of X^η -smooth integers, that is to say those numbers whose prime factors are all less than X^η . Let $H \geq X^{\theta+\varepsilon}$ with $\theta := \frac{1}{2} + \eta$. As is well known, the density of S_η in $[X, X+H]$ is asymptotic to the Dickman function $\rho(1/\eta)$ evaluated at $1/\eta$. We conjecture that the methods of this paper can be used to establish a bound of the form

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X < n \leq X+H} (1_{S_\eta}(n) - \rho(1/\eta)) \overline{F}(g(n)\Gamma) \right|^* \ll_{\varepsilon, d, D, \eta} \delta^{-O_{d,D}(1)} H \log^{-c} X$$

for some absolute constant $c > 0$ under the hypotheses of Theorem 1.1.

Indeed, a Heath–Brown type decomposition, involving only $(1, x^{1/2-\eta}, x^{1/2})$ type II sums and a (somewhat) small exceptional set, was constructed in [40, Lemma 11.5]; the exceptional set was only shown to be small on long intervals such as $[1, X]$ in that paper, but it is likely that one can show the set to also be small on the shorter interval $\{X < n \leq X+H\}$.

There are however some further technical difficulties in implementing our methods here. The first (and less serious) issue is that one would need to verify that the type II sums $f(n)$ produced by [40, Lemma 11.5] obey the bound (4.9); we believe that this is likely to be achievable after some computation. The second and more significant difficulty is that one would need an approximant $1_{S_\eta}^\#$ obeying a major arc estimate of the shape

$$\left| \sum_{X < n \leq X+H} 1_{S_\eta}(n) - 1_{S_\eta}^\#(n) \right|^* \ll_A H \log^{-A} X$$

for any $A > 0$ (possibly after removing a small exceptional set from S_η), in the spirit of Theorem 3.1 and Corollary 3.10.

The constant $\rho(1/\eta)$ is an obvious candidate for such an approximant, but unfortunately such an estimate is only valid for small values of A ; see [31, Theorem 1.8]. Thus, as in the previous discussion for the indicator of the sums of two squares, a more complicated approximant is likely to be required; the

function $\Lambda(x, y)$ appearing in [31, Theorem 1.8] will most likely become involved. See also [53] for some recent estimates on the distribution of smooth numbers in short intervals or arithmetic progressions (in a slightly different regime in which the X'' threshold for smoothness is replaced by a smaller quantity).

Acknowledgments. We are greatly indebted to Maksym Radziwiłł for many helpful discussions during the course of this project and would like to thank Lilian Matthiesen for discussions concerning [51]. We are grateful to the anonymous referee for a careful reading of the paper and for numerous helpful comments and corrections.

Conflicts of interest. None.

Financial support. KM was supported by Academy of Finland grant no. 285894. XS was supported by NSF grant DMS-1802224. TT was supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and by NSF grant DMS-1764034. JT was supported by a Titchmarsh Fellowship, Academy of Finland grant no. 340098, and funding from European Union's Horizon Europe research and innovation programme under Marie Skłodowska-Curie grant agreement No 101058904.

References

- [1] J. Andrade and K. Smith. On Additive Divisor Sums and minorants of divisor functions. *arXiv e-prints*, page arXiv:1903.01566, March 2019.
- [2] T. Austin. On the norm convergence of non-conventional ergodic averages. *Ergodic Theory Dynam. Systems*, 30(2):321–338, 2010.
- [3] R. C. Baker and G. Harman. The three primes theorem with almost equal summands. *R. Soc. Lond. Philos. Trans. Ser. A Math. Phys. Eng. Sci.*, 356(1738):763–780, 1998.
- [4] R. C. Baker, G. Harman, and J. Pintz. The difference between consecutive primes. II. *Proc. London Math. Soc.* (3), 83(3):532–562, 2001.
- [5] J. B. Conrey and S. M. Gonek. High moments of the Riemann zeta-function. *Duke Math. J.*, 107(3):577–604, 2001.
- [6] H. Davenport. On some infinite series involving arithmetical functions. II. *Quart. J. Math. Oxf.*, 8:313–320, 1937.
- [7] R. de la Bretèche and G. Tenenbaum. Remarks on the Selberg-Delange method. *Acta Arith.*, 200(4):349–369, 2021.
- [8] P. Dodos and V. Kanellopoulos. Uniformity norms, their weaker versions, and applications. *Acta Arith.*, 203(3):251–270, 2022.
- [9] W. Duke, J. B. Friedlander, and H. Iwaniec. A quadratic divisor problem. *Invent. Math.*, 115(2):209–217, 1994.
- [10] A.-M. Ernvall-Hytönen and K. Karppinen. On short exponential sums involving Fourier coefficients of holomorphic cusp forms. *Int. Math. Res. Not. IMRN*, (10):Art. ID. rnm022, 44, 2008.
- [11] M. Filaseta and O. Trifonov. On gaps between squarefree numbers. II. *J. London Math. Soc.* (2), 45(2):215–221, 1992.
- [12] N. Frantzikinakis. Multiple recurrence and convergence for Hardy sequences of polynomial growth. *J. Anal. Math.*, 112:79–135, 2010.
- [13] N. Frantzikinakis, B. Host, and B. Kra. Multiple recurrence and convergence for sequences related to the prime numbers. *J. Reine Angew. Math.*, 611:131–144, 2007.
- [14] N. Frantzikinakis, E. Lesigne, and M. Wierdl. Random differences in Szemerédi's theorem and related results. *J. Anal. Math.*, 130:91–133, 2016.
- [15] H. Furstenberg and B. Weiss. A mean ergodic theorem for $(1/N) \sum_{n=1}^N f(T^n x)g(T^{n^2} x)$. In *Convergence in ergodic theory and probability (Columbus, OH, 1993)*, volume 5 of *Ohio State Univ. Math. Res. Inst. Publ.*, pages 193–227. de Gruyter, Berlin, 1996.
- [16] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. *Ann. of Math.* (2), 167(2):481–547, 2008.
- [17] B. Green and T. Tao. Linear equations in primes. *Ann. of Math.* (2), 171(3):1753–1850, 2010.
- [18] B. Green and T. Tao. The Möbius function is strongly orthogonal to nilsequences. *Ann. of Math.* (2), 175(2):541–566, 2012.
- [19] B. Green and T. Tao. The quantitative behaviour of polynomial orbits on nilmanifolds. *Ann. of Math.* (2), 175(2):465–540, 2012.
- [20] B. Green and T. Tao. On the quantitative distribution of polynomial nilsequences—erratum. *Ann. of Math.* (2), 179(3):1175–1183, 2014.
- [21] B. Green, T. Tao, and T. Ziegler. An inverse theorem for the Gowers $U^{s+1}[N]$ -norm. *Ann. of Math.* (2), 176(2):1231–1372, 2012.
- [22] B. Hall. *Lie groups, Lie algebras, and representations*, volume 222 of *Graduate Texts in Mathematics*. Springer, Cham, second edition, 2015. An elementary introduction.
- [23] G. H. Hardy and J. E. Littlewood. The Approximate Functional Equation in the Theory of the Zeta-Function, with Applications to the Divisor-Problems of Dirichlet and Piltz. *Proc. London Math. Soc.* (2), 21:39–74, 1923.
- [24] G. Harman. *Prime-detecting sieves*, volume 33 of *London Mathematical Society Monographs Series*. Princeton University Press, Princeton, NJ, 2007.
- [25] X. He and M. Wang. Discorrelation of multiplicative functions with nilsequences and its application on coefficients of automorphic L -functions. *Mathematika*, 69(1):250–285, 2023.

- [26] X. He and Z. Wang. Möbius disjointness for nilsequences along short intervals. *Trans. Amer. Math. Soc.*, 374(6):3881–3917, 2021.
- [27] D. R. Heath-Brown. Mean values of the zeta function and divisor problems. In *Recent progress in analytic number theory, Vol. 1 (Durham, 1979)*, pages 115–119. Academic Press, London-New York, 1981.
- [28] D. R. Heath-Brown. The ternary Goldbach problem. *Rev. Mat. Iberoamericana*, 1(1):45–59, 1985.
- [29] D. R. Heath-Brown. The number of primes in a short interval. *J. Reine Angew. Math.*, 389:22–63, 1988.
- [30] K. Henriot. Nair-Tenenbaum bounds uniform with respect to the discriminant. *Math. Proc. Cambridge Philos. Soc.*, 152(3):405–424, 2012.
- [31] A. Hildebrand and G. Tenenbaum. Integers without large prime factors. *J. Théor. Nombres Bordeaux*, 5(2):411–484, 1993.
- [32] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. of Math.* (2), 161(1):397–488, 2005.
- [33] M. N. Huxley. On the difference between consecutive primes. *Invent. Math.*, 15:164–170, 1972.
- [34] M. N. Huxley. Exponential sums and lattice points. III. *Proc. London Math. Soc.* (3), 87(3):591–609, 2003.
- [35] A. Ivić. The general additive divisor problem and moments of the zeta-function. In *New trends in probability and statistics, Vol. 4 (Palanga, 1996)*, pages 69–89. VSP, Utrecht, 1997.
- [36] A. Ivić. *The Riemann zeta-function*. Dover Publications, Inc., Mineola, NY, 2003. Theory and applications, Reprint of the 1985 original.
- [37] H. Iwaniec and E. Kowalski. *Analytic number theory*, volume 53 of *American Mathematical Society Colloquium Publications*. American Mathematical Society, Providence, RI, 2004.
- [38] A. Kanigowski. Prime orbits for some smooth flows on \mathbb{T}^2 . *arXiv e-prints*, page arXiv:2005.09403, May 2020.
- [39] A. Kanigowski, M. Lemańczyk, and M. Radziwiłł. Prime number theorem for analytic skew products. *To appear in Ann. of Math.*, 2023.
- [40] O. Klurman, A. P. Mangerel, and J. Teräväinen. Multiplicative functions in short arithmetic progressions. *Proc. Lond. Math. Soc.*, 127(3):366–446, 2023.
- [41] G. Kolesnik. On the estimation of multiple exponential sums. In *Recent progress in analytic number theory, Vol. 1 (Durham, 1979)*, pages 231–246. Academic Press, London-New York, 1981.
- [42] A. N. Le. Nilsequences and multiple correlations along subsequences. *Ergodic Theory Dynam. Systems*, 40(6):1634–1654, 2020.
- [43] K. Matomäki, J. Maynard, and X. Shao. Vinogradov’s theorem with almost equal summands. *Proc. Lond. Math. Soc.* (3), 115(2):323–347, 2017.
- [44] K. Matomäki and M. Radziwiłł. A note on the Liouville function in short intervals. *arXiv e-prints*, page arXiv:1502.02374, Feb 2015.
- [45] K. Matomäki, M. Radziwiłł, X. Shao, T. Tao, and J. Teräväinen. Singmaster’s conjecture in the interior of Pascal’s triangle. *Q. J. Math.*, 73(3):1137–1177, 2022.
- [46] K. Matomäki, M. Radziwiłł, X. Shao, T. Tao, and J. Teräväinen. Higher uniformity of arithmetic functions in short intervals II. Almost all intervals. *Under preparation*, 2023.
- [47] K. Matomäki, M. Radziwiłł, and T. Tao. Correlations of the von Mangoldt and higher divisor functions I. Long shift ranges. *Proc. Lond. Math. Soc.* (3), 118(2):284–350, 2019.
- [48] K. Matomäki, M. Radziwiłł, T. Tao, J. Teräväinen, and T. Ziegler. Higher uniformity of bounded multiplicative functions in short intervals on average. *Ann. of Math.* (2), 197(2):739–857, 2023.
- [49] K. Matomäki and X. Shao. Discorrelation between primes in short intervals and polynomial phases. *Int. Math. Res. Not. IMRN*, (16):12330–12355, 2021.
- [50] K. Matomäki and J. Teräväinen. On the Möbius function in all short intervals. *J. Eur. Math. Soc. (JEMS)*, 25(4):1207–1225, 2023.
- [51] L. Matthiesen. Generalized Fourier coefficients of multiplicative functions. *Algebra Number Theory*, 12(6):1311–1400, 2018.
- [52] L. Matthiesen. Linear correlations of multiplicative functions. *Proc. Lond. Math. Soc.* (3), 121(2):372–425, 2020.
- [53] L. Matthiesen and M. Wang. Smooth numbers are orthogonal to nilsequences. *arXiv e-prints*, page arXiv:2211.16892, November 2022.
- [54] H. L. Montgomery and R. C. Vaughan. *Multiplicative number theory. I. Classical theory*, volume 97 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2007.
- [55] N. Ng and M. Thom. Bounds and conjectures for additive divisor sums. *Funct. Approx. Comment. Math.*, 60(1):97–142, 2019.
- [56] K. Ramachandra. Some problems of analytic number theory. *Acta Arith.*, 31(4):313–324, 1976.
- [57] K. A. Ribet. On l -adic representations attached to modular forms. *Invent. Math.*, 28:245–275, 1975.
- [58] O. Robert and P. Sargos. Three-dimensional exponential sums with monomials. *J. Reine Angew. Math.*, 591:1–20, 2006.
- [59] X. Shao and J. Teräväinen. The Bombieri-Vinogradov theorem for nilsequences. *Discrete Anal.*, pages Paper No. 21, 55, 2021.
- [60] P. Shiu. A Brun-Titchmarsh theorem for multiplicative functions. *J. Reine Angew. Math.*, 313:161–170, 1980.
- [61] T. Tao and J. Teräväinen. Quantitative bounds for Gowers uniformity of the Möbius and von Mangoldt functions. *arXiv e-prints*, page arXiv:2107.02158, July 2021.

- [62] G. Tenenbaum. *Introduction to analytic and probabilistic number theory*, volume 46 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Translated from the second French edition (1995) by C. B. Thomas.
- [63] T. D. Wooley and T. D. Ziegler. Multiple recurrence and convergence along the primes. *Amer. J. Math.*, 134(6):1705–1732, 2012.
- [64] T. Zhan. On the representation of large odd integer as a sum of three almost equal primes. *Acta Math. Sinica (N.S.)*, 7(3):259–272, 1991. A Chinese summary appears in *Acta Math. Sinica* 35 (1992), no. 4, 575.