



# Higher uniformity of arithmetic functions in short intervals

## II. Almost all intervals

Kaisa Matomäki<sup>1</sup> · Maksym Radziwiłł<sup>2</sup> · Xuancheng Shao<sup>3</sup> · Terence Tao<sup>4</sup> ·  
Joni Teräväinen<sup>5</sup>

Received: 19 November 2024 / Accepted: 15 January 2026  
© The Author(s) 2026

### Abstract

We study higher uniformity properties of the von Mangoldt function  $\Lambda$ , the Möbius function  $\mu$ , and the divisor functions  $d_k$  on short intervals  $(x, x + H]$  for almost all  $x \in [X, 2X]$ . Let  $\Lambda^\sharp$  and  $d_k^\sharp$  be suitable approximants of  $\Lambda$  and  $d_k$ ,  $G/\Gamma$  a filtered nilmanifold, and  $F : G/\Gamma \rightarrow \mathbb{C}$  a Lipschitz function. Then our results imply for instance that when  $X^{1/3+\varepsilon} \leq H \leq X$  we have, for almost all  $x \in [X, 2X]$ ,

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H} (\Lambda(n) - \Lambda^\sharp(n)) \overline{F}(g(n)\Gamma) \right| \ll H \log^{-A} X$$

for any fixed  $A > 0$ , and that when  $X^\varepsilon \leq H \leq X$  we have, for almost all  $x \in [X, 2X]$ ,

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H} (d_k(n) - d_k^\sharp(n)) \overline{F}(g(n)\Gamma) \right| = o(H \log^{k-1} X).$$

As a consequence, we show that the short interval Gowers norms  $\|\Lambda - \Lambda^\sharp\|_{U^s(X, X+H]}$  and  $\|d_k - d_k^\sharp\|_{U^s(X, X+H]}$  are also asymptotically small for any fixed  $s$  in the same ranges of  $H$ . This in turn allows us to establish the Hardy–Littlewood conjecture and the divisor correlation conjecture with a short average over one variable. Our main new ingredients are type II estimates obtained by developing a “contagion lemma” for nilsequences and then using this to “scale up” an approximate functional equation for the nilsequence to a larger scale. This extends an approach developed by Walsh for Fourier uniformity.

### Contents

1	Introduction	1
1.1	Previous results	1
1.2	Gowers uniformity	1

Extended author information available on the last page of the article

1.3	Applications . . . . .	
1.4	The case of $d_2$ revisited . . . . .	
1.5	Methods of proof . . . . .	
1.6	Notation . . . . .	
2	Basic tools . . . . .	
2.1	Total variation . . . . .	
2.2	Vinogradov lemma . . . . .	
2.3	Equidistribution on nilmanifolds . . . . .	
2.4	Combinatorial lemmas . . . . .	
3	Major arc estimates . . . . .	
3.1	Lemmas on Dirichlet polynomials . . . . .	
3.2	Type II major arc estimates . . . . .	
3.3	Type I major arc estimates . . . . .	
3.4	Theorem 3.1 in the case of small $ T $ . . . . .	
3.5	Theorem 3.1 in the case of large $ T $ . . . . .	
4	Reduction to type II estimates . . . . .	
4.1	Combinatorial decompositions . . . . .	
4.2	Divisor sums in short intervals . . . . .	
4.3	Deduction of Theorem 1.1 . . . . .	
5	Contagion lemmas . . . . .	
5.1	Walsh's contagion lemma . . . . .	
5.2	Monomial contagion . . . . .	
5.3	Nilsequence contagion . . . . .	
6	Proof of type II case . . . . .	
7	Controlling the Gowers uniformity norms . . . . .	
7.1	Gowers uniformity of the von Mangoldt and Möbius functions . . . . .	
7.2	Gowers uniformity of the divisor functions . . . . .	
8	Applications . . . . .	
8.1	The polynomial phase case . . . . .	
8.2	Calculating the main terms . . . . .	
8.3	Proofs of the correlation asymptotics . . . . .	
9	Fixed Fourier estimate for $d_2$ . . . . .	
	References . . . . .	
	Authors and Affiliations . . . . .	

## 1 Introduction

This paper is a sequel to our previous paper [30], which was focused on decorrelation of arithmetic functions  $f: \mathbb{N} \rightarrow \mathbb{C}$  with arbitrary nilsequences  $n \mapsto F(g(n)\Gamma)$  in all short intervals  $(X, X + H]$  with  $H = X^\theta$ . In this paper we consider the corresponding question for *almost all* short intervals  $(x, x + H]$  with  $x \in [X, 2X]$ , thus we permit a small number of exceptional short intervals on which there might be strong correlation. It turns out that in this setting we can reduce the exponent  $\theta$  significantly.

As in the previous paper we will restrict our attention to the following model examples of functions  $f$  and their approximants  $f^\sharp$ :

- $\mu$ : The *Möbius function*  $\mu(n)$ , defined as  $(-1)^j$  when  $n$  is the product of  $j$  distinct primes, and 0 otherwise, with the approximant  $\mu^\sharp$  set equal to zero.
- $\Lambda$ : 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, with approximant

$$\Lambda^\sharp(n) := \frac{P(R)}{\varphi(P(R))} 1_{(n, P(R))=1},$$

where  $P(w) := \prod_{p < w} p, \quad R := \exp((\log X)^{1/10}).$  (1.1)

$d_k$ : The  $k^{\text{th}}$  *divisor function*  $d_k(n)$ , defined to equal to the number of representations of  $n$  as the product  $n = n_1 \cdots n_k$  of  $k$  natural numbers, where  $k \geq 2$  is a fixed natural number. (In particular, all implied constants in our asymptotic notation are allowed to depend on  $k$ .) The approximant  $d_k^\sharp$  will be taken to be

$$d_k^\sharp(n) := \sum_{\substack{m \leq R_k^{2k-2} \\ m|n}} P_m(\log n),$$
 (1.2)

where  $R_k$  is the parameter

$$R_k := X^{\frac{\varepsilon}{10k}},$$

where  $0 < \varepsilon \leq 1$  is small but fixed<sup>1</sup> and the polynomials  $P_m(t)$  (which have degree at most  $k - 1$ ) are given by the formula

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

We refer the reader to the discussion in [30, Sect. 3.1] for a justification for choosing these approximants  $\mu^\sharp, \Lambda^\sharp, d_k^\sharp$ .

By a “nilsequence”, we mean a function of the form  $n \mapsto F(g(n)\Gamma)$ , where  $G/\Gamma$  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 Sect. 2.3, but for now we remark (as mentioned for example in [30, Sect. 1]) 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.

---

<sup>1</sup>In [30] we stated our results for  $\varepsilon = 1$ , but as noted in [30, Sect. 1], the arguments work equally well with smaller choices of  $\varepsilon$ , so long as the constants in the asymptotic notation are allowed to depend on  $\varepsilon$ .

As in [30], for technical reasons, it can be beneficial to consider “maximal discorrelation” estimates. Loosely following Robert and Sargos [34], we adopt the notation that, for an interval  $I$ ,

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

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

Now we are ready to state our main theorem.<sup>2</sup>

**Theorem 1.1** (Discorrelation estimate) *Let  $X \geq 3$  and  $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$  for some  $0 \leq \theta < 1$  and  $\varepsilon > 0$ . Let  $\delta \in (0, 1/2)$ . Let  $G/\Gamma$  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  $\theta = 1/3$  and  $A > 0$ , then*

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H} \mu(n) \overline{F}(g(n)\Gamma) \right|^* \leq \frac{H}{\log^A X}$$

*for all  $x \in [X, 2X]$  outside of a set of measure  $O_A(\delta^{-O_{d,D,\varepsilon}(1)} X \log^{-A} X)$ .*

(ii) *If  $\theta = 1/3$  and  $A > 0$ , then*

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H} (\Lambda(n) - \Lambda^\sharp(n)) \overline{F}(g(n)\Gamma) \right|^* \leq \frac{H}{\log^A X}$$

*for all  $x \in [X, 2X]$  outside of a set of measure  $O_A(\delta^{-O_{d,D,\varepsilon}(1)} X \log^{-A} X)$ .*

(iii) *If  $\theta = 1/3$  and  $k \geq 2$ , then for some constant  $c_{k,d,D,\varepsilon} > 0$  we have*

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

*for all  $x \in [X, 2X]$  outside of a set of measure  $O(\delta^{-O_{d,D,\varepsilon}(1)} X^{1-c_{k,d,D,\varepsilon}})$ .*

(iv) *If  $\theta = 0$ ,  $A > 0$ , and  $\eta > 0$ , and additionally  $F$  is 1-bounded, then*

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H} \mu(n) \overline{F}(g(n)\Gamma) \right|^* \leq \eta H$$

*for all  $x \in [X, 2X]$  outside of a set of measure  $O_A(\delta^{-O_{d,D,\varepsilon,\eta}(1)} X \log^{-A} X)$ .*

<sup>2</sup>For definitions of terms such as “filtered nilmanifold” and  $\text{Poly}(\mathbb{Z} \rightarrow G)$ , see Definitions 2.7 and 2.6 below. For our conventions for asymptotic notation, see Sect. 1.6.

(v) If  $\theta = 0$ ,  $A > 0$ ,  $k \geq 2$  and  $\eta > 0$ , and additionally  $F$  is 1-bounded, then

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H} (d_k(n) - d_k^\sharp(n)) \overline{F}(g(n)\Gamma) \right|^* \leq \eta H \log^{k-1} X$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_A(\delta^{-O_{d,D,\varepsilon,\eta}(1)} X \log^{-A} X)$ .

By the preceding remarks, Theorem 1.1 works in particular with polynomial phase twists. Hence we have the following corollary (proven in Sect. 8.1), which is new already for linear phases.

**Corollary 1.2** (Discorrelation of  $\mu$  and  $\Lambda$  with polynomial phases) *Let  $X \geq 3$  and  $X^{1/3+\varepsilon} \leq H \leq X^{1-\varepsilon}$  for some  $\varepsilon > 0$ . Let  $d \geq 1$ , and for each  $x \in [X, 2X]$  let  $P_x: \mathbb{Z} \rightarrow \mathbb{R}$  be a polynomial of degree  $d$ . Also let  $A > 0$ .*

(i) *We have*

$$\left| \sum_{x < n \leq x+H} \mu(n) e(P_x(n)) \right| \leq \frac{H}{\log^A X}$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_{A,d,\varepsilon}(X \log^{-A} X)$ .

(ii) *Either we have*

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

for all  $x \in [X, 2X]$  outside of a set of measure  $O_{A,\varepsilon,d}(X \log^{-A} X)$  or there exists an integer  $1 \leq q \ll (\log X)^{O_{A,d,\varepsilon}(1)}$  such that

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

for measure  $\gg_{A,d,\varepsilon} X \log^{-A} X$  of  $x \in [X, 2X]$ , where  $\alpha_{j,x}$  is the degree  $j$  coefficient of the polynomial  $n \mapsto P_x(x+n)$  and  $\|y\|_{\mathbb{R}/\mathbb{Z}}$  denotes the distance from  $y$  to the nearest integer(s).

In case of correlations of  $d_2$  with a fixed linear phase, we can obtain power-saving estimates for almost all sums of length  $X^\varepsilon$ ; see Theorem 1.8 below.

### 1.1 Previous results

We start by discussing the results concerning the von Mangoldt function  $\Lambda$ . Huxley’s classical zero density estimate can be used to show (see for example [11, Sect. 9.1]) that for any  $A > 0$  and  $X^{1/6+\varepsilon} \leq H \leq X$  we have

$$\left| \sum_{x < n \leq x+H} (\Lambda(n) - 1) \right| \leq H \log^{-A} X \tag{1.5}$$

for all  $A > 0$  and for all  $x \in [X, 2X]$  outside of an exceptional set of measure  $O_{A,\varepsilon}(X \log^{-A} X)$ ; the exponent  $1/6$  was recently improved to  $2/15$  by Guth and Maynard [9]. Moreover, under<sup>3</sup> the Riemann hypothesis, the regime of  $H$  can be extended to  $X^\varepsilon \leq H \leq X$  (in fact, one has a power-saving bound for the left-hand side of (1.5) then). Using the fundamental lemma of the sieve, one can replace the approximant 1 here by the more refined approximant  $\Lambda^\sharp$ , and a modification of the arguments used to prove (1.5) also gives for  $H$  in the same regime the maximal analogue

$$\left| \sum_{x < n \leq x+H} (\Lambda(n) - \Lambda^\sharp(n)) \right|^* \leq H \log^{-A} X \quad (1.6)$$

for all  $x \in [X, 2X]$  outside of an exceptional set of measure  $O_{A,\varepsilon}(X \log^{-A} X)$ .

Next we consider discorrelation of the von Mangoldt function against Fourier phases and nilsequences. In the prequel [30] to this paper we proved that

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{X < n \leq X+H} (\Lambda(n) - \Lambda^\sharp(n)) \overline{F}(g(n)\Gamma) \right|^* \ll_{A,\varepsilon,F,G/\Gamma} H \log^{-A} X \quad (1.7)$$

for  $X^{5/8+\varepsilon} \leq H \leq X^{1-\varepsilon}$  for any filtered nilmanifold  $G/\Gamma$  and Lipschitz function  $F: G/\Gamma \rightarrow \mathbb{C}$ . This generalized the work of Zhan [46] concerning the linear phase case  $\overline{F}(g(n)\Gamma) = e(\alpha n)$  for some  $\alpha \in \mathbb{R}$ .

These are currently the best known discorrelation estimates for  $\Lambda$  that hold for all short intervals. To the best of our knowledge, no further improvement to the range of validity of (1.7) exists in the literature, even in the case of linear phase functions and almost all short intervals.

Let us next discuss the Möbius function, for which stronger results are known. The result (1.6) works with  $\Lambda - \Lambda^\sharp$  replaced by  $\mu$  with a very similar proof, but two of us [21] (see also [20]) established unconditionally that in the much wider regime  $X^\varepsilon \leq H \leq X$  we have

$$\left| \sum_{x < n \leq x+H} \mu(n) \right| \leq H \log^{-c} X$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_\varepsilon(X \log^{-c} X)$  for some absolute constant  $c > 0$ . In fact, in [21] similar results (but with weaker discorrelation and exceptional set bounds) were obtained for any  $H = H(X) \leq X$  that goes to infinity as  $X \rightarrow \infty$ . In the sequel [22], improved bounds were obtained for the measure of the exceptional set of  $x \in [X, 2X]$ .

Concerning correlations with nilsequences, (1.7) works with  $\Lambda$  replaced by  $\mu$ , and in [30] we reached intervals of length  $H \geq X^{3/5+\varepsilon}$  with weaker error terms, strengthening the earlier works [23, 24] concerning polynomial phases.

<sup>3</sup>One can also obtain similar results assuming just the Lindelöf hypothesis or the density hypothesis, though perhaps without the power saving.

The work [29] of four of us together with Ziegler (combined with a routine Fourier expansion to handle the maximal truncation) gives for  $X^\epsilon \leq H \leq X^{1-\epsilon}$  the bound

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H} \mu(n) \overline{F}(g(n)\Gamma) \right|^* = o_{X \rightarrow \infty, \epsilon}(H) \tag{1.8}$$

for all  $x \in [X, 2X]$  outside of a set of measure  $o_{\epsilon, d, F, G/\Gamma}(X)$ , whenever  $G/\Gamma$  is a filtered nilmanifold and  $F: G/\Gamma \rightarrow \mathbb{C}$  is Lipschitz. This implies a qualitative version of Theorem 1.1(iv) where the bound for the measure of the exceptional set is  $o_{X \rightarrow \infty; \epsilon, d, D}(X)$ .

The bound (1.8) was a generalization of work of three of us [28] concerning linear phases. A simplified proof of this linear phase case was recently established by Walsh [42], based on exploiting the phenomenon of ‘‘contagiousness’’ of local frequencies. The proof of (1.8) given in [29] is quite complicated; in this paper we give a new proof of (1.8), using the ideas of Walsh [42] to simplify the arguments, which also gives a slightly stronger outcome in that we obtain better control on the exceptional set.

For applications toward Chowla’s conjecture demonstrated in [39], it is important that there is a supremum in (1.8). However, the  $H$ -range would need to be significantly improved in order to deduce the (logarithmic) Chowla conjecture. See [29] and [43, 44] for some progress on this in the case of polynomial and linear phases, respectively.

The variant of (1.8) without the supremum is easier. Regarding this version, X. He and Z. Wang [12] have established that, for any fixed filtered nilmanifold  $G/\Gamma$ ,  $g \in \text{Poly}(\mathbb{Z} \rightarrow G)$ , and  $F: G/\Gamma \rightarrow \mathbb{C}$  Lipschitz and any  $H = H(X) \leq X$  tending to infinity, we have

$$\left| \sum_{x < n \leq x+H} \mu(n) \overline{F}(g(n)\Gamma) \right|^* = o_{X \rightarrow \infty}(H)$$

for all  $x \in [X, 2X]$  outside of a set of measure  $o(X)$ . This generalized the work of three of us [25] concerning linear phases which in turn generalized [21].

Finally, we review the known results for the divisor functions  $d_k$  for fixed  $k \geq 2$ . As is well known, the mean value of  $d_k$  on  $[1, X]$  is  $\asymp \log^{k-1} X$ . Similarly, for  $X^\epsilon \leq H \leq X$  we have

$$\sum_{X < n \leq X+H} d_k(n) \asymp_\epsilon H \log^{k-1} X;$$

the upper bound part of this follows from Shiu’s bound [35] for sums of multiplicative functions on short intervals, and the lower bound part follows by considering a minorant of  $d_k(n)$  given by restricting  $k - 1$  of the  $k$  variables in the definition of  $d_k(n)$  to be  $\leq X^{\epsilon/(2k)}$ .

For  $k = 2$ , one can use the second moment estimates of Jutila [17] for the divisor problem in short intervals to show that in the regime  $X^\varepsilon \leq H \leq X^{1-\varepsilon}$  one has

$$\left| \sum_{x < n \leq x+H} (d_2(n) - d_2^\sharp(n)) \right| \leq H X^{-c\varepsilon}$$

for some absolute constant  $c > 0$  and for all  $x \in [X, 2X]$  outside a set of measure  $\ll X^{1-c\varepsilon}$ . Improving on work of three of the authors [27], Mangerel [19] and Y.-C. Sun [36] have recently extended the result of [21] on bounded multiplicative functions in short intervals to divisor-bounded multiplicative functions, and in particular Sun’s work can be used to show that, for any fixed  $k \geq 2$ ,  $\varepsilon > 0$  and any  $(\log X)^{k \log k - k + 1 + \varepsilon} \leq H \leq X$ , we have

$$\left| \sum_{x < n \leq x+H} (d_k(n) - d_k^\sharp(n)) \right| = o_{X \rightarrow \infty}(H \log^{k-1} X)$$

for all  $x \in [X, 2X]$  outside of a set of measure  $o_{X \rightarrow \infty}(X)$ .

For correlations of  $d_k$  with polynomial phases, using the approach introduced in [43], M. Wang [45] showed that in the regime  $\exp(C(\log X)^{1/2}(\log \log X)^{1/2}) \leq H \leq X$  with  $C$  large enough, we have for almost all  $x \in [X, 2X]$  the estimate

$$\sup_{\substack{P \in \mathbb{R}[y] \\ \deg(P) \leq s}} \left| \sum_{x < n \leq x+H} (d_k(n) - d_k^\sharp(n)) e(P(n)) \right| = o_s(H \log^{k-1} X).$$

In [45], a different approximant than  $d_k^\sharp$  was used (which is simpler but does not allow as good error terms), but for the sake of comparison we have stated the result with the qualitatively equivalent approximant  $d_k^\sharp$ . The result of [45] thus applies to shorter intervals than Theorem 1.1(v), but is only available for polynomial phases.

Other than the result of [45], the only results we are aware of in previous literature apply to all short intervals, rather than almost all short intervals. In particular, in the prequel [30] to this paper, we showed that for  $X^{\theta_k + \varepsilon} \leq H \leq X^{1-\varepsilon}$  we have

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

for any filtered nilmanifold  $G/\Gamma$  and Lipschitz function  $F : G/\Gamma \rightarrow \mathbb{C}$ , some constant  $c_{k, G/\Gamma} > 0$ , with  $\theta_2 = 1/3$ ,  $\theta_3 = 5/9$ , and  $\theta_k = 5/8$  for  $k \geq 4$ . In particular, the  $k = 2$  case of Theorem 1.1(iii) follows from our earlier work, even without the  $x$ -average.

### 1.2 Gowers uniformity

As in [30], our decorrelation estimates with nilsequences for arithmetic functions in short intervals imply that their Gowers norms over short intervals are small.

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

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

where  $\vec{\omega} = (\omega_1, \dots, \omega_s)$ ,  $|\vec{\omega}| := \omega_1 + \dots + \omega_s$ , and  $\mathcal{C} : z \mapsto \bar{z}$  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

$$\|f\|_{U^s(x, x+H]} := \frac{\|f 1_{(x, x+H]}\|_{U^s(\mathbb{Z})}}{\|1_{(x, x+H]}\|_{U^s(\mathbb{Z})}}$$

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

Using the inverse theorem for the Gowers norms (see Proposition 7.3) and a construction of pseudorandom majorants in Sect. 7, we can deduce a Gowers uniformity estimate from Theorem 1.1. There and later we use the approximant

$$\Lambda_w(n) := \frac{W}{\varphi(W)} 1_{(n, W)=1}, \tag{1.9}$$

with  $W := \prod_{p \leq w} p$  and  $X$  large in terms of  $w$ .

**Theorem 1.3** (Gowers uniformity estimate) *Let  $X \geq H \geq 3$  satisfy  $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$  for some  $0 \leq \theta < 1$  and some fixed  $\varepsilon > 0$ . Let  $s \geq 1$  and  $k \geq 2$  be fixed integers, and let  $A > 0$ . Let  $w \geq 1$  and let  $\Lambda_w$  be as in (1.9), and assume that  $X$  is large enough in terms of  $w$ .*

(i) *If  $\theta = 1/3$ , then*

$$\begin{aligned} & \|\Lambda - \Lambda_w\|_{U^s(x, x+H]} = o_{w \rightarrow \infty}(1), \\ & \max_{\substack{1 \leq a \leq W \\ (a, W)=1}} \left\| \frac{\varphi(W)}{W} \Lambda(W \cdot + a) - 1 \right\|_{U^s(x, x+H]} = o_{w \rightarrow \infty}(1) \end{aligned}$$

*for all  $x \in [X, 2X]$  outside of an exceptional set of measure  $O_{A, \varepsilon}(X \log^{-A} X)$ .*

(ii) *If  $\theta = 0$ , then*

$$\|\mu\|_{U^s(x, x+H]} = o_{X \rightarrow \infty}(1)$$

*for all  $x \in [X, 2X]$  outside of an exceptional set of measure  $O_{A, \varepsilon}(X \log^{-A} X)$ .*

(iii) If  $\theta = 0$ ,  $k \geq 2$  and  $W'$  satisfies  $W \mid W' \mid W^{\lfloor w \rfloor}$ , then

$$\begin{aligned} \|d_k - d_k^\sharp\|_{U^s(x, x+H)} &= o_{X \rightarrow \infty}(\log^{k-1} X), \\ \max_{\substack{1 \leq a \leq W' \\ (a, W')=1}} \left\| \left( \frac{W}{\varphi(W)} \right)^{k-1} d_k(W' \cdot + a) - \frac{\log^{k-1} X}{(k-1)!} \right\|_{U^s(x, x+H)} & \quad (1.10) \\ &= o_{w \rightarrow \infty}(\log^{k-1} X) \end{aligned}$$

for all  $x \in [X, 2X]$  outside of an exceptional set of measure  $O_{A, \varepsilon}(X \log^{-A} X)$ .

Theorem 1.3(ii) was earlier established in [29] but with a qualitative exceptional set bound of  $o_{X \rightarrow \infty; \varepsilon}(X)$ . The argument used to prove Theorem 1.3(ii) (which develops an idea of Walsh [42]) in fact gives a different, simpler proof of the main result in [29].

In the case  $s = 2$ , the inverse theorem for the Gowers norm has polynomial dependencies by a standard Fourier-analytic argument, so one could strengthen Theorem 1.3(i) to

$$\|\Lambda - \Lambda_w\|_{U^2(x, x+H)} \leq w^{-c}$$

with the same conditions, but now in the larger regime  $w \ll_{A, \varepsilon} \log^A X$ . Similarly, Theorem 1.1(iii) gives that for  $X^{1/3+\varepsilon} \leq H \leq X^{1-\varepsilon}$  we have

$$\|d_k - d_k^\sharp\|_{U^2(x, x+H)} \leq X^{-c_{k, \varepsilon}}$$

for some constant  $c_k > 0$  and for all  $x \in [X, 2X]$  outside of a set of measure  $O_\varepsilon(X^{1-c_{k, \varepsilon}})$ .

**Remark 1.4** It is natural to wonder if the quasipolynomial inverse theorem of Leng–Sah–Sawhney [18] would allow quantifying the Gowers norm bounds in Theorem 1.3. While we believe a quantification would in principle be possible, the error terms obtained would likely be rather weak (perhaps triple-logarithmic or so), the largest difficulty being tracking the dimension dependence of the “nilsequence contagion argument” in Sect. 5. Another bottleneck for the Möbius and divisor functions is that, in order to obtain savings of order  $O(\varepsilon)$  using the Turán–Kubilius identity, we need to handle type II sums of length  $X^{\varepsilon'}$  where  $\varepsilon' = \exp(-\varepsilon^{-2})$ . The arguments in Sect. 6 involving iterated “scaling up” of a phase relation are not quantitatively very effective and would limit the speed at which  $\varepsilon$  can go to 0. For these reasons, and in order not to further complicate the paper, we have chosen to keep the Gowers norm results qualitative.

### 1.3 Applications

Theorem 1.3 allows us to study  $\ell$ -point correlations with only one averaging variable — Theorem 1.3(i) combined with the generalized von Neumann theorem and (somewhat involved) sieve-theoretic computations implies the following.

**Theorem 1.5** (*ℓ*-point Hardy–Littlewood with one averaging variable) *Let ℓ ∈ ℕ and ε > 0 be fixed, and let X<sup>1/3+ε</sup> ≤ H ≤ X<sup>1-ε</sup>. Then we have*

$$\sum_{n \leq X} \Lambda(n)\Lambda(n+h) \cdots \Lambda(n+(\ell-1)h) = (1 + o_{X \rightarrow \infty}(1)) \mathfrak{S}(0, h, \dots, (\ell-1)h) X$$

for proportion 1 - o<sub>X→∞</sub>(1) of integers 1 ≤ h ≤ H, where

$$\mathfrak{S}(h_1, \dots, h_\ell) := \prod_p \left( 1 - \frac{|\{h_1, \dots, h_\ell\} \pmod p|}{p} \right) \left( 1 - \frac{1}{p} \right)^{-\ell} \tag{1.11}$$

is the usual singular series.

Allowing *H* to be fixed in Theorem 1.5 would roughly correspond to the Hardy–Littlewood prime tuples conjecture [10]. The coefficients 0, 1, . . . , ℓ - 1 here can be replaced by any other fixed, distinct integers if desired. Previously, the case *H* = *X* and ℓ arbitrary was known by work of Green, the fourth author and Ziegler [5, 6, 8], and the case *H* = *X*<sup>8/33+ε</sup> and ℓ = 2 was known by work of three of the authors [26]. For ℓ ≥ 3 and *H* ≤ *X*<sup>1-ε</sup>, we are not aware of any previous results of this shape in the literature, although using results of part I of this series [30], together with arguments from Sect. 8 of the current paper, it would be possible to deduce the case *H* = *X*<sup>5/8+ε</sup> and ℓ arbitrary. Thus we obtain a significant improvement over what the methods of part I would give.

We get a similar result for the divisor function using Theorem 1.3(iii), the generalized von Neumann theorem and some divisor correlation computations.

**Theorem 1.6** (Divisor correlation conjecture with one averaging variable) *Let k, ℓ ∈ ℕ and ε > 0 be fixed, and let X<sup>ε</sup> ≤ H ≤ X<sup>1-ε</sup>. Then we have*

$$\sum_{n \leq X} d_k(n)d_k(n+h) \cdots d_k(n+(\ell-1)h) = \left( \frac{C(h, k, \ell)}{((k-1)!)^\ell} + o_{X \rightarrow \infty}(1) \right) (\log X)^{k\ell-\ell}$$

for proportion 1 - o<sub>X→∞</sub>(1) of integers 1 ≤ h ≤ H, where

$$C(h, k, \ell) := \prod_p \lim_{Y \rightarrow \infty} \frac{\mathbb{E}_{n \leq Y} d_{k,p}(n)d_{k,p}(n+h) \cdots d_{k,p}(n+(\ell-1)h)}{(\mathbb{E}_{n \leq Y} d_{k,p}(n))^\ell} \tag{1.12}$$

and *d*<sub>*k,p*</sub> is the multiplicative function given on prime powers by *d*<sub>*k,p*</sub>(*q*<sup>*a*</sup>) =  $\binom{k-1+a}{k-1}$  if *q* = *p* and *d*<sub>*k,p*</sub>(*q*<sup>*a*</sup>) = 1 otherwise.

**Remark 1.7** It is possible to evaluate the constants *C*(*h*, *k*, ℓ) explicitly with sufficient effort. See [40] for the evaluation of *C*(*h*, *k*, 2).

Allowing *H* to be fixed in Theorem 1.6, and taking ℓ = 2, would correspond to (leading order) verification of the well known divisor correlation conjecture [26, Conjecture 1.1], [2, Conjecture 3], [15, 41]. Previously, the case *H* = *X* and ℓ arbitrary

was established in the work of Matthiesen [31], and the case  $H = (\log X)^{10000k \log k}$  and  $\ell = 2$  was proven by three of the authors [27]. Moreover, it would be possible to deduce the case  $H = X^{3/5+\varepsilon}$  and  $\ell$  arbitrary from part I of this series [30] combined with Sect. 8 of the current paper. Thus, we again obtain a significant improvement over what methods of part I would give.

#### 1.4 The case of $d_2$ revisited

In the case of  $d_2$  and a fixed linear phase, we prove in Sect. 9 the following result that has better savings than what Theorem 1.1(v) implies:

**Theorem 1.8** (Improved disconnection for  $d_2$  against fixed linear phase) *Let  $X \geq 3$  and  $X^\varepsilon \leq H \leq X^{1-\varepsilon}$  for some  $\varepsilon \in (0, \frac{1}{300})$ . Also let  $\alpha \in \mathbb{R}$ . Then we have*

$$\left| \sum_{x < n \leq x+H} (d_2(n) - d_2^\sharp(n)) e(n\alpha) \right|^* dx \leq H X^{-\varepsilon/1000}$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_\varepsilon(X^{1-\varepsilon/1000})$ .

The proof method of Theorem 1.8 is rather different from that of Theorem 1.1, relying on more classical inputs such as the fourth moment of Dirichlet  $L$ -functions.

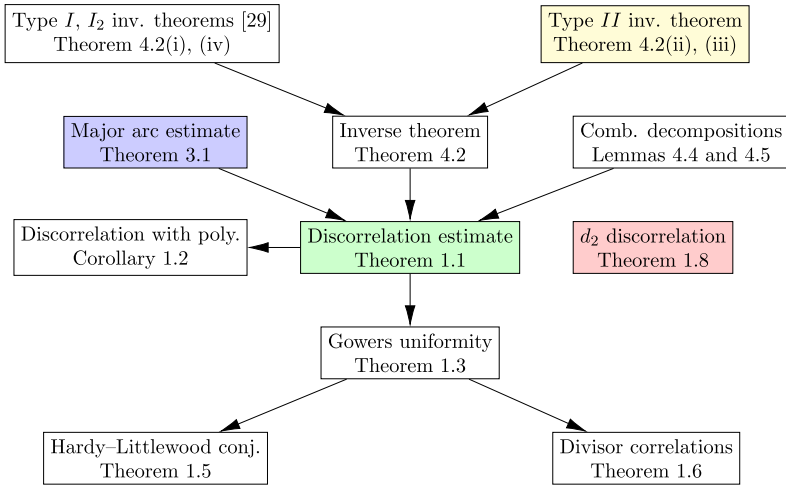
**Remark 1.9** The most notable aspect of Theorem 1.8 is the power-saving bound for the disconnection. It is plausible that, with some work, the methods in [27] could prove a variant of this theorem in which  $d_2$  can be replaced by  $d_k$  for an arbitrary  $k \geq 2$ , and  $H$  is permitted to be as small as  $\log^{C_k} X$  for some  $C_k$  depending on  $k$ , but with the savings  $X^{-\eta}$  significantly weakened to  $\log^{-c_k} X$  for some  $c_k > 0$ . However, we do not attempt to establish this claim here.

#### 1.5 Methods of proof

We now describe (in somewhat informal terms) the general strategy of proof of Theorem 1.1, although for various technical reasons the actual rigorous proof will not quite follow the intuitive plan that is outlined here. See also Fig. 1 for a high-level depiction of the proof and Figs. 2 to 4 for the structure of proofs of several key intermediate theorems. (These figures could also be combined into one giant diagram if desired.)

The first step in the case of Theorem 1.1(i)–(iii), which is standard, is to apply Heath–Brown’s identity (Lemma 2.17) and decompose  $\mu$ ,  $\Lambda$ ,  $d_k$  and their approximants (up to small error terms) into three standard types of sums for divisor-bounded coefficients  $\alpha$ ,  $\beta$ :

- (I) *Type I* sums, which are essentially of the form  $\alpha * 1$  for some arithmetic function  $\alpha: \mathbb{N} \rightarrow \mathbb{C}$  supported on the interval  $[1, HX^{-\varepsilon/2}]$ .
- (I<sub>2</sub>) *Type I<sub>2</sub>* sums, which are essentially of the form  $\alpha * 1 * 1$  for some arithmetic function  $\alpha: \mathbb{N} \rightarrow \mathbb{C}$  supported on the interval  $[1, H^{3/2} X^{-1/2-\varepsilon/2}]$ .



**Fig. 1** Logical relationships used to both prove and apply the main estimate, Theorem 1.1. Theorem 3.1 and Theorem 4.2(ii), (iii) will themselves require lengthy proofs; see Figs. 2 and 4 respectively. Theorem 1.8 is proven by different (and more classical) methods

(II) *Type II* sums, which are of the form  $\alpha * \beta$  for some arithmetic functions  $\alpha, \beta: \mathbb{N} \rightarrow \mathbb{C}$  with  $\alpha$  supported on the interval  $[X^{\varepsilon/2}, HX^{-\varepsilon/2}]$ .

On the other hand, the first step in proving Theorem 1.1(iv)–(v) is to apply an identity of Turán–Kubilius type, which leads only to type I sums in the range  $[1, X^{\varepsilon/2}]$  and to type II sums with one variable in the range  $[X^\kappa, X^{\varepsilon/2}]$ , with  $\kappa > 0$  an arbitrarily small constant. We formalize these decompositions as Lemmas 4.4 and 4.5.

We then split our considerations to the *major arc case* in which the nilsequence  $F(g(n)\Gamma)$  is more-or-less replaced by  $n^{iT}$  for some  $T = X^{O(1)}$ , and the *minor arc case* in which  $F$  has mean zero and  $g(n)\Gamma$  is highly equidistributed in the nilmanifold  $G/\Gamma$ .

The major arc case of Theorem 1.1 is formalized as Theorem 3.1, and is proven in Sect. 3. Here we use Parseval’s identity to relate the variance of the short interval sum to Dirichlet polynomial estimates. We split this case into two subcases: the case of small  $|T|$  (i.e.,  $|T| \leq X^{1+\varepsilon/2}/H$ ) and the case of large  $|T|$  (i.e.,  $|T| > X^{1+\varepsilon/2}/H$ ). If  $|T|$  is small, then  $n^{iT}$  does not vary much on intervals of the form  $[x, x + HX^{-\varepsilon/2}]$ , which allows us to effectively reduce to the case  $T = 0$ . This case could be deduced from known estimates on arithmetic functions in almost all short intervals, although with the large  $|T|$  case in mind we found it convenient to reprove these estimates. If instead  $|T|$  is large, we split into type I and II sums in the usual fashion. The type II sums are handled using Dirichlet polynomial estimates of Baker–Harman–Pintz [1]. After this, we are left with handling only a very narrow type I range where we have a smooth variable of length  $> X^{1-\varepsilon}$ . This case in turn can be handled by Taylor expanding  $n^{iT} = e(\frac{T}{2\pi} \log n)$  on a short interval around  $x$  as a polynomial exponential phase and using van der Corput’s exponential sum bound to show that the sequence  $(\frac{T}{2\pi x} \bmod 1)_{x \in [X, 2X]}$  (corresponding to the linear term in the Taylor expansion) equidistributes.

To prove Theorem 1.1(i)–(iii) in the remaining, minor arc, cases, in case of type  $I$  and type  $I_2$  sums, we simply use our estimates from [30]. Unfortunately, for these contributions we have not found a way to use the additional averaging over  $x$  to extend the range where  $\alpha$  can be supported.

The remaining task is then to handle the contribution of type  $II$  minor arc sums. Let us consider sums of the type

$$\sum_{\substack{x < ab \leq x+H \\ A_{II} < a \leq 2A_{II}}} \alpha(a)\beta(b)\overline{F}(g_x(ab)\Gamma), \quad (1.13)$$

with  $H^{\varepsilon/2} \leq A_{II} \leq HX^{-\varepsilon/2}$ , where  $g_x: \mathbb{Z} \rightarrow G$  is a polynomial sequence, for all  $x \in [X, 2X]$  outside a small exceptional set. To treat these sums, we can use a Fourier decomposition and the equidistribution theory of nilmanifolds to reduce (roughly speaking) to treating the following two special cases of these sums:

- *Abelian type II minor arc sums* in which  $F(g_x(n)\Gamma) = e(P_x(n))$  is a polynomial phase that does not “look like” a character  $n^{iT}$  (or more generally  $\chi(n)n^{iT}$  for some Dirichlet character  $\chi$  of small conductor) in the sense that the Taylor coefficients of  $e(P_x(n))$  around  $X$  do not align with the corresponding Taylor coefficients of such characters.
- *Nonabelian type II minor arc sums*, in which  $g_x(n)\Gamma$  is highly equidistributed in the nilmanifold  $G/\Gamma$ , which is arising from a nonabelian 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).

To handle both these cases, we extend the methodology of Walsh [42] to develop two tools: a theory of “nilsequence contagion” and a combinatorial “scaling” argument.

To state the nilsequence contagion result (Theorem 5.11) that we establish, we need to introduce a relation  $\sim_{I,\eta}$  between polynomial sequences on  $G$ , where  $I$  is an interval and  $\eta \in (0, 1)$ . We say that  $g_1 \sim_{I,\eta} g_2$  for two polynomial sequences  $g_1, g_2: \mathbb{Z} \rightarrow G$  if  $g_1 = \varepsilon g_2 \gamma$  for some  $(\eta^{-1}, I)$ -smooth polynomial sequence  $\varepsilon$  and some  $\eta^{-1}$ -rational polynomial sequence  $\gamma$ . We are then interested in what can be said of a family of polynomial maps  $(g_n)_{n \in [P, 2P] \cap \mathbb{N}}$ ,  $g_n: \mathbb{Z} \rightarrow G$  if we have the relation

$$g_n(n'\cdot) \sim_{\frac{1}{nm'} I, \eta} g_{n'}(n\cdot) \quad (1.14)$$

for proportion  $\geq \eta$  of pairs  $(n, n') \in ([P, 2P] \cap \mathbb{N})^2$ . We claim that, under suitable conditions on the parameters  $P, |I|, \eta$ , there must exist a polynomial sequence  $g_*$  such that

$$g_n \sim_{\frac{1}{n} I, \eta} g_*(n\cdot)$$

for  $\gg \eta^{O(1)}P$  choices of  $n \in [P, 2P] \cap \mathbb{N}$  (and these indeed are clearly solutions to (1.14) if we adjust the value of  $\eta$  by a polynomial amount).

Walsh showed in [42] (see also [43, 44]) that an analogous claim holds in the case where  $g_n(y) = \alpha_n y$  is a linear polynomial on  $\mathbb{R}/\mathbb{Z}$ , and this was one key ingredient

that enabled him to give a simpler proof of the main result of [28]. Our result extends this to the case of arbitrary polynomial sequences on Lie groups and in addition, importantly for the proof of Theorem 1.1, our result allows extra flexibility already in the case of linear polynomials (we can improve the dependencies between the different parameters  $P, |I|, \eta$ , resulting in improved exponents for Theorem 1.1).

The proof of the nilsequence contagion result involves first showing a version for classical polynomials (Theorem 5.2) by some combinatorial arguments that give iteratively stronger conclusions until we reach the desired conclusion. We then prove the general case by considering a certain two-variable family of subspaces of  $\text{Poly}(\mathbb{Z} \rightarrow G)$  and performing a double induction with respect to this family, where the starting case is trivial, the final case is the full nilsequence contagion result, and the passage from one case to the next is enabled by the polynomial case.

The scaling argument executed in Sect. 6 is a combinatorial argument that together with the nilsequence contagion lemma shows that if the type II sum (1.13) is large, then the phase function  $x \mapsto g_x$ , initially defined on  $[X, 2X]$ , can be extended to a larger scale  $[XA_{II}^K, C_K XA_{II}^K]$  for any fixed  $K$ , such that the phases  $g_x$  at scale  $X$  and the phases  $g_z$  at scale  $XA_{II}^K$  are related by an “approximate functional equation” roughly of the form

$$g_z(m \cdot) \sim \frac{1}{m} (z, z + HA_{II}^K, (\log X)^{-O_K(1)} g_x$$

for “many”  $m \asymp A_{II}^K$ . Since  $A_{II} \geq H^{\epsilon/2} \geq X^\kappa$  for some constant  $\kappa > 0$ , in all cases of Theorem 1.1 we can find  $K \ll_\epsilon 1$  such that  $A_{II}^K \geq X$ ; thus, we can reach a very large scale where the phase functions “communicate” with those at scale  $[X, 2X]$ . We then employ a technical proposition (Proposition 2.16), related to the non-abelian type II estimate established in the prequel paper [30], to show that the functional equation (1.14) has no solutions in the non-abelian minor arc case where  $G$  is non-abelian with one-dimensional center and  $g_x \Gamma$  is highly equidistributed. In the complementary abelian case, we can reduce to  $G/\Gamma = \mathbb{R}/\mathbb{Z}$ . We then show that the only solutions are essentially those where  $g_x$  is the truncated Taylor series around  $x$  of the function  $n \mapsto T \log n$  for some  $T \ll X^{O(1)}$ ; for this a key ingredient is Proposition 2.5 that relates approximately dilation-invariant polynomials to Taylor polynomials of the logarithm function. This then brings us back to the major arc case that was discussed above.

**Remark 1.10** There is an obstacle in improving the exponent  $\theta$  in Theorem 1.1(i) below  $\theta = 1/3$ . Namely, for  $H = X^{1/3-\epsilon}$ , we do not know how to handle  $d_3$ -type sums

$$\sum_{\substack{X < abc \leq X+H \\ X^{1/3-\epsilon/2} < a, b, c \leq X^{1/3+\epsilon/2}}} \overline{F}(g_x(abc)\Gamma).$$

arising from Heath-Brown’s identity. Our type II estimates are not applicable anymore since none of the variables is supported on numbers  $\leq HX^{-\eta}$  for any  $\eta > 0$ . On the other hand, our type  $I_2$  estimates from our previous paper [30] rely on a certain decomposition of hyperbolic neighborhoods into arithmetic progressions which is only non-trivial in the regime  $\theta \geq 1/3$ .

## 1.6 Notation

We use  $Y \ll Z$ ,  $Y = O(Z)$ , or (when  $Y \geq 0$ )  $Z \gg Y$  to denote the estimate  $|Y| \leq CZ$  for some constant  $C$ . If we permit this constant to depend on one or more parameters we shall indicate this by appropriate subscripts, thus for instance  $O_{\varepsilon, A}(Z)$  denotes a quantity bounded in modulus by  $C_{\varepsilon, A}Z$  for some quantity  $C_{\varepsilon, A}$  depending only on  $\varepsilon$ ,  $A$ . We write  $Y \asymp Z$  for  $Y \ll Z \ll Y$ , and we write  $y \sim Y$  for  $Y < y \leq 2Y$ . When working with  $d_k$ , all implied constants are permitted to depend on  $k$ . We also write  $X = o_{w \rightarrow \infty}(Y)$  (respectively  $X = o_{\delta \rightarrow 0}(Y)$ ) to denote an estimate of the form  $|X| \leq c(w)Y$  where  $c(w) \rightarrow 0$  as  $w \rightarrow \infty$  (respectively,  $|X| \leq c(\delta)Y$  where  $c(\delta) \rightarrow 0$  as  $\delta \rightarrow 0$ ); again, if the implied decay rate  $c$  needs to depend on additional parameters, we indicate this with appropriate subscripts.

If  $x$  is a real number (respectively 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 (respectively to zero). We let  $\{x\}$  be the *signed* fractional part of  $x$ , that is to say the unique real number in  $(-1/2, 1/2]$  that is equal to  $x$  modulo 1. In particular  $\|x\|_{\mathbb{R}/\mathbb{Z}} = |\{x\}|$ . We use  $\log$  to denote the natural logarithm function  $x \mapsto \log x$ .

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  $1_S(n) := 1_{n \in S}$ .

We write  $|E|$  for the cardinality of a finite set  $E$ . If  $I$  is an interval, we use  $|I|$  to denote its length. We write  $\text{meas}(\mathcal{B})$  for the Lebesgue measure of a set  $\mathcal{B} \subset \mathbb{R}$ . If  $\vec{\omega} = (\omega_1, \dots, \omega_s) \in \{0, 1\}^s$  is a vector, we write  $|\vec{\omega}| := \omega_1 + \dots + \omega_s$ .

Unless otherwise specified, all sums range over natural number values, except for sums over  $p$  which are understood to range over primes. We use  $d \mid n$  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  $(f * g)(n) := \sum_{d \mid n} f(d)g(n/d)$  to denote the Dirichlet convolution of two arithmetic functions  $f, g: \mathbb{N} \rightarrow \mathbb{C}$ . We use  $d \mid W^\infty$  to denote the claim that  $d \mid W^m$  for sufficiently large  $m$ , or equivalently that all prime factors of  $d$  divide  $W$ . We also sometimes write  $P(w) = \prod_{p < w} p$ .

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

## 2 Basic tools

In this section we state various basic tools and definitions that we will need in this paper. Most of these are quoted from our previous paper [30], but some (in particular, Proposition 2.5) were not present in that work.

### 2.1 Total variation

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

**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

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

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

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

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

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

Furthermore, from the identity  $ab - a'b' = (a - a')b + (b - b')a'$  we see that

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

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

We recall some basic properties of maximal summation:

**Lemma 2.2** (Basic properties of maximal sums) [30, Lemma 2.2]

- (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

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

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

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

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

and more generally

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

for any  $q \geq 1$ .

### 2.2 Vinogradov lemma

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

$$\|f\|_{C^d(I)} := \sup_{0 \leq j \leq d} \sup_{n \in I} |I|^j \|\partial_1^j f(n)\|_{\mathbb{R}/\mathbb{Z}} \tag{2.1}$$

where  $\partial_1$  is the difference operator<sup>4</sup>  $\partial_1 f(n) := f(n + 1) - f(n)$ . We make a similar definition if  $f$  takes values in  $\mathbb{R}$  instead of  $\mathbb{R}/\mathbb{Z}$  (where now we replace  $\|\cdot\|_{\mathbb{R}/\mathbb{Z}}$  by the absolute value). In the case where  $f$  is a polynomial of degree  $d$ , we have  $\partial_1^j f = 0$  for all  $j > d$ , and in this case we write  $\|f\|_{C^d(I)}$  as  $\|f\|_{C^\infty(I)}$ . We remark that this definition of  $\|f\|_{C^\infty(I)}$  deviates slightly from that in [7, Definition 2.7]; in particular, we allow the index  $j$  to equal zero and we allow  $n$  to range over  $I$  instead of setting  $n$  to equal 0. If  $I = [x, x + H]$  with  $x, H \geq 1$  and  $f(n) = \sum_{k=0}^d \alpha_k \binom{n-n_0}{k}$  for some integer  $n_0 \in I$ , then

$$\|f\|_{C^\infty(I)} \asymp_d \max_{0 \leq j \leq d} H^j \|\alpha_j\|_{\mathbb{R}/\mathbb{Z}}. \tag{2.2}$$

The lower bound is clear (specialize  $n$  to  $n_0$  in (2.1)). The upper bound comes from using the Taylor expansion identity

$$\partial_1^j f(n) = \sum_{k=j}^d \binom{n-n_0}{k-j} \partial_1^k f(n_0)$$

and the triangle inequality.

If  $f$  is a polynomial of degree  $d$ , for any intervals  $I \subset I'$  we have

$$\|f\|_{C^\infty(I)} \leq \|f\|_{C^\infty(I')} \ll_d \left( \frac{|I'|}{|I|} \right)^d \|f\|_{C^\infty(I)}; \tag{2.3}$$

<sup>4</sup>In our previous paper [30] we used the convention  $\partial_1 f(n) = f(n) - f(n - 1)$  instead, but this makes little difference to the final statements, and we have switched to this convention to align with other literature in the subject.

indeed, the first inequality is trivial, and for the second one we can use (2.2). We also claim the dilation property

$$\|f(a \cdot)\|_{C^\infty(\frac{1}{a}I)} \ll_d \|f\|_{C^\infty(I)} \tag{2.4}$$

for any  $1 \leq a \leq |I|$ , where here and later  $\frac{1}{a}I := \{\frac{t}{a} : t \in I\}$ . In view of (2.2), it suffices to show that for every  $0 \leq k \leq d$ , the dilated binomial coefficient  $\binom{an}{k}$  has an expansion of the form  $\sum_{j=0}^k c_{j,k} \binom{n}{j}$  with  $c_{j,k} = O_d(a^k)$ . But this is clear since  $\binom{an}{k}$  is a degree  $k$  polynomial in  $n$  with coefficients  $O_d(a^k)$ .

We have the following convenient form of the Vinogradov lemma:

**Lemma 2.3** (Vinogradov lemma) [30, Lemma 2.3] *Let  $\varepsilon, \delta \in (0, 1/2)$ ,  $d \geq 0$ , and let  $P : \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  be a polynomial of degree at most  $d$ . Let  $I$  be an interval with  $|I| \geq 1$ , and suppose that*

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

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

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

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

As a useful corollary of Lemma 2.3 we have, roughly speaking, that if many dilates of a polynomial are smooth, then the polynomial itself is smooth, providing a converse of sorts to (2.4).

**Corollary 2.4** (Concatenating dilated smoothness) [30, Corollary 2.4] *Let  $\delta \in (0, 1/2)$ ,  $d \geq 0$ , 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*

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

for at least  $\delta A$  integers  $a$  in  $[A, 2A]$ . Then either we have  $|I| \ll_d \delta^{-O_d(1)} A$ , or else

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

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

For the proof of our type II estimates, we also need the following proposition (which can be seen as an abstraction of [23, Proposition 2.2]) that relates polynomials  $P$  satisfying a certain dilation invariance to multiples  $T \log$  of the logarithm function.

**Proposition 2.5** (Approximately dilation-invariant polynomials behave like the logarithm) *Let  $d \geq 0$ . Let  $X \geq H$ ,  $A \geq 2$  and  $0 < \delta \leq \frac{1}{\log X}$  with  $H \geq \delta^{-C_d} \max(A, X/A)$  for a large enough constant  $C_d$ . Let  $P : \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  be a polynomial of degree*

*d.* Cover  $(A, 2A]$  by  $O(X/H)$  intervals  $I_{A'} = (A', A'(1 + \frac{H}{X})]$  with  $A \leq A' \leq 2A$ , with each point belonging to at most  $O(1)$  of these intervals. Suppose that for at least  $\delta X/H$  of the intervals  $I_{A'}$  there exist at least  $\delta|I_{A'}|^2$  integer pairs  $(a_1, a_2) \in I_{A'}^2$  for which we have

$$\|P(a_1 \cdot) - P(a_2 \cdot)\|_{C^\infty(X/A', X/A'+H/A')} \leq \frac{1}{\delta}. \tag{2.5}$$

Then there exist a real number  $T$  satisfying  $|T| \ll_d \delta^{-O_d(1)}(X/H)^{d+1}$  and an integer  $1 \leq q \ll_d \delta^{-O_d(1)}$  such that

$$\|qP(\cdot) - T \log(\cdot)\|_{C^d(X, X+H)} \ll_d \delta^{-O_d(1)}.$$

Note that the function  $T \log(\cdot)$  is dilation-invariant up to constants:

$$T \log(a_1 \cdot) - T \log(a_2 \cdot) = T \log \frac{a_1}{a_2}.$$

As a consequence, any polynomial  $P$  which is close to  $T \log$  in  $C^d(X, X + H]$  norm would also obey an estimate similar to (2.5) for all  $a_1, a_2 \in I_{A'}$ , thanks to (2.4). The above proposition can be viewed as a sort of converse to this assertion.

**Proof** We may assume that  $\delta > 0$  is smaller than any fixed absolute constant. We allow implied constants to depend on  $d$ . By rounding, we may assume that  $X$  is an integer.

By the pigeonhole principle, we can find  $a_2 \in I_{A'}$  and a subinterval  $I'_{A'} = (A'', A''(1 + \delta^3 \frac{H}{X})]$  such that (2.5) holds for  $\gg \delta^4 |I_{A'}|$  choices of  $a_1 \in I'_{A'}$ . For each such  $a_1$ , we write

$$P(a_1 n) - P(a_2 n) - (P(a_1 X/A'') - P(a_2 X/A'')) = \sum_{j=1}^d \beta_j(a_1, a_2)(n - X/A'')^j$$

for some real numbers  $\beta_j(a_1, a_2)$ . From (2.5) and (2.2) we see that  $(H/A')^j \|d! \beta_j(a_1, a_2)\|_{\mathbb{R}/\mathbb{Z}} \ll 1/\delta$  for all  $1 \leq j \leq d$ . By the triangle inequality, this implies that for all  $n = X/A'' + O(\delta^2 H/A')$  we have

$$\|d!(P(a_1 n) - P(a_2 n) - (P(a_1 X/A'') - P(a_2 X/A'')))\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{1}{\delta} \sum_{j=1}^d \delta^{2j} \ll \delta.$$

We conclude that for any  $a_1$  as above we have

$$\left| \sum_{X < a_1 n \leq X + \delta^2 H} e(d!P(a_1 n) - d!P(a_2 n)) \right| \gg \delta^2 H/A. \tag{2.6}$$

Summing over the possible choices of  $a_1$ , this implies that

$$\left| \sum_{a_1 \in I'_A} \sum_{X < a_1 n \leq X + \delta^2 H} \alpha_{a_1} \beta_n e(d! P(a_1 n)) \right| \gg \delta^6 H$$

for some bounded coefficients  $\alpha_{a_1}, \beta_n$ .

We can then invoke [23, Proposition 2.2] and conclude that if we write

$$P(y) = \sum_{j=0}^d \alpha_j (y - X)^j \pmod{1}$$

with  $\alpha_j \in \mathbb{R}/\mathbb{Z}$ , then for some integer  $1 \leq q \ll \delta^{-O(1)}$  we have

$$\|q(j\alpha_j + X(j + 1)\alpha_{j+1})\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)} X/H^{j+1}, \tag{2.7}$$

for all  $1 \leq j \leq d$ , with the convention that  $\alpha_{d+1} = 0$ . By backward induction on  $j$ , we may thus lift from  $\mathbb{R}/\mathbb{Z}$  to  $\mathbb{R}$  and write  $(-1)^j q j \alpha_j = \beta_j \pmod{1}$  for some real numbers  $\beta_j$  obeying

$$\beta_j - X\beta_{j+1} = O(\delta^{-O(1)} X/H^{j+1})$$

for all  $j = 1, \dots, d$ , with the convention that  $\beta_{d+1} = 0$ . If we write  $\beta_1 = T/X$  for a real number  $T$ , then a forward induction in  $j$  gives

$$\beta_j = \frac{T}{X^j} + O(\delta^{-O(1)}/H^j)$$

for all  $1 \leq j \leq d + 1$ ; in particular substituting  $j = d + 1$  we have  $T \ll \delta^{-O(1)} \times (X/H)^{d+1}$ . By construction of  $\beta_j$ , we now have

$$qd! \alpha_j = d! T \frac{(-1)^j}{j X^j} + O(\delta^{-O(1)}/H^j) \pmod{1}$$

for  $j = 1, \dots, d$ .

Let  $\partial_1 f(h) = f(h) - f(h - 1)$  be the differencing operator in the  $h$  variable. Then, by the Taylor series of the logarithm function, for all  $h = O(H)$  and  $1 \leq i \leq d$  we have

$$\begin{aligned} \partial_1^i [d! q P(X + h)] &= \sum_{j=1}^d d! q \alpha_j \partial_1^i h^j \\ &= \sum_{j=1}^d \frac{d! (-1)^j}{j} \frac{T}{X^j} \partial_1^i h^j + O(\delta^{-O(1)}/H^i) \pmod{1} \\ &= \sum_{j=1}^{\infty} \frac{d! (-1)^j}{j} \frac{T}{X^j} \partial_1^i h^j + O(\delta^{-O(1)}/H^i) \pmod{1} \end{aligned}$$

$$\begin{aligned}
&= \partial_1^i [d!T \log(1 + h/X)] + O(\delta^{-O(1)}/H^i) \pmod{1} \\
&= \partial_1^i [d!T \log(X + h)] + O(\delta^{-O(1)}/H^i) \pmod{1}.
\end{aligned}$$

Making the substitution  $n := X + h$ , we conclude that

$$\|d!qP(\cdot) - d!T \log(\cdot)\|_{C^d(X, X+H)} \ll \delta^{-O(1)}$$

and the claim follows after some relabeling.  $\square$

### 2.3 Equidistribution on nilmanifolds

We now recall some of the basic notation and results from [7] on equidistribution of polynomial maps on nilmanifolds.

**Definition 2.6** (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) together 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  $g(n) = g_0 g_1^{\binom{n}{1}} \dots g_d^{\binom{n}{d}}$  where  $g_i \in G_i$  for all  $0 \leq i \leq d$ ; the collection of such maps will be denoted by  $\text{Poly}(\mathbb{Z} \rightarrow G)$ .

The well known Lazard–Leibman theorem (see e.g. [7, Proposition 6.2]) states that  $\text{Poly}(\mathbb{Z} \rightarrow G)$  is a group under pointwise multiplication; also, from [7, Corollary 6.8] it follows 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 formula<sup>5</sup> it follows that the exponential map  $\exp: \log G \rightarrow G$  is a homeomorphism and hence has an inverse  $\log: G \rightarrow \log G$ .

**Definition 2.7** (Filtered nilmanifolds) Let  $d, D \geq 1$  and  $0 < \delta < 1$ . A *filtered nilmanifold*  $G/\Gamma$  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  $\Gamma$ ) 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:

<sup>5</sup>More precisely, the Baker–Campbell–Hausdorff formula, which has only finitely many non-zero terms in the connected case, provides an associative polynomial multiplication law on  $\log G$  that agrees locally with the multiplication law on  $G$  after conjugating by the exponential map, thus providing a local isomorphism of Lie groups; see e.g., [38, §1.2, 2.4]. As  $\log G$  and  $G$  are both simply connected, one can extend this local isomorphism to a global one by monodromy.

- (i) 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  $1/\delta$ .
- (ii) For all  $0 \leq i \leq D$ , the vector space  $G_i$  is spanned by the  $X_j$  with  $D - \dim G_i < j \leq D$ .
- (iii) We have  $\Gamma = \{\exp(n_1 X_1) \cdots \exp(n_D X_D) : n_1, \dots, n_D \in \mathbb{Z}\}$ .

One easily sees that  $G/\Gamma$  has the structure of a smooth compact  $D$ -dimensional manifold, to which we can associate 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  $G/\Gamma$  by the formula  $d_{G/\Gamma}(x, y) := \inf_{g \in \Gamma} d_G(x, h\Gamma = y)$ . The Lipschitz norm of a function  $F : G/\Gamma \rightarrow \mathbb{C}$  is defined as

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

A horizontal character  $\eta$  associated to a filtered nilmanifold is a continuous homomorphism<sup>6</sup>  $\eta : G \rightarrow \mathbb{R}$  that maps  $\Gamma$  to the integers.

An element  $\gamma$  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  $G'/\Gamma'$  of  $G/\Gamma$  (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 being rational numbers of height at most  $M$ .

One easily sees 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  $O_{d,D}(\delta^{-O_{d,D}(1)})$ .

**Remark 2.8** In particular, if  $G = \mathbb{R}$  and  $\Gamma = \mathbb{Z}$ , all horizontal characters are of the form  $y \mapsto ry$  with  $r \in \mathbb{Z}$ .

From several applications of Baker–Campbell–Hausdorff formula (see e.g. [29, Appendix B]) we see that if  $G$  has degree at most  $d$  and  $\gamma_1, \gamma_2$  are  $M$ -rational, then  $\gamma_1 \gamma_2$  is  $O_d(M^{O_d(1)})$ -rational.

The following theorem gives a basic dichotomy between equidistribution and smoothness:

**Theorem 2.9** (Quantitative Leibman theorem) [30, Theorem 2.7] *Let  $0 < \delta < 1/2$ , let  $d, D \geq 1$ , let  $I$  be an interval with  $|I| \geq 1$ , and let  $G/\Gamma$  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 of mean zero (i.e.,  $\int_{G/\Gamma} F d\mu_{G/\Gamma} = 0$ ).*

<sup>6</sup>This is a lift of the convention in [7], in which a horizontal character is a continuous homomorphism from  $G$  to  $\mathbb{R}/\mathbb{Z}$  that annihilates  $\Gamma$ .

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  $\eta : G \rightarrow \mathbb{R}$  of Lipschitz norm  $O_{d,D}(\delta^{-O_{d,D}(1)})$  such that

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

Let  $G/\Gamma$  be a filtered nilmanifold of dimension  $D$  and complexity at most  $1/\delta$ . 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  $1/\delta$ . In [7, Proposition A.10] it is shown that  $G'/\Gamma'$  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  $G'/\Gamma'$  as a subnilmanifold of  $G/\Gamma$ .

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  $\eta$  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.

If we quotient out a filtered nilmanifold by a normal subgroup, we obtain another filtered nilmanifold, with polynomial bounds on complexity:

**Lemma 2.10** (Quotienting by a normal subgroup) [30, Lemma 2.8] *Let  $G/\Gamma$  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*

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

for any  $g, h \in G$ .

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*  $\xi$  if one has the identity

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

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

We have the following convenient decomposition:

**Proposition 2.11** (Central Fourier approximation) [30, Proposition 2.9] *Let  $d, D \geq 1$  and  $0 < \delta < 1$ . Let  $G/\Gamma$  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  $\xi$  ranges over central frequencies of Lipschitz norm at most  $O_{d,D}(\delta^{-O_{d,D}(1)})$ , and each  $F_{\xi}$  has Lipschitz norm  $O_{d,D}(\delta^{-O_{d,D}(1)})$  and oscillates with central frequency  $\xi$ . Furthermore, if  $F$  has mean zero, then the same holds for all of the  $F_{\xi}$ .

We now recall some properties of polynomial sequences.

**Definition 2.12** (Smoothness, total equidistribution, rationality) Let  $G/\Gamma$  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  $G/\Gamma$  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 for some  $1 \leq r \leq M$  one has  $g(n)^r \in \Gamma$  for each  $n \in \mathbb{Z}$ .

From Taylor expansion and the Baker–Campbell–Hausdorff formula it is not difficult to see that if  $G/\Gamma$  has degree at most  $d$  and  $g$  is  $M$ -rational, then the map  $n \mapsto g(n)\Gamma$  is  $q$ -periodic for some period  $1 \leq q \ll_d M^{O_d(1)}$ ; see, e.g., [7, Lemma A.12].

We also have the following basic facts about smooth and rational polynomials for any interval  $I$ :

**Lemma 2.13** *Suppose that  $G/\Gamma$  is a filtered nilmanifold of degree and dimension  $O(1)$  and complexity at most  $M$  for some  $M \geq 1$ .*

- (i) *(Smooth polynomials form an approximate group) The constant polynomial 1 is  $M$ -smooth on  $I$ , and if  $g, h$  are  $M$ -smooth on  $I$ , then  $g^{-1}$  and  $gh$  are  $M^{O(1)}$ -smooth.*
- (ii) *(Rational polynomials form an approximate group) The constant polynomial 1 is  $M$ -rational, and if  $g, h$  are  $M$ -rational, then  $g^{-1}$  and  $gh$  are  $M^{O(1)}$ -rational.*

**Proof** Claim (i) follows from the triangle inequality, the right-invariance of the metric  $d_G$ , and the approximate left-invariance of  $d_G$  (see [7, Lemma A.5]). Claim (ii) is [7, Lemma A.11(v)].  $\square$

**Lemma 2.14** [30, Lemma 2.11] *Let  $d, D \geq 1$  and  $0 < \delta < 1$ . Let  $G/\Gamma$  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*

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

for some non-trivial horizontal character  $\eta: G \rightarrow \mathbb{R}$  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)  $\varepsilon$  is  $(\delta^{-O_{d,D}(1)}, I)$ -smooth;
- (ii)  $g'$  takes values in  $G' = \ker \eta$ ;
- (iii)  $\gamma$  is  $\delta^{-O_{d,D}(1)}$ -rational.

We also rely on the following large sieve inequality for nilsequences, which is a more quantitative variant of the one in [29, Proposition 4.11] and will be deduced from [30, Proposition 2.15].

**Proposition 2.15** (Large sieve) *Let  $d, D \geq 1$  and  $0 < \delta < 1$ . Let  $G/\Gamma$  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  $I$  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  $\xi$ . Let  $f: \mathbb{Z} \rightarrow \mathbb{C}$  be a function with  $\sum_{n \in I} |f(n)|^2 \leq \frac{1}{8}|I|$ . Then there exist  $K \ll_{d,D} \delta^{-O_{d,D}(1)}$  and elements  $g_1, \dots, g_K \in \text{Poly}(\mathbb{Z} \rightarrow G)$  such that the following holds. Whenever  $g \in \text{Poly}(\mathbb{Z} \rightarrow G)$  satisfies*

$$\left| \sum_{n \in I} f(n) \overline{F}(g(n)\Gamma) \right|^* \geq \delta |I|, \quad (2.8)$$

at least one of the following holds.

- (i) *There exists a non-trivial horizontal character  $\eta: G \rightarrow \mathbb{R}$  having Lipschitz norm  $O_{d,D}(\delta^{-O_{d,D}(1)})$  such that  $\|\eta \circ g\|_{C^\infty(I)} \ll_{d,D} \delta^{-O_{d,D}(1)}$ .*

(ii) *There exists  $i \in \{1, \dots, K\}$  such that we can write*

$$g = \varepsilon g_i \gamma$$

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

Informally speaking, this proposition asserts that a given function  $f$  can only correlate with a bounded number of essentially distinct non-trivial nilsequences  $F(g(\cdot)\Gamma)$ ; it is a higher-order analogue of the classical large sieve inequality, which asserts (roughly speaking) that  $f$  will only have a bounded number of essentially distinct frequencies where the Fourier transform is large.

**Proof** This follows quickly from [30, Proposition 2.15]. We may assume that  $\delta < 1/2$ . Let  $C$  be large enough in terms of  $d, D$ , and let  $C'$  be large enough in terms of  $C$ . Let  $\mathcal{G} = \{g_1, \dots, g_J\}$  be a maximally large collection of elements of  $\text{Poly}(\mathbb{Z} \rightarrow G)$  such that

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

for all  $1 \leq j \leq J$  and such that for any  $h, h' \in \mathcal{G}$  with  $h \neq h'$  we have the following properties:

- ( $P_C$ ) For any non-trivial horizontal character  $\eta: G \rightarrow \mathbb{R}$  of Lipschitz norm  $\leq \delta^{-C}$  we have  $\|\eta \circ h \bmod 1\|_{C^\infty(I)} > \delta^{-C}$ ;
- ( $Q_C$ ) If  $h = \varepsilon h' \gamma$  with  $\varepsilon, \gamma \in \text{Poly}(\mathbb{Z} \rightarrow G)$ , then either  $\varepsilon$  is not  $(\delta^{-C}, I)$ -smooth or  $\gamma$  is not  $\delta^{-C}$ -rational.

We may assume that the collection  $\{g_1, \dots, g_J\}$  is non-empty. If  $C$  is large enough, by [30, Proposition 2.15] we have  $J \leq \delta^{-C}$ . Now, let  $g \in \text{Poly}(\mathbb{Z} \rightarrow G)$  be any polynomial sequence satisfying (2.8). If  $g$  fails property  $P_C$ , we are done. If the pair  $(g, g_{j'})$  fails property  $Q_{C'}$  for some  $1 \leq j' \leq J$ , we are again done. Finally, we are left with the case where  $g$  satisfies  $P_C$  and  $(g, g_{j'})$  satisfies  $Q_{C'}$  for any  $1 \leq j' \leq J$  (and therefore by Lemma 2.13 also  $(g_j, g)$  satisfies  $Q_C$  for any  $1 \leq j \leq J$ , thanks to the assumption that  $C'$  is large in terms of  $C$ ). But since property  $Q_{C'}$  implies property  $Q_C$ , this contradicts the maximality of the collection  $\{g_1, \dots, g_J\}$ . The claim follows. □

The following proposition from [30] will be an important ingredient in the proof of our type II estimate, although it will be applied in a different context to the one in [30], namely in order to analyze the outcome of Walsh’s contagion argument. It will play a role in the nonabelian type II theory similar to the role Proposition 2.5 will play in the abelian type II theory.

**Proposition 2.16** (Abstract non-abelian type II inverse theorem) [30, Proposition 6.1] *Let  $C \geq 1, d, D \geq 1, X \geq H \geq 2, X \geq A \geq 1, 0 < \delta < \frac{1}{\log X}$ , and let  $G/\Gamma$  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  $\frac{1}{C}\delta^C |I_{A'}|^2$  integer 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  $J_{A'}$  denoting the interval

$$J_{A'} := \left( \left(1 - \frac{10H}{X}\right) \frac{X}{A'}, \left(1 + \frac{10H}{X}\right) \frac{X}{A'} \right).$$

Then either

$$H \ll_{d,D,C} \delta^{-O_{d,D,C}(1)} \max(A, X/A)$$

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

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

### 2.4 Combinatorial lemmas

The following lemma is Heath-Brown’s identity in the form of [30, Lemma 2.16], but with one additional case.

**Lemma 2.17** *Let  $X \geq 2$ , and let  $L \in \mathbb{N}$  be fixed. We may find a collection  $\mathcal{F}$  of  $O((\log X)^{O(1)})$  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

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

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  $\mu(n)1_{(N_i, 2N_i]}(n)$ . Further,  $N_1 N_2 \dots N_\ell \asymp X$ , 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 with  $\mu$  in place of  $\Lambda$  (but  $(\log n)1_{(N_i, 2N_i]}(n)$  does not appear) and with  $d_k$  in place of  $\Lambda$  (but  $(\log n)1_{(N_i, 2N_i]}(n)$  and  $\mu(n)1_{(N_i, 2N_i]}(n)$  do not appear).

**Proof** The cases of  $\mu$  and  $\Lambda$  are [30, Lemma 2.16]. For the case of  $d_k$ , note that  $d_k$ , which is the  $k$ -fold convolution of 1, can be decomposed into a sum of  $O(\log^k X)$  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  $N_1 N_2 \dots N_k \asymp X$ . □

We shall also frequently use Shiu’s bound (see [35, Theorem 1]).

**Lemma 2.18** [30, Lemma 2.17] *Let  $A \geq 1$  and  $\varepsilon > 0$  be fixed. Let  $X \geq H \geq X^\varepsilon$  and let  $1 \leq q \leq H^{1-\varepsilon}$  be an integer. Let  $f$  be a non-negative multiplicative function such that  $f(p^\ell) \leq A^\ell$  for every prime  $p$  and every  $\ell \in \mathbb{N}$ , and suppose that  $f(n) \ll_c n^c$  for every  $c > 0$ . Then for any integer  $a$  coprime to  $q$  we have*

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

We will also need the fact that the approximants  $\Lambda^\sharp$  and  $d_k^\sharp$  are essentially type  $I$  sums.

**Lemma 2.19** *Let  $k \geq 2$  be a fixed integer, let  $\varepsilon > 0$  be fixed, and let  $X \geq 3$ .*

- (i) *There exists a sequence  $a(n)$  supported on  $[1, X^{\varepsilon/5}]$  with  $a(n) \ll \log X$  such that, for every  $1 \leq n \leq 3X$ , we have  $\Lambda^\sharp(n) = (a * 1)(n) + E(n)$ , where  $E$  is negligible in the sense that for any  $X^\varepsilon \leq H \leq x \leq 2X$  we have*

$$\sum_{x < n \leq x+H} |E(n)| \ll H \exp(-\log^{1/20} X).$$

- (ii) *There exists  $J \ll 1$  and sequences  $a_1(n), \dots, a_J(n)$  supported on  $[1, X^{\varepsilon/5}]$  with  $|a_j(m)| \leq d_{k-1}(m)$  such that, for each  $1 \leq n \leq 3X$ , we have  $d_k^\sharp(n) = \sum_{1 \leq j \leq J} (a_j * \psi_j)(n)$ , where  $\psi_j(n) = (\log^{\ell_j} n) / (\log^{\ell_j} X)$  for some integer  $0 \leq \ell_j \leq k - 1$ .*

**Proof** (i) Recall the definitions  $R = \exp((\log X)^{1/10})$  and  $P(w) = \prod_{p < w} p$  from (1.1). Defining

$$E(n) = \Lambda^\sharp(n) - \frac{P(R)}{\varphi(P(R))} \sum_{\substack{d|n \\ d \leq X^{\varepsilon/5} \\ d|P(R)}} \mu(d),$$

the function  $\Lambda^\sharp(n) - E(n)$  is of the desired form  $a * 1$ , and by Möbius inversion and [30, Lemma 2.18] for  $X^\varepsilon \leq H \leq x \leq 2X$  we have

$$\sum_{x < n \leq x+H} |E(n)| \leq \frac{P(R)}{\varphi(P(R))} \sum_{\substack{x < dn \leq x+H \\ d > X^{\varepsilon/5} \\ d|P(R)}} 1 \ll_\varepsilon H \exp(-\log^{1/20} X). \tag{2.9}$$

- (ii) By (1.2) and (1.3), we have

$$d_k^\sharp(n) = \sum_{j=0}^{k-1} \frac{\binom{k}{j}}{(k-j-1)!}$$

$$\times \sum_{\substack{n=n_1 \cdots n_{k-1} d \\ n_1, \dots, n_{j-1} \leq R_k \\ R_k < n_{j+1}, \dots, n_{k-1} \leq R_k^2}} \left( \frac{\log(R_k^{k-j} n_{j+1} \cdots n_{k-1}) + \log d}{\log R_k} \right)^{k-j-1}.$$

Applying the binomial formula to the inner sum and merging the variables  $n_1, \dots, n_{k-1}$ , the claim follows.  $\square$

### 3 Major arc estimates

The purpose of this section is to handle the case of Theorem 1.1 where  $F(g(n)\Gamma)$  is major arc in the sense that it behaves like  $n^{iT}$  in arithmetic progressions to small modulus (the case  $T = 0$  would correspond to the classical major arcs in the circle method), for some  $T$  of polynomial size. To this end, we prove the following theorem.

**Theorem 3.1** (Major arc estimate) *Let  $X \geq 3$  and  $X^{\theta+\varepsilon} \leq H \leq X$  for some  $\theta \in (0, 1)$  and some small  $\varepsilon \in (0, 1)$ . Let  $T$  be a real number with  $|T| \leq X^C$  for some  $C > 0$ .*

(i) (Major arc discorrelation) *Let  $\theta = 1/3$ . Then, for any  $A > 0$ , one has*

$$\left| \sum_{x < n \leq x+H} \mu(n) n^{-iT} \right|^* \leq H \log^{-A} X$$

and

$$\left| \sum_{x < n \leq x+H} (\Lambda(n) - \Lambda_k^\sharp(n)) n^{-iT} \right|^* \leq H \log^{-A} X$$

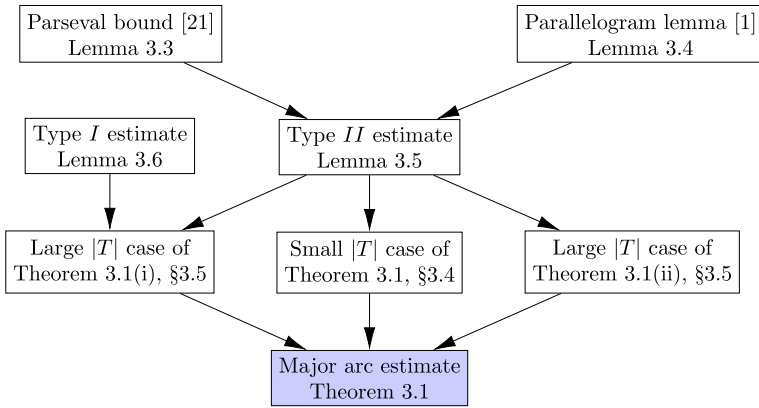
for all  $x \in [X, 2X]$  outside of a set of measure  $O_{A,\varepsilon,C}(X \log^{-A} X)$ . Furthermore, for any integer  $k \geq 2$  and some constant  $c_{k,C} > 0$  depending only on  $k, C$ , one has

$$\left| \sum_{x < n \leq x+H} (d_k(n) - d_k^\sharp(n)) n^{-iT} \right|^* \leq H X^{-c_{k,C}\varepsilon}$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_{\varepsilon,C}(X^{1-c_{k,C}\varepsilon})$ .

(ii) (Major arc discorrelation with weight on prime factors) *Let  $\theta = 0$  and  $\kappa \in (0, \varepsilon/20)$ . Then, for any  $A > 0$  and any integer  $k \geq 2$ , one has*

$$\left| \sum_{x < n \leq x+H} \mu(n) n^{-iT} \sum_{X^\kappa < p \leq X^{\varepsilon/10}} 1_{p|n} \right|^* \leq H \log^{-A} X$$



**Fig. 2** Proof of Theorem 3.1. The Vinogradov–Korobov bound (Lemma 3.2) and Heath–Brown identity (Lemma 2.17) are also used at several points in the proof, but are not depicted on this diagram to reduce clutter

and

$$\left| \sum_{x < n \leq x+H} (d_k(n) - d_k^\sharp(n))n^{-iT} \sum_{X^\kappa < p \leq X^{\varepsilon/10}} 1_{p|n} \right|^* \leq H \log^{-A} X$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_{A,\varepsilon,\kappa,C}(X \log^{-A} X)$ .

The proof strategy of Theorem 3.1 is to apply combinatorial decompositions to  $\mu, \Lambda - \Lambda^\sharp, d_k - d_k^\sharp$  (possibly with an additional weight as in part (ii)) and to prove estimates for the resulting type I and II major arc sums. These sums will be handled in Sects. 3.3 and 3.2, respectively, and then we are left with some reductions (involving the removal of maximal summation) and combinatorial considerations in Sects. 3.4–3.5. For the logic of the proof, see Fig. 2.

### 3.1 Lemmas on Dirichlet polynomials

We begin with the following standard pointwise bounds for Dirichlet polynomials of  $\mu, \Lambda, d_k$ .

**Lemma 3.2** (Vinogradov–Korobov bound) *Let  $k \in \mathbb{N}$  be fixed. Let  $f \in \{\mu, \Lambda, d_k\}$ . Let  $X \geq 3, A > 0$ , and define  $W_f$  by the formulae*

$$W_\mu := 0; \quad W_\Lambda := \log^A X; \quad W_{d_k} := X^{c_k},$$

where  $c_k > 0$  is a small enough constant. Then we have

$$\sup_{W_f \leq |T| \leq X^A} \left| \sum_{X < n \leq 2X} \frac{f(n)}{n^{1+iT}} \right|^* \ll_A \frac{\log^{O(1)} X}{(W_f + \log^A X)^{1/2}}.$$

**Proof** For some interval  $I \subset [X, 2X]$  and some integers  $q \geq r \geq 1$ , we have

$$\left| \sum_{X < n \leq 2X} \frac{f(n)}{n^{1+iT}} \right|^* = \left| \sum_{\substack{n \in I \\ n \equiv r \pmod{q}}} \frac{f(n)}{n^{1+iT}} \right|. \tag{3.1}$$

We may assume that  $q \ll (W_f + \log^A X)^{1/2}$ , as otherwise the claimed estimate follows directly from the triangle inequality (and Lemma 2.18 in the case  $f = d_k$ ). Let  $d = (r, q)$  and  $r' = r/d, q' = q/d$ . Then using the orthogonality of characters, the right-hand side of (3.1) becomes

$$\leq \frac{1}{\varphi(q')} \sum_{\chi \pmod{q'}} \left| \sum_{m \in d^{-1}I} \frac{f(dm)\chi(m)}{m^{1+iT}} \right|. \tag{3.2}$$

Now the claim follows in the case of  $f \in \{\mu, \Lambda\}$  from [30, Lemma 3.9](ii)–(iii) (which is an application of the Vinogradov–Korobov zero free region) and partial summation. The case  $f = d_k$  follows from [30, Lemma 3.9](i) and partial summation after first writing  $d_k(n) = \sum_{n=n_1 \dots n_k} 1$  in (3.2) and pulling out all but the largest of the variables  $n_i$  outside the absolute values.  $\square$

We will also need the following standard estimate, which relates the variance of a sequence in short intervals to the mean square of its Dirichlet polynomial.

**Lemma 3.3** (Parseval-type bound for variance) *Let  $X \geq 3$  and  $1 \leq W \leq X^{1/10}$ . Let  $C > 0$  and  $|a(n)| \leq d_2(n)^C$ . Let<sup>7</sup>  $A$  denote the Dirichlet polynomial*

$$A(s) := \sum_{X/8 < n \leq 8X} a(n)n^{-s}.$$

(i) *For  $1 \leq H_1 \leq H_2 \leq X/W^3$ , one has*

$$\begin{aligned} & \frac{1}{X} \int_X^{2X} \left| \frac{1}{H_1} \sum_{x < n \leq x+H_1} a(n) - \frac{1}{H_2} \sum_{x < n \leq x+H_2} a(n) \right|^2 dx \\ & \ll C \left( \frac{\log^{2C-1} X}{W} \right)^2 + \max_{T \geq \frac{X}{H_1}} \frac{X/H_1}{T} \int_{W \leq |t| \leq T} |A(1+it)|^2 dt. \end{aligned}$$

(ii) *For  $1 \leq H_1 \leq X$ , one has*

$$\frac{1}{X} \int_X^{2X} \left| \frac{1}{H_1} \sum_{x < n \leq x+H_1} a(n) \right|^2 dx \ll \max_{T \geq \frac{X}{H_1}} \frac{X/H_1}{T} \int_{|t| \leq T} |A(1+it)|^2 dt.$$

<sup>7</sup>For some applications where  $a$  is a convolution of two dyadically supported sequences, it is convenient that  $A(s)$  involves a sum over an interval somewhat longer than  $[X, 2X]$ ; therefore we use the interval  $[X/8, 8X]$  here.

**Proof** This is very similar to [21, Lemma 14]. Let us give some details for part (i); part (ii) is similar but easier.

For any  $x \geq 1$  such that  $x, x + H_1, x + H_2$  are not integers, Perron’s formula gives

$$\begin{aligned} & \frac{1}{H_1} \sum_{x < n \leq x + H_1} a(n) - \frac{1}{H_2} \sum_{x < n \leq x + H_2} a(n) \\ &= \frac{1}{2\pi} \int_{\mathbb{R}} A(1 + it) \left( \frac{(x + H_1)^{1+it} - x^{1+it}}{H_1(1 + it)} - \frac{(x + H_2)^{1+it} - x^{1+it}}{H_2(1 + it)} \right) dt \\ &:= I_1(x) + I_2(x), \end{aligned}$$

where  $I_1(x)$  is the integral over  $|t| \leq W$  and  $I_2(x)$  is the integral over  $|t| > W$ . By Taylor approximation, for any real number  $|y| \leq 1$  and complex number  $s$ , we have  $(1 + y)^s = 1 + sy + O((|s|^2 + 1)y^2)$ . Substituting this into the integrand in  $I_1(x)$  and recalling that  $H_2 \leq X/W^3$ , we obtain

$$\begin{aligned} |I_1(x)|^2 &\ll \left( \int_{|t| \leq W} |A(1 + it)| (|t| + 1) \frac{H_2}{x} dt \right)^2 \ll \left( W^2 \frac{H_2}{X} \max_{|t| \leq W} |A(1 + it)| \right)^2 \\ &\ll \left( W^{-1} \sum_{X/8 < n \leq 8X} \frac{|a(n)|}{n} \right)^2 \ll \left( \frac{\log^{2^c - 1} X}{W} \right)^2, \end{aligned}$$

using in the last step the standard divisor sum bound  $\sum_{X/8 < n \leq 8X} d_2(n)^C/n \ll \log^{2^c - 1} X$ .

To handle the contribution of  $I_2(x)$ , it suffices to show for  $j \in \{1, 2\}$  that

$$\begin{aligned} & \int_X^{2X} \left| \int_{|t| > W} A(1 + it) \frac{(x + H_j)^{1+it} - x^{1+it}}{H_j(1 + it)} dt \right|^2 dx \\ &\ll \max_{T \geq \frac{X}{H_1}} \frac{X^2/H_1}{T} \int_{W \leq |t| \leq T} |A(1 + it)|^2 dt. \end{aligned}$$

This estimate is contained in the proof of [21, Lemma 14] (and is proven by inserting a smooth majorant for  $1_{[X, 2X]}(x)$  into the integral over  $x$  and opening the square).  $\square$

For proving type II major arc estimates, we need a lemma of Baker, Harman and Pintz [1] on mean values of products of Dirichlet polynomials.

**Lemma 3.4** (Baker–Harman–Pintz parallelogram lemma) *Let  $\theta \in (1/2, 1)$ ,  $\varepsilon > 0$  and  $K \geq 2$  be fixed. Let  $X \geq 3$ . Let  $|a(n)|, |b(n)|, |c(n)| \leq d_2(n)^K$ , and define*

$$\begin{aligned} A(s) &= \sum_{M < m \leq KM} a(m)m^{-s}, & B(s) &= \sum_{N < n \leq KN} b(n)n^{-s}, \\ C(s) &= \sum_{R < r \leq KR} c(r)r^{-s}, \end{aligned}$$

where  $M, N, R \geq 1$  satisfy  $MNR \asymp X$ . Suppose that  $M = X^{\alpha_1}$ ,  $N = X^{\alpha_2}$ , where  $\alpha_1, \alpha_2 \in [0, 1]$  satisfy

$$|\alpha_1 - \alpha_2| < 2\theta - 1 - \frac{\varepsilon}{10}, \quad \alpha_1 + \alpha_2 > 1 - \min \left\{ 4\theta - 2, \frac{4\theta - 1}{3}, \frac{24\theta - 13}{3} \right\} + \frac{\varepsilon}{10}.$$

Suppose also that for some  $0 \leq T_0 \leq X^{1-\theta}$  and  $1 \leq W \leq X^{\varepsilon/1000}$  we have

$$\max_{T_0 \leq |t| \leq X^{1-\theta}} |C(1 + it)| \leq \frac{1}{W^{1/3}}. \tag{3.3}$$

Then we have

$$\int_{T_0}^{X^{1-\theta}} |A(1 + it)||B(1 + it)||C(1 + it)| dt \ll \frac{\log^{O(1)} X}{W^{1/3}}. \tag{3.4}$$

**Proof** This is essentially [11, Lemma 7.3] with  $g = 1$  (alternatively, see [1, Lemma 9]), the proof of which is based on mean and large value estimates for Dirichlet polynomials. One minor difference is that in that lemma one had  $T_0 = \exp(\log^{1/3} X)$  and  $W = \log^{-A} X$  (and  $\varepsilon = o(1)$ ), but the same proof works with the more general choices above. Another minor difference is that in [11, Lemma 7.3] the integral in (3.4) was over the 1/2-line rather than the 1-line (and similarly (3.3) was replaced with an analogous estimate on the 1/2-line), but this also makes no difference in the proof.  $\square$

### 3.2 Type II major arc estimates

In this subsection, we consider the kind of type II major arc sums that arise in the proof of Theorem 3.1. The following lemma is used in the proof of Theorem 3.1 to control type II sums.

**Lemma 3.5** (Type II major arc estimate) *Let  $\varepsilon > 0$  be small and let  $C > 0$ . Also let  $|a(m)|, |b(m)| \leq d_2(m)^C$ ,  $X \geq 3$  and  $1 \leq W \leq X^{\varepsilon/1000}$ .*

(i) *If  $X^\varepsilon \leq M \leq X^{1/2}$ ,  $X/W^4 \geq H_2 \geq H_1 \geq X^{1/3+\varepsilon}$ , and*

$$\begin{aligned} \sup_{W \leq |t| \leq XW/H_1} \left| \sum_{m \sim M} \frac{a(m)}{m^{1+it}} \right| &\leq \frac{1}{W^{1/3}}, \\ \sup_{W \leq |t| \leq XW/H_1} \left| \sum_{X/(3M) < n \leq 3X/M} \frac{b(n)}{n^{1+it}} \right| &\leq \frac{1}{W^{1/3}}, \end{aligned} \tag{3.5}$$

we have

$$\left| \sum_{\substack{x < mn \leq x+H_1 \\ m \sim M}} a(m)b(n) - \frac{H_1}{H_2} \sum_{\substack{x < mn \leq x+H_2 \\ m \sim M}} a(m)b(n) \right| \leq \frac{H_1}{W^{1/10}}$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_{\varepsilon, C}(X(\log^{O_C(1)} X)/W^{1/10})$ .

- (ii) *The same conclusion also holds assuming  $X/W^4 \geq H_2 \geq H_1 \geq X^{2\epsilon}$ , and  $X^\epsilon \leq M \leq H_1 X^{-\epsilon}$ , and the first estimate in (3.5) (without assuming  $H_1 \geq X^{1/3+\epsilon}$  or the second estimate in (3.5)).*
- (iii) *If  $X^\epsilon \leq M \leq X^{1/2}$ ,  $X \geq H_1 \geq X^{1/3+\epsilon}$ , and*

$$\begin{aligned} \sup_{|t| \leq XW/H_1} \left| \sum_{m \sim M} \frac{a(m)}{m^{1+it}} \right| &\leq \frac{1}{W^{1/3}}, \\ \sup_{|t| \leq XW/H_1} \left| \sum_{X/(3M) < n \leq 3X/M} \frac{b(n)}{n^{1+it}} \right| &\leq \frac{1}{W^{1/3}}, \end{aligned} \tag{3.6}$$

then we have

$$\left| \sum_{\substack{x < mn \leq x + H_1 \\ m \sim M}} a(m)b(n) \right| \leq \frac{H_1}{W^{1/10}}$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_{\epsilon, C}(X \log^{O_C(1)} X) / W^{1/10}$ .

- (iv) *The same conclusion also holds assuming  $X \geq H_1 \geq X^{2\epsilon}$ , and  $X^\epsilon \leq M \leq H_1 X^{-\epsilon}$ , and the first estimate in (3.6) (without assuming  $H_1 \geq X^{1/3+\epsilon}$  or the second estimate in (3.6)).*

**Proof of Lemma 3.5** Let  $K_C > 0$  be a large enough constant which may depend on  $C$ . We may assume that  $W \geq \log^{K_C} X$ , since otherwise, by adjusting the  $O_C(1)$  constants in the lemma, we may take the exceptional set there to contain all of  $[X, 2X]$ . Applying Chebyshev’s inequality, it suffices to show in the case of Lemma 3.5(i)–(ii) that

$$\begin{aligned} \frac{1}{X} \int_X^{2X} \left| \sum_{\substack{x < mn \leq x + H_1 \\ m \sim M}} a(m)b(n) - \frac{H_1}{H_2} \sum_{\substack{x < mn \leq x + H_2 \\ m \sim M}} a(m)b(n) \right|^2 dx \\ \ll H_1^2 \frac{\log^{O_C(1)} X}{W^{3/10}}, \end{aligned} \tag{3.7}$$

and in the case of Lemma 3.5(iii)–(iv) that

$$\frac{1}{X} \int_X^{2X} \left| \sum_{\substack{x < mn \leq x + H_1 \\ m \sim M}} a(m)b(n) \right|^2 dx \ll H_1^2 \frac{\log^{O_C(1)} X}{W^{3/10}}. \tag{3.8}$$

Let  $N := X/M$  and

$$F(s) = \sum_{\substack{X/8 < mn \leq 8X \\ m \sim M \\ N/3 < n \leq 3N}} a(m)b(n)(mn)^{-s}.$$

Then  $F(s) = A(s)B(s)$ , where

$$A(s) := \sum_{m \sim M} a(m)m^{-s}, \quad B(s) := \sum_{N/3 < n \leq 3N} b(n)n^{-s}.$$

By Lemma 3.3(i), the claim (3.7) reduces to showing that

$$\int_{W \leq |t| \leq T} |A(1+it)B(1+it)|^2 dt \ll \frac{TH_1}{X} \cdot \frac{\log^{O_C(1)} X}{W^{3/10}} \tag{3.9}$$

uniformly for all  $T \geq X/H_1$ . Similarly, by Lemma 3.3(ii), the claim (3.8) reduces to showing that

$$\int_{-T}^T |A(1+it)B(1+it)|^2 dt \ll \frac{TH_1}{X} \cdot \frac{\log^{O_C(1)} X}{W^{3/10}} \tag{3.10}$$

uniformly for all  $T \geq X/H_1$ .

Note that by the mean value theorem for Dirichlet polynomials (see e.g. [16, Theorem 9.1]) and a standard bound for moments of the divisor function we have

$$\begin{aligned} \int_{-T}^T |A(1+it)B(1+it)|^2 dt &\ll (T+X) \sum_{X/8 < \ell \leq 8X} \frac{|\sum_{\ell=mn} a(m)b(n)|^2}{\ell^2} \\ &\ll (T+X) \sum_{X/8 < \ell \leq 8X} \frac{d_2(\ell)^{4C+2}}{\ell^2} \\ &\ll \left(1 + \frac{T}{X}\right) \log^{O_C(1)} X. \end{aligned}$$

Since  $W \leq H_1^{1/20}$ , it follows that (3.9) and (3.10) hold in the case  $T \geq XW^{3/10}/H_1$ . We may therefore assume from now on that

$$T < XW^{3/10}/H_1. \tag{3.11}$$

If  $M \leq H_1/W^{3/10}$ , we can use the pointwise bound from (3.5) on  $|A(1+it)|$  and the mean value theorem for Dirichlet polynomials to bound

$$\begin{aligned} \int_{W \leq |t| \leq T} |A(1+it)B(1+it)|^2 dt &\ll \frac{1}{W^{2/3}} \int_{-T}^T |B(1+it)|^2 dt \\ &\ll \frac{T + \frac{X}{M}}{W^{2/3}} \sum_{X/(8M) < n \leq 8X/M} \frac{|b(n)|^2}{n^2} \\ &\ll \frac{\log^{O_C(1)} X}{W^{2/3}}, \end{aligned}$$

so (3.9) holds. We obtain (3.10) similarly. Hence parts (ii) and (iv) of the lemma follow (since there we assume that  $X^\epsilon \leq M \leq H_1 X^{-\epsilon}$ ).

We are left with proving parts (i) and (iii) (where we assume  $X \geq H_1 \geq X^{1/3+\varepsilon}$  and  $X^\varepsilon \leq M \leq X^{1/2}$ ). By (3.11) we have  $T < X^{2/3}$ . It now suffices to prove that

$$\int_{W \leq |t| \leq X^{2/3}} |A(1+it)B(1+it)|^2 dt \ll \frac{\log^{O_C(1)} X}{W^{3/10}} \tag{3.12}$$

in the case of Lemma 3.5(i) and

$$\int_{|t| \leq X^{2/3}} |A(1+it)B(1+it)|^2 dt \ll \frac{\log^{O_C(1)} X}{W^{3/10}} \tag{3.13}$$

in the case of Lemma 3.5(iii).

In what follows, denote  $M = X^{\alpha_1}$ ,  $N = X^{\alpha_2}$ , so that the coefficients of  $A(s)$  are supported on  $[X^{\alpha_1}, 2X^{\alpha_1}]$  and the coefficients of  $B(s)$  are supported on  $[X^{\alpha_2}/3, 3X^{\alpha_2}]$ . The estimates (3.12) and (3.13) follow from<sup>8</sup> Lemma 3.4. Indeed, we can apply that lemma with  $X^2$  in place of  $X$ , and with  $\theta = 2/3$  and  $T_0 \in \{W, 0\}$ . Applying it to the Dirichlet polynomials  $A(s)$ ,  $A(s)$ ,  $B(s)^2$ , and recalling (3.5), (3.6), we obtain the claim if

$$2\alpha_1 > 1 - \frac{4 \cdot \frac{2}{3} - 1}{3} + \frac{\varepsilon}{10} = \frac{4}{9} + \frac{\varepsilon}{10}.$$

We are left with the case  $\alpha_1 \leq 2/9 + \varepsilon/20$ . But in this case applying Lemma 3.4 to the Dirichlet polynomials  $B(s)$ ,  $B(s)$ ,  $A(s)^2$  instead, we obtain the claim if

$$2\alpha_2 > 1 - \frac{4 \cdot \frac{2}{3} - 1}{3} + \frac{\varepsilon}{10} = \frac{4}{9} + \frac{\varepsilon}{10},$$

which must now hold since  $\alpha_1 + \alpha_2 = 1$ . This completes the proof. □

### 3.3 Type I major arc estimates

Next we shall prove some type I major arc estimates that are useful in the case when  $|T|$  is large.

#### 3.3.1 Type I major arc estimates without weights

The following lemma asserts, roughly speaking, that type I major arc sums on short intervals are small as soon as  $n^{iT}$  exhibits non-trivial oscillation on the interval.

**Lemma 3.6** (Type I major arc estimate) *Let  $C > 0$  and  $\varepsilon > 0$ . Let  $2 \leq X^\varepsilon \leq H \leq X^{1-\varepsilon}$ ,  $0 < \delta < 1/\log X$  and  $1 \leq M \leq HX^{-\varepsilon/2}$ . Let  $|a(m)| \leq d_2(m)^C$ , and suppose that  $a$  is supported on  $[M, 2M]$ . Suppose that for some real number  $T \in [-X^C, X^C]$  we have*

$$\left| \sum_{x < n \leq x+H} (a * 1)(n)n^{-iT} \right|^* \geq \delta H$$

<sup>8</sup>Alternatively, one could use [11, (9.2.3) in Lemma 9.4]. We have chosen to reduce matters to Lemma 3.4 as that lemma is better known.

for all  $x \in [X, 2X]$  in a set of measure  $\geq \delta X$ . Then we have

$$|T| \ll_{\varepsilon, C} \delta^{-O_{\varepsilon, C}(1)} \frac{X}{H}. \quad (3.14)$$

In fact, only the case  $M \leq X^\varepsilon$  of this lemma will be needed, since the type II region in Lemma 3.5 is very wide.

**Proof** We allow all implied constants in this proof to depend on  $\varepsilon, C$ . On increasing the  $O_{\varepsilon, C}(1)$  constant in (3.14) if necessary, we may assume that  $X$  is large enough in terms of  $\varepsilon, C$  and that

$$\delta^{-1} \leq X^{c_{C, \varepsilon}} \quad (3.15)$$

for a suitably small constant  $c_{C, \varepsilon} > 0$ .

By Taylor approximation, for  $y \in [x, x + H]$  and for any  $x \in [X, 2X]$  and integer  $J \geq 1$ , we can write

$$y^{iT} = e\left(\frac{T}{2\pi} \log y\right) = e(P_x(y)) + O_J\left(T\left(\frac{H}{X}\right)^{J+1}\right) \quad (3.16)$$

where

$$P_x(y) := \frac{T}{2\pi} \log x + \frac{T}{2\pi} \sum_{j=1}^J \frac{(-1)^{j-1}}{jx^j} (y-x)^j$$

is the Taylor polynomial of degree  $J$  of the function  $\frac{T}{2\pi} \log(\cdot)$  at  $x$ . We choose  $J = O(1)$  to be large enough so that the error term in (3.16) becomes  $O(X^{-1})$ . Then, for all  $x \in [X, 2X]$  in a set  $E$  of measure  $\geq \delta X$ , we have

$$\left| \sum_{x < n \leq x+H} (a * 1)(n) e(-P_x(n)) \right|^* \gg \delta H.$$

Since  $H \in \mathbb{N}$ , without loss of generality we may assume that  $E$  is a union of  $\gg \delta X$  intervals of the form  $[m, m + 1)$  with  $m \in \mathbb{N}$ . By assumptions, the sum above is a  $(\delta^{O(1)}, 2M)$  type I sum in the notation of [30, Definition 4.1(i)]. Since  $M \leq HX^{-\varepsilon/2}$ , it follows from the type I estimate of [30, Theorem 4.2(i)] (together with Remark 2.8) that for all  $x \in E$  there is some integer  $1 \leq k_x \ll \delta^{-O(1)}$  such that

$$\left\| k_x \frac{T}{2\pi j x^j} \right\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{\delta^{-O(1)}}{H^j} \quad (3.17)$$

for all  $1 \leq j \leq J$ . Actually we will only need the  $j = 1$  case of this bound. By the pigeonhole principle, we can find an integer  $1 \leq k \ll \delta^{-O(1)}$  such that

$$\left\| \frac{kT}{2\pi x} \right\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{\delta^{-O(1)}}{H} \quad (3.18)$$

for  $\gg \delta^{O(1)} X$  integers  $x$  in  $[X, 2X]$ .

If we have

$$\frac{k|T|}{2\pi X} < \frac{1}{2},$$

then (3.18) simplifies to

$$\frac{k|T|}{2\pi x} \ll \frac{\delta^{-O(1)}}{H},$$

and (3.14) follows. Thus we may assume that

$$\frac{k|T|}{2\pi X} \geq \frac{1}{2}.$$

If we then define  $h := \lfloor \sqrt{H} \rfloor$ , then  $X^{1+\varepsilon/4} \ll hk|T| \ll X^{O_C(1)}$  and

$$\left\| \frac{hkT}{2\pi x} \right\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{\delta^{-O(1)}}{H^{1/2}} \quad (3.19)$$

for  $\gg \delta^{O(1)}X$  integers  $x$  in  $[X, 2X]$ .

From van der Corput's exponential sum bound [16, Theorem 8.20] we have

$$\sum_{X < n \leq 2X} e\left(\frac{m h k T}{2\pi n}\right) \ll X^{1-c'_{C,\varepsilon}} \quad (3.20)$$

for all  $1 \leq m \leq X^{c'_{C,\varepsilon}}$  and some  $c'_{C,\varepsilon} > 0$ . Applying the Erdős–Turán inequality [33, Corollary 1.1 in Sect. 1] and (3.20), we conclude that the bound (3.19) can only hold for at most

$$\ll \frac{\delta^{-O(1)}}{H^{1/2}} X + X^{1-c'_{C,\varepsilon}} \log X$$

integers  $x \in [X, 2X]$ . But this is a contradiction if the constant  $c_{C,\varepsilon}$  in (3.15) is small enough.  $\square$

### 3.4 Theorem 3.1 in the case of small $|T|$

The proof of Theorem 3.1 splits into the “small  $|T|$  case”

$$|T| \leq \frac{X}{H} X^{\varepsilon/2} \quad (3.21)$$

and the “large  $|T|$  case”

$$|T| > \frac{X}{H} X^{\varepsilon/2}. \quad (3.22)$$

In this section we treat the case of small  $|T|$ ; the large  $|T|$  case will be handled in Sect. 3.5.

Let  $\varepsilon, \kappa > 0$  be small, let  $A, C > 0$  be large, and let  $c_{k,C} > 0$  be small enough. To have a unified notation to treat all cases, we set  $l := 1$  when proving Theorem 3.1(i) and  $l := 2$  when proving Theorem 3.1(ii), fix  $f \in \{\mu, \Lambda, d_k\}$  for  $l = 1$  and  $f \in \{\mu, d_k\}$  for  $l = 2$ , and introduce the parameters and weights

$$\begin{aligned} \delta &:= \begin{cases} \log^{-A} X & \text{if } f \in \{\mu, \Lambda\} \text{ or } l = 2, \\ X^{-c_{k,C}\varepsilon} & \text{if } f = d_k \text{ and } l = 1, \end{cases} & (3.23) \\ g(n) &:= \begin{cases} 1 & \text{if } l = 1, \\ \sum_{X^\kappa < p \leq X^{\varepsilon/10}} 1_{p|n} & \text{if } l = 2, \end{cases} \\ H' &:= H/X^{\varepsilon/2} \geq \begin{cases} X^{1/3+\varepsilon/2} & \text{if } l = 1, \\ X^{\varepsilon/2} & \text{if } l = 2, \end{cases} \\ H^* &:= X^{0.99}. \end{aligned}$$

The condition (3.21) can now be written as

$$|T| \leq \frac{X}{H'}. \tag{3.24}$$

By Chebyshev’s inequality, it suffices to show that

$$\frac{1}{X} \int_X^{2X} \left( \left| \sum_{x < n \leq x+H} (f(n) - f^\sharp(n))g(n)n^{-iT} \right|^* \right)^2 dx \ll_{A,\varepsilon,\kappa} \delta^3 H^2. \tag{3.25}$$

From (3.24) we see that the function  $n \mapsto n^{-iT}$  has total variation  $O(1)$  on any interval  $I \subset [X, 4X]$  of length  $H'$ . Thus, by covering  $(x, x + H]$  by  $O(H/H')$  intervals of length  $H'$  and using Lemma 2.2(i)–(ii), we have

$$\begin{aligned} &\left| \sum_{x < n \leq x+H} (f(n) - f^\sharp(n))g(n)n^{-iT} \right|^* \\ &\ll \sum_{0 \leq \ell \leq \frac{H}{H'}} \left| \sum_{x+\ell H' < n \leq x+(\ell+1)H'} (f(n) - f^\sharp(n))g(n) \right|^* \end{aligned}$$

and hence by Cauchy–Schwarz

$$\begin{aligned} &\left( \left| \sum_{x < n \leq x+H} (f(n) - f^\sharp(n))g(n)n^{-iT} \right|^* \right)^2 \\ &\ll \frac{H}{H'} \sum_{0 \leq \ell \leq \frac{H}{H'}} \left( \left| \sum_{x+\ell H' < n \leq x+(\ell+1)H'} (f(n) - f^\sharp(n))g(n) \right|^* \right)^2. \end{aligned}$$

Integrating and applying a translation, it thus suffices to show that

$$\frac{1}{X} \int_X^{3X} \left( \left| \sum_{x < n \leq x+H'} (f(n) - f^\sharp(n))g(n) \right|^* \right)^2 dx \ll_{A,\varepsilon,\kappa} \delta^3 (H')^2. \tag{3.26}$$

Thus we have eliminated the  $n^{-iT}$  factor, at the slight cost of worsening  $H$  to  $H'$ .

Now we deal with the maximal truncation  $|\cdot|^*$ . Observe the bounds  $g(n) \ll_\kappa 1$  for  $n \leq 4X$  and  $d_k^\sharp(n) \ll d_k(n) \ll n^{o(1)}$  (see [30, (3.14)]). Hence, by Shiu’s bound (Lemma 2.18), we have

$$\sum_{\substack{x < n \leq x+H' \\ n \equiv r \pmod{q}}} |f(n) - f^\sharp(n)|g(n) \ll \left( \frac{H'}{\varphi(q)} + (H')^{1/100} \right) \log^{O(1)} X \tag{3.27}$$

(say) for any  $1 \leq r \leq q$ . Thus by (1.4) we have

$$\begin{aligned} & \left| \sum_{x < n \leq x+H'} (f(n) - f^\sharp(n))g(n) \right|^* \\ & \ll \sum_{1 \leq r \leq q \leq \delta^{-4}} \sup_{0 \leq H'' \leq H'} \left| \sum_{\substack{x < n \leq x+H'' \\ n \equiv r \pmod{q}}} (f(n) - f^\sharp(n))g(n) \right| + \delta^2 H' \log^{O(1)} X. \end{aligned}$$

By a further application of Shiu’s bound, one can round  $H''$  to the nearest multiple of  $\delta^{10} H'$  without significantly worsening the error term. Thus we have

$$\begin{aligned} & \left| \sum_{x < n \leq x+H'} (f(n) - f^\sharp(n))g(n) \right|^* \\ & \ll \sum_{1 \leq r \leq q \leq \delta^{-4}} \sum_{\substack{0 \leq H'' \leq H' \\ H'' \in (\delta^{10} H') \cdot \mathbb{Z}}} \left| \sum_{\substack{x < n \leq x+H'' \\ n \equiv r \pmod{q}}} (f(n) - f^\sharp(n))g(n) \right| \\ & \quad + \delta^2 H' \log^{O(1)} X. \end{aligned} \tag{3.28}$$

The contribution of the  $\delta^2 H' \log^{O(1)} X$  error to (3.26) is acceptable if  $A$  is large enough. As there are  $O(\delta^{-18})$  summands in (3.28), it will suffice by the Cauchy–Schwarz inequality to show that

$$\frac{1}{X} \int_X^{3X} \left| \sum_{\substack{x < n \leq x+H'' \\ n \equiv r \pmod{q}}} (f(n) - f^\sharp(n))g(n) \right|^2 dx \ll_{A,\varepsilon,\kappa} \delta^{39} (H')^2 \tag{3.29}$$

for all  $1 \leq r \leq q \leq \delta^{-4}$  and all  $H'' \in [0, H']$ . By (3.27) we may assume that  $H'' \geq \delta^{100} H' \gg H/X^{2\varepsilon/3}$ .

Fix  $r, q, H''$ . From [30, Theorem 3.1 with  $\varepsilon = 1/200$ ] (and the fact that  $g(n)$  is a sum of  $O(X^{\varepsilon/10})$  indicators of arithmetic progressions, and we assume that our  $\varepsilon$  is

small), we have the long interval estimate

$$\left| \sum_{x < n \leq x + H^*} (f(n) - f^\sharp(n))g(n) \right|^* \ll_{A,\varepsilon} \delta^{100} H^*.$$

for all  $x \in [X, 3X]$ . Hence, by the triangle inequality, (3.29) follows if we show that

$$\begin{aligned} & \frac{1}{X} \int_X^{3X} \left| \sum_{\substack{x < n \leq x + H'' \\ n \equiv r \pmod{q}}} (f(n) - f^\sharp(n))g(n) \right. \\ & \quad \left. - \frac{H''}{H^*} \sum_{\substack{x < n \leq x + H^* \\ n \equiv r \pmod{q}}} (f(n) - f^\sharp(n))g(n) \right|^2 dx \tag{3.30} \\ & \ll_{A,\varepsilon,\kappa} \delta^{39} (H'')^2. \end{aligned}$$

We now claim the estimate

$$\left| \sum_{\substack{x < n \leq x + H'' \\ n \equiv r \pmod{q}}} f^\sharp(n)g(n) - \frac{H''}{H^*} \sum_{\substack{x < n \leq x + H^* \\ n \equiv r \pmod{q}}} f^\sharp(n)g(n) \right| \ll_{A,\varepsilon} \delta^{100} H'' \tag{3.31}$$

for all  $x \in [X, 3X]$ . In the case  $f = \mu$ , this bound is trivial since  $\mu^\sharp$  vanishes. In the case  $f \in \{\Lambda, d_k\}$ ,  $l = 1$  (so that  $g = 1$ ), this follows from [30, Lemma 3.3], noting that  $H'' \geq X^{1/4}$  when  $l = 1$ . The only remaining case to verify is when  $f = d_k$ ,  $l = 2$  (note that we do not permit  $f = \Lambda$  when  $l = 2$ ). In that case, by Lemma 2.19(ii),  $d_k^\sharp$  is a sum of  $O(1)$  type I sums  $a_j * \psi$  where  $\psi = \log^{\ell_j}$  for some integers  $0 \leq \ell_j \ll 1$ , some divisor-bounded sequences  $a_j$  supported on  $[1, X^{\varepsilon/5}]$ . Moreover,  $g$  is a sum of  $O(X^{\varepsilon/10})$  indicators of arithmetic progressions. Hence, (3.31) follows in this case by using

$$\left| \sum_{\substack{n \in I \\ n \equiv r \pmod{q}}} 1_{m|n} - \frac{|I|}{|J|} \sum_{\substack{n \in J \\ n \equiv r \pmod{q}}} 1_{m|n} \right| \ll 1$$

for any nonempty intervals  $I, J$  with  $|I| \leq |J|$ , and for various integers  $1 \leq m \leq X^{\varepsilon/5 + \varepsilon/10}$ , and applying partial summation.

In view of (3.31) and the triangle inequality, we have reduced (3.30) to showing that

$$\frac{1}{X} \int_X^{3X} \left| \sum_{\substack{x < n \leq x + H'' \\ n \equiv r \pmod{q}}} f(n)g(n) - \frac{H''}{H^*} \sum_{\substack{x < n \leq x + H^* \\ n \equiv r \pmod{q}}} f(n)g(n) \right|^2 dx \ll_{A,\varepsilon,\kappa} \delta^{39} (H'')^2 \tag{3.32}$$

We now treat the  $l = 1, 2$  cases separately.

### 3.4.1 Proof of (3.32) for $l = 1$

By Lemma 2.17 with  $L = 3$ , for  $X/2 \leq n \leq 4X$  we can decompose  $f$  into the sum of  $O(\log^{O(1)} X)$  functions  $h$ , each of which is bounded by  $d_2^{O(1)}$  and is of one of the following forms:

Type I:  $h = a * \psi$ , with  $a$  supported on a dyadic subinterval of  $[1, X^\varepsilon]$ , and  $\psi \equiv 1$  or  $\psi \equiv \log^j$  with  $0 \leq j \ll 1$ .

Type II:  $h = a * b$ , with each of  $a, b$  a convolution of at most five functions, each of which is a restriction of one of the functions  $1, \log, \mu$  to an interval, with  $a$  and  $b$  each being supported on some dyadic subinterval of  $[c_0 X^\varepsilon, c_0^{-1} X^{1-\varepsilon}]$  for some constant  $c_0 > 0$ .

By the triangle inequality, to prove (3.32), it will suffice to show that

$$\frac{1}{X} \int_X^{3X} \left| \sum_{\substack{x < n \leq x + H'' \\ n \equiv r \pmod{q}}} h(n) - \frac{H''}{H^*} \sum_{\substack{x < n \leq x + H^* \\ n \equiv r \pmod{q}}} h(n) \right|^2 dx \ll_{A, \varepsilon, \kappa} \delta^{40} (H'')^2. \tag{3.33}$$

Recall that by the choice of  $\delta$  in (3.23), the bound in (3.33) amounts to a power-saving if  $f = d_k$  and  $l = 1$ , and to an arbitrary power of logarithm saving otherwise. By Lemma 3.2, if  $h = a * b$  and we are in the type II case, the sequences  $a$  and  $b$  satisfy (3.5) with  $W = \log^{1000A} X$  in the case  $f \in \{\Lambda, \mu\}$  and with  $W = X^{c_k}$  for some sufficiently small  $c_k$  in the case  $f = d_k$ . Hence, if  $h$  is of type II, then the claim (3.33) follows from Lemma 3.5(i).

Consider then the case where  $h = a * \psi$  is of type I with  $a$  supported on  $[M, 2M]$  with  $1 \leq M \leq X^\varepsilon$  and  $\psi \equiv \log^j$  for some  $j \ll 1$ . Note that if  $j \geq 1$ , we have

$$\psi(\ell) = j \int_1^X \frac{(\log u)^{j-1}}{u} 1_{u \leq m} du.$$

Using this decomposition and the estimate  $\sum_{\ell \in I \cap P} 1 = |I \cap P| + O(1)$  for any interval  $I$  and arithmetic progression  $P$ , for any  $M \in [1, X^\varepsilon]$  we have

$$\begin{aligned} \sum_{\substack{x < n \leq x + H'' \\ n \equiv r \pmod{q}}} h(n) &= \sum_{m \sim M} a(m) \sum_{\substack{x/m < \ell \leq (x + H'')/m \\ \ell m \equiv r \pmod{q}}} \psi(\ell) \\ &= \sum_{m \sim M} a(m) \frac{H''}{H^*} \sum_{\substack{x/m < \ell \leq (x + H^*)/m \\ \ell m \equiv r \pmod{q}}} \psi(\ell) + O\left(\left(\frac{H''}{H^*} + 1\right) M X^{o(1)}\right) \\ &= \frac{H''}{H^*} \sum_{\substack{x < n \leq x + H^* \\ n \equiv r \pmod{q}}} h(n) + O\left(\left(\frac{H''}{H^*} + 1\right) M X^{o(1)}\right). \end{aligned}$$

This implies (3.33) and hence concludes the proof of (3.32) in the  $l = 1$  case.

### 3.4.2 Proof of (3.32) for $l = 2$

Now let  $f \in \{\mu, d_k\}$ . Since  $f$  is multiplicative, we have the decomposition

$$\begin{aligned}
 f(n)g(n) &= \sum_{\substack{n=pm \\ X^\kappa < p \leq X^{\varepsilon/10}}} f(p)f(m) + O(d_k(n)(\log X)1_{p^2|n \text{ for some } p > X^\kappa}) \\
 &:= (a * b)(n) + O(d_k(n)(\log X)1_{p^2|n \text{ for some } p > X^\kappa}),
 \end{aligned}
 \tag{3.34}$$

where  $a(n) = f(n)1_{n \in \mathbb{P} \cap (X^\kappa, X^{\varepsilon/10}]}$ . We substitute this into (3.32), noting that the  $O(\cdot)$  error term contributes  $\ll (H'' X^{-\kappa/3})^2 \log^{O(1)} X$  by the triangle inequality and Shiu’s bound (Lemma 2.18). Hence, it suffices to prove (3.32) with  $f \cdot g$  replaced by the type II sum  $a * b$ . Now (3.32) follows from Lemma 3.5(ii) (after decomposing  $a$  into  $O(\log X)$  dyadically supported sequences), using Lemma 3.2 to verify the first estimate in (3.5).

### 3.5 Theorem 3.1 in the case of large $|T|$

Now we handle the large  $T$  case  $|T| \geq \frac{X}{H} X^{\varepsilon/2}$ .

We recall the notions of type I and type II sums from Sect. 3.4.1. We begin by noting that  $d_k^\sharp$  and  $\Lambda^\sharp$  can be approximated by suitable type I sums. By Lemma 2.19(i), for  $n \in [1, 3X]$  we can approximate

$$\Lambda^\sharp(n) = (a * 1)(n) + E(n),$$

where  $|a(n)| \ll \log X$ ,  $a$  is supported on  $[1, X^{\varepsilon/5}]$ , and for  $X^\varepsilon \leq H \leq x \leq 2X$  we have  $\sum_{x < n \leq x+H} |E(n)| \ll_\varepsilon H \exp(-\log^{1/20} X)$ . By Lemma 2.19(ii), for  $n \in [1, 3X]$  we can write as a sum of  $J = O(1)$  type I sums in the form

$$d_k^\sharp(n) = \sum_{j=1}^J (a_j * \psi_j)(n), \tag{3.35}$$

where  $a_j$  are divisor-bounded sequences supported on  $[1, X^{\varepsilon/5}]$ , and  $\psi_j(n) = \log^j n$ .

We begin with Theorem 3.1(i). We apply Lemma 2.17 and the triangle inequality to reduce to proving Theorem 3.1(i) with  $f - f^\sharp$  for  $f \in \{\mu, \Lambda, d_k\}$  replaced by a type I sum  $a * \psi$  or by a type II sum  $a * b$  (adjusting  $A$  or  $c_{k,C}$  if necessary).

Now, in the case of type I sums, the claim follows from Lemma 3.6 (we can get rid of the possible log factor in the type I convolution by writing

$$a * \log^m = m \int_1^{3X} (a * 1_{[1, \cdot]}) \log^{m-1}(t) \frac{dt}{t}$$

and applying the triangle inequality). In the case of type II sums, if  $f = d_k$ , we can apply Lemma 3.5(iii) with  $W = X^{\varepsilon/4}$  (recalling that  $H \geq X^{1/3+\varepsilon}$ ), since the assumption (3.5) with  $W = X^{\varepsilon/4}$  holds for the sequences  $a(n)n^{iT}$  and  $b(n)n^{iT}$  by Lemma 3.2. If instead  $f \in \{\mu, \Lambda\}$ , we can similarly apply Lemma 3.5(iii) but with

$W = \log^{100A} X$ , again using Lemma 3.2 to verify the assumption (3.5). This completes the proof of Theorem 3.1(i).

We then turn to Theorem 3.1(ii). Recall that  $g(n) = \sum_{X^\kappa < p \leq X^{\varepsilon/10}} 1_{p|n}$ . Applying (3.34) and Shiu’s bound (Lemma 2.18), it suffices to prove Theorem 3.1(ii) with  $f \cdot g$  for  $f \in \{\mu, d_k\}$  replaced by a type II sum  $a * b$ , where  $a(n) = f(n)1_{n \in \mathbb{P} \cap (X^\kappa, X^{\varepsilon/10}]}$  and  $|b(n)| \ll d_2(n)^{O(1)}$ . We also want to show that the term  $d_k^\sharp \cdot g$  arising in the case  $f = d_k$  is a linear combination of type I sums. To this end, note that by (3.35) we can write

$$\begin{aligned}
 d_k^\sharp(n) &= \sum_{X^\kappa < p \leq X^{\varepsilon/10}} 1_{p|n} \\
 &= \sum_{X^\kappa < p \leq X^{\varepsilon/10}} \sum_{j=1}^J \sum_{n=pm}^J a_j(m)\psi_j(\ell)(1_{p|m} + 1_{p|\ell}) \\
 &\quad + O(X^{o(1)} 1_{p^2|n \text{ for some } X^\kappa < p \leq X^{\varepsilon/10}}) \\
 &= \sum_{j=1}^J \sum_{X^\kappa < p \leq X^{\varepsilon/10}} \left( \sum_{n=pm'\ell} a_j(pm')\psi_j(\ell) + \sum_{n=pm'\ell} a_j(m)\psi_j(p\ell') \right) \\
 &\quad + O(X^{o(1)} 1_{p^2|n \text{ for some } X^\kappa < p \leq X^{\varepsilon/10}}),
 \end{aligned} \tag{3.36}$$

where  $a_j$  is a divisor-bounded sequence supported on  $[1, X^{\varepsilon/5}]$  and  $\psi_j(n) = \log^j n$ . The contribution of the  $O(\cdot)$  error term here is negligible by Lemma 2.18. Also note that by the binomial formula  $\psi_j(p\ell')$  is a linear combination of functions of the form  $\psi_r(p)\psi_{j-r}(\ell')$  with  $0 \leq r \leq j$ . We conclude that  $d_k^\sharp \cdot g$  is a sum of  $O(1)$  type I sums.

We then use Lemma 3.5(ii) to handle the type II sums arising from  $f \cdot g$  (noting that the first estimate in (3.5) holds with  $W = \log^{100A} X$  for  $a(n)n^{-iT}$  by the assumption  $|T| \geq \frac{X}{H} X^{\varepsilon/2}$ , Lemma 3.2 and partial summation). Furthermore, we use Lemma 3.6 for the type I sums arising from  $d_k^\sharp \cdot g$ . This concludes the proof of Theorem 3.1(ii).

### 4 Reduction to type II estimates

To complement the major arc estimates in Theorem 3.1, we will establish some “inverse theorems” that yield decorrelation between an arithmetic function  $f$  and a nilsequence  $F(g(n)\Gamma)$  assuming that  $f$  is of “type I”, “type II”, or “type  $I_2$ ”, and the nilsequence is “minor arc” in a suitable sense. To make this precise, we recall some definitions from [30]:

**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  $(\delta, A_I)$  type I sum is an arithmetic function of the form  $f = \alpha * \beta$ , where  $\alpha$  is supported on  $[1, A_I]$ , and one has

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

and

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

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  $\alpha$  is supported on  $[A_{II}^-, A_{II}^+]$ , and one has the bound (4.1) and

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

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

- (iii) (Type I<sub>2</sub> sum) A  $(\delta, A_{I_2})$  type I<sub>2</sub> sum is an arithmetic function of the form  $f = \alpha * \beta_1 * \beta_2$ , where  $\alpha$  is supported on  $[1, A_{I_2}]$  and satisfies (4.1) for all  $A \geq 1$ , and  $\beta_1, \beta_2$  satisfy the bound (4.2) for some  $1 \leq q \leq \frac{1}{\delta}$ .

We are now ready to state the inverse theorem.

**Theorem 4.2** (Inverse theorem) *Let  $d, D \geq 1, 3 \leq H \leq X, 0 < \delta < \frac{1}{\log X}$ , and let  $G/\Gamma$  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*

$$\left| \sum_{x < n \leq x+H} f(n) F(g_x(n)\Gamma) \right|^* \geq \delta H$$

for all  $x$  in a subset  $E$  of  $[X, 2X]$  of measure at least  $\delta X$  where, for each  $x \in E$ ,  $g_x : \mathbb{Z} \rightarrow G$  is a polynomial map.

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

$$H \ll_{d,D} \delta^{-O_{d,D}(1)} A_I \tag{4.3}$$

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

$$\|\eta \circ g_x\|_{C^\infty(x, x+H)} \ll_{d,D} \delta^{-O_{d,D}(1)}$$

for all  $x$  in a subset of  $E$  of measure  $\gg_{d,D} \delta^{O_{d,D}(1)} X$ .

- (ii) (Type II inverse theorem, non-abelian case) If  $\varepsilon > 0$ ,  $f$  is a  $(\delta, A_{II}^-, A_{II}^+)$  type II sum for some  $A_{II}^+ \geq A_{II}^- \geq X^\varepsilon$ ,  $G$  is non-abelian with one-dimensional center, and  $F$  oscillates with a central frequency  $\xi$  of Lipschitz norm at most  $1/\delta$ , then either

$$H \ll_{d,D,\varepsilon} \delta^{-O_{d,D,\varepsilon}(1)} A_{II}^+ \tag{4.4}$$

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

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

for all  $x$  in a subset of  $E$  of measure  $\gg_{d,D,\varepsilon} \delta^{O_{d,D,\varepsilon}(1)} X$ .

- (iii) (Type II inverse theorem, abelian case) If  $\varepsilon > 0$ ,  $f$  is a  $(\delta, A_{II}^-, A_{II}^+)$  type II sum for some  $A_{II}^+ \geq A_{II}^- \geq X^\varepsilon$ , and for every  $x \in E$  one has  $F(g_x(n)\Gamma) = e(P_x(n))$  for some polynomial  $P_x: \mathbb{Z} \rightarrow \mathbb{R}$  of degree at most  $d$ , then either (4.4) holds, or else there exists a real number  $|T| \ll_{d,\varepsilon} \delta^{-O_{d,\varepsilon}(1)} (X/H)^{d+1}$  such that

$$\|e(P_x(n))n^{-iT}\|_{\text{TV}((x,x+H] \cap \mathbb{Z}; q)} \ll_{d,\varepsilon} \delta^{-O_{d,\varepsilon}(1)}$$

for some integer  $1 \leq q \ll_{d,\varepsilon} \delta^{-O_{d,\varepsilon}(1)}$  and for all  $x$  in a subset of  $E$  of measure  $\gg_{d,\varepsilon} \delta^{O_{d,\varepsilon}(1)} X$ .

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

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

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

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

for all  $x$  in a subset of  $E$  of measure  $\gg_{d,D} \delta^{O_{d,D}(1)} X$ .

**Remark 4.3** In the proof of Theorem 4.2, we may clearly assume that  $H \in \mathbb{N}$  and then that the map  $x \mapsto g_x$  is constant on all intervals of the form  $[m, m + 1)$  with  $m \in \mathbb{N}$ . This makes it easy to check the measurability of various subsets of  $E$  defined in terms of  $g_x$ .

We observe that parts (i), (iv) of Theorem 4.2 follow easily from parts (i), (iv) of [30, Theorem 4.2]. For instance, if the hypotheses of Theorem 4.2(i) hold and (4.3) does not hold, then from [30, Theorem 4.2(i)], we see that for each  $x \in E$  there exists a non-trivial horizontal character  $\eta_x: G \rightarrow \mathbb{R}$  of Lipschitz norm  $O_{d,D}(\delta^{-O_{d,D}(1)})$  such that

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

There are only  $O_{d,D}(\delta^{-O_{d,D}(1)})$  possible choices of  $\eta_x$ , so Theorem 4.2(i) follows from the pigeonhole principle and Remark 4.3, which guarantees the measurability

of those subsets of the set of  $x \in E$  for which  $\eta_x$  takes a given value. Theorem 4.2(iv) follows similarly from [30, Theorem 4.2(iv)].

Thus it only remains to establish parts (ii) and (iii) of Theorem 4.2, which do not follow from their counterparts in [30, Theorem 4.2], since the conclusion

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

in [30, Theorem 4.2(ii)–(iii)] is always satisfied when  $H \leq X^{1/2}$ . We shall prove Theorem 4.2(ii)–(iii) in Sect. 6, based on work in Sect. 5.

In this section we show how Theorem 4.2, when combined with the major arc estimates from the previous section, implies Theorem 1.1.

### 4.1 Combinatorial decompositions

We start by describing the combinatorial decompositions (Lemmas 4.4 and 4.5 below) that allow us to reduce sums involving  $\mu$ ,  $\Lambda$ ,  $d_k$  to type I, type II, and type  $I_2$  sums. Lemma 4.4 will be used to prove Theorem 1.1(i)–(iii) and Lemma 4.5 will be used to prove Theorem 1.1(iv)–(v).

As in [30], we notice that while the model function  $\Lambda^\sharp$  is not a type I sum, we can approximate it well by the type I sum

$$\Lambda_I^\sharp(n) := \frac{P(R)}{\varphi(P(R))} \sum_{\substack{d \leq X^{\varepsilon/5} \\ d|(n, P(R))}} \mu(d), \tag{4.5}$$

where we recall that  $R = \exp((\log X)^{1/10})$ . Indeed, by the proof of Lemma 2.19(i) we see that for  $x \in [X, 2X]$  we have

$$\sum_{x < n \leq x+H} |\Lambda_I^\sharp(n) - \Lambda^\sharp(n)| \ll H \exp(-(\log X)^{1/20}). \tag{4.6}$$

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

We state two combinatorial decompositions, one that applies to all of our functions of interest and another one that is more flexible (in the sense of allowing a more restricted type II range) but applies only to the functions  $\mu$  and  $d_k$  (with an additional weight).

**Lemma 4.4** (Combinatorial decompositions of  $\mu$ ,  $\Lambda$ ,  $\Lambda_I^\sharp$ ,  $d_k$  and  $d_k^\sharp$ ) *Let  $\varepsilon > 0$  and  $k \geq 2$  be fixed. Let  $g \in \{\mu, \Lambda, \Lambda_I^\sharp, d_k, d_k^\sharp\}$ . There is a set  $\mathcal{F}$  of size  $O((\log X)^{O(1)})$  consisting of functions  $f : \mathbb{N} \rightarrow \mathbb{R}$  such that, for each  $n \in [X/2, 3X]$ , we have*

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

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

- (i)  $f$  is a  $(\log^{-O(1)} X, O(X^{1/3+\varepsilon/2}))$  type I sum;

- (ii)  $f$  is a  $(\log^{-O(1)} X, O(X^{\varepsilon/2})$  type  $I_2$  sum.
- (iii)  $f$  is a  $(\log^{-O(1)} X, X^{\varepsilon/10}, O(X^{1/3}))$  type II sum.

**Lemma 4.5** (Flexible combinatorial decompositions of  $\mu$  and  $d_k$ ) *Let  $k \geq 1$  be fixed. Let  $A > 0, \varepsilon > 0$ , and  $\eta \in (0, \varepsilon/(10k)]$ . Let  $X \geq H \geq X^\varepsilon$  and  $g \in \{\mu, d_k\}$ , and write  $c_g = 0$  if  $g = \mu$  and  $c_g = k - 1$  if  $g = d_k$ . Also let*

$$Y_g = \sum_{X^{\exp(-\eta^{-2})} < p \leq X^{\varepsilon/10}} \frac{c_g + 1}{p}.$$

*There exist  $K \ll 1$  and functions  $f_j: \mathbb{N} \rightarrow \mathbb{R}$  for  $1 \leq j \leq K$  such that each  $f_j$  is either a  $(\log^{-O(1)} X, X^{\varepsilon/2})$  type I sum or a  $(\log^{-O(1)} X, X^{\exp(-\eta^{-2})}, X^{\varepsilon/10})$  type II sum, and such that the following holds. For any sequence  $\{\omega_n\}$  with  $|\omega_n| \leq 1$  and any  $x \in [X, 2X]$ , we have*

$$\begin{aligned} & \sum_{x < n \leq x+H} (g(n) - g^\sharp(n))\omega_n \\ &= \frac{1}{Y_g} \sum_{x < n \leq x+H} (g(n) - g^\sharp(n)) \sum_{X^{\exp(-\eta^{-2})} < p \leq X^{\varepsilon/10}} 1_{p|n}\omega_n + O(\eta H \log^{c_g} X) \end{aligned}$$

*for all  $x \in [X, 2X]$  outside of a set of measure  $O_{A,\varepsilon,\eta}(X \log^{-A} X)$ , and additionally*

$$\begin{aligned} & \sum_{x < n \leq x+H} (g(n) - g^\sharp(n)) \sum_{X^{\exp(-\eta^{-2})} < p \leq X^{\varepsilon/10}} 1_{p|n}\omega_n \\ &= \sum_{j=1}^K \sum_{x < n \leq x+H} f_j(n)\omega_n + O(HX^{-\exp(-\eta^{-2})/2}) \end{aligned}$$

*for all  $x \in [X, 2X]$ .*

Lemma 4.4 is proved below, and the proof of Lemma 4.5 is given in the next subsection.

Unlike in [30, Sect. 4.1], we do not need to impose in our combinatorial decompositions conditions on decorrelation of type II sums with  $n^{iT}$ . This is thanks to our major arc estimate (Theorem 3.1) which shows that functions such as  $\Lambda(n) - \Lambda^\sharp(n)$  are decorrelated with  $n^{iT}$ . For further discussion on this matter, see [30, Remark 4.6].

We will prove Lemma 4.4 by first decomposing the relevant functions into certain Dirichlet convolutions (using Lemma 2.17 in the proof of Lemma 4.4 for  $\mu$  and  $\Lambda$ ). It is then easy to see that we get appropriate sums.

**Proof of Lemma 4.4** The function  $\Lambda_I^\sharp$  is clearly a  $(\log^{-O(1)} X, X^{\varepsilon/5})$  type I sum by its definition (4.5). Similarly, the function  $d_k^\sharp$  is a sum of  $O(1)$  such sums by Lemma 2.19(ii).

For  $\Lambda$ ,  $\mu$ , and  $d_k$ , we apply Lemma 2.17 with  $L = \lceil 10/\varepsilon \rceil$ . Each of the  $O((\log X)^{O(1)})$  components  $f \in \mathcal{F}$  takes the form

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

for some  $\ell \leq 2L$ , 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,

$$N_1 N_2 \dots N_\ell \asymp X, \tag{4.7}$$

and  $N_i \leq X^{\varepsilon/10}$  for each  $i$  with  $a^{(i)}(n) = \mu(n)1_{(N_i, 2N_i]}(n)$ . Consequently whenever  $N_i > X^{\varepsilon/10}$ , we have  $\|a^{(i)}\|_{\text{TV}(\mathbb{N})} \ll \log X$ .

Without loss of generality we can assume that

$$N_1 \geq N_2 \geq \dots \geq N_\ell. \tag{4.8}$$

Consider first the case  $N_1 > X^{2/3-\varepsilon/2}$ . Since  $2/3 - \varepsilon/2 > \varepsilon/10$  and thus  $\|a^{(1)}\|_{\text{TV}(\mathbb{N})} \ll \log X$ , from (4.7) we see that  $f$  is a  $(\log^{-O(1)} X, O(X^{1/3+\varepsilon/2}))$  type I sum of the form  $\alpha * \beta$  with  $\beta = a^{(1)}$  and  $\alpha = a^{(2)} * \dots * a^{(\ell)}$ .

Henceforth we may assume that  $N_j \leq X^{2/3-\varepsilon/2}$  for each  $j$ . Next consider the case that  $N_1 N_2 > X^{1-\varepsilon/2}$ . Since  $N_1, N_2 \leq X^{2/3-\varepsilon/2}$ , this implies that  $N_1, N_2 > X^{1/3} > X^{\varepsilon/10}$  and thus  $\|a^{(1)}\|_{\text{TV}(\mathbb{N})}, \|a^{(2)}\|_{\text{TV}(\mathbb{N})} \ll \log X$ . Hence, by (4.7), the function  $f$  is a  $(\log^{-O(1)} X, O(X^{\varepsilon/2}))$  type  $I_2$  sum of the form  $f = \alpha * \beta_1 * \beta_2$ , with  $\beta_1 = a^{(1)}$ ,  $\beta_2 = a^{(2)}$ .

In the remaining case, we have  $N_1 N_2 \leq X^{1-\varepsilon/2}$  and hence necessarily  $\ell \geq 3$  and  $X/(N_1 N_2) \gg X^{\varepsilon/2}$ . By (4.8) we have  $N_3 \ll X^{1/3}$  and so there exists  $j \in \{3, \dots, \ell\}$  such that  $X^{\varepsilon/10} \leq N_3 \dots N_j \ll X^{1/3}$  and thus we have a  $(\log^{-O(1)} X, X^{\varepsilon/10}, O(X^{1/3}))$  type II sum of the form  $f = \alpha * \beta$ , where  $\alpha = a^{(3)} * \dots * a^{(j)}$ .  $\square$

### 4.2 Divisor sums in short intervals

In this subsection, we prove Lemma 4.5. We first need an auxiliary result about sums of  $d_k$  in almost all short intervals, which will also be needed later in Sect. 7.

**Lemma 4.6** (Divisor functions in almost all short intervals) *Let  $X \geq 3$ ,  $\varepsilon, \eta > 0$ ,  $k \in \mathbb{N}$ , and  $A > 0$ . Let  $H \in [X^\varepsilon, X^{1-\varepsilon}]$ . Then*

$$\max_{\substack{a, q \in \mathbb{N} \\ a, q \leq \log X \\ (a, q) = 1}} \left| \sum_{x < n \leq x+H} \left( \frac{q}{\varphi(q)} \right)^{k-1} d_k(qn + a) - \frac{1}{(k-1)!} H \log^{k-1} X \right| \leq \eta H \log^{k-1} X \tag{4.9}$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_{A, \varepsilon, \eta, k}(X \log^{-A} X)$ .

We remark that the strong bound on the exceptional set in this lemma is important in what follows, and hence we cannot apply the results of [19, 36] on short sums of  $d_k$  that work in much shorter intervals. We also note that in this section we only need the  $q = 1$  case of the lemma, whereas in Sect. 7 we need the general case.

**Proof of Lemma 4.6** We can clearly assume that  $\eta$  is fixed and small in terms of  $\varepsilon$ . By the union bound, it suffices to show that (4.9) holds without the maximum for any given coprime  $a, q$  in this range.

Fix large  $A \geq 1$  and small  $\varepsilon > 0$ . Let  $H_1 = X/\log^{1000A} X$ . We first claim that for  $x \in [qX, 3qX]$  we have

$$\left| \sum_{\substack{x < n \leq x + qH_1 \\ n \equiv a \pmod{q}}} \left( \left( \frac{q}{\varphi(q)} \right)^{k-1} d_k(n) - \frac{\log^{k-1} X}{(k-1)!} \right) \right| \leq \frac{\eta}{100} H_1 \log^{k-2+\varepsilon} X. \tag{4.10}$$

This estimate follows from standard estimates for the  $d_k$  divisor function in arithmetic progressions; for example [4] gives for coprime  $1 \leq a \leq q \leq x^{\delta_k}$  the estimate

$$\sum_{\substack{n \leq x \\ n \equiv a \pmod{q}}} d_k(n) = \frac{x}{\varphi(q)} \sum_{\substack{n \leq x \\ (n,q)=1}} d_k(n) + O(x^{1-\delta_k}) = \frac{x}{\varphi(q)} P_{k,q}(\log x) + O(x^{1-\delta_k}) \tag{4.11}$$

for some  $\delta_k > 0$  and  $P_{k,q}$  given by,

$$P_{k,q}(t) = \text{Res}_{s=1} \left( \prod_{p|q} \left( 1 - \frac{1}{p^s} \right)^k \zeta(s)^k e^{(s-1)t} \right).$$

Since  $\zeta(s) = (s-1)^{-1} + O(1)$  in the neighborhood of  $s = 1$ , using Cauchy’s formula on the circle  $|z - 1| = 1/t$  we conclude that  $P_{k,q}$  is a polynomial of degree  $k - 1$ , and uniformly in  $1 \leq q \leq x$  and  $t \geq 1$ ,

$$P_{k,q}(t) = \left( \frac{\varphi(q)}{q} \right)^k \frac{t^{k-1}}{(k-1)!} + \sum_{j=0}^{k-2} c_{j,q} t^j, \quad |c_{j,q}| \ll \left( \frac{q}{\varphi(q)} \right)^{10k}.$$

The claim (4.10) follows by subtracting (4.11) at two points, and using the above formula for  $P_{k,q}(t)$  together with the fact that  $q/\varphi(q) \ll \log \log x$  for any  $1 \leq q \leq x$ .

Now, by (4.10), it suffices to show that

$$\left| \sum_{\substack{x < n \leq x + H \\ n \equiv a \pmod{q}}} d_k(n) - \frac{H}{H_1} \sum_{\substack{x < n \leq x + H_1 \\ n \equiv a \pmod{q}}} d_k(n) \right| \leq \frac{\eta}{2} H \log^{k-1} X$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_{A,\varepsilon,\eta,k}(X \log^{-A} X)$ . Let  $\mathcal{S}$  be the set of integers having at least one prime factor in  $(H^\eta, H^{1/2}]$ . Applying Shiu’s bound (Lemma 2.18) to the multiplicative function  $d_k \cdot (1 - 1_{\mathcal{S}})$  (recalling that we can assume that  $\eta$  is small in terms of  $\varepsilon$  and thus small compared to the implied constant in

Shiu’s bound), we reduce to showing that

$$\left| \sum_{\substack{x < n \leq x+H \\ n \equiv a \pmod q}} d_k(n) 1_{\mathcal{S}}(n) - \frac{H}{H_1} \sum_{\substack{x < n \leq x+H_1 \\ n \equiv a \pmod q}} d_k(n) 1_{\mathcal{S}}(n) \right| \leq \frac{\eta}{3} H \log^{k-1} X$$

with the same exceptional set bound. Using the orthogonality of characters, it is enough to show the estimate

$$\left| \sum_{x < n \leq x+H} d_k(n) \chi(n) 1_{\mathcal{S}}(n) - \frac{H}{H_1} \sum_{x < n \leq x+H_1} d_k(n) \chi(n) 1_{\mathcal{S}}(n) \right| \leq \frac{\eta}{3} H \log^{k-1} X \tag{4.12}$$

for all Dirichlet characters  $\chi \pmod q$ .

Denoting by  $\omega_{(P_1, P_2]}(n)$  the number of prime factors of  $n$  from  $(P_1, P_2]$ , we can write

$$\begin{aligned} & d_k(n) \chi(n) 1_{\mathcal{S}}(n) \\ &= \sum_{\substack{n=pm \\ H^{\eta^2} < p \leq H^{1/2}}} \frac{k \chi(p) d_k(m) \chi(m)}{1 + \omega_{(H^{\eta^2}, H^{1/2}]}(m)} + O(d_k(n) 1_{p^2|n \text{ for some } p \in (H^{\eta^2}, H^{1/2}]}) \\ &:= (\alpha * \beta)(n) + O(d_k(n) 1_{p^2|n \text{ for some } p \in (H^{\eta^2}, H^{1/2}]}), \end{aligned} \tag{4.13}$$

where  $\alpha(n) = k \chi(n) 1_{\mathbb{P} \cap (H^{\eta^2}, H^{1/2}]}$  and  $|\beta(n)| \ll d_k(n)$ . The error term in (4.13) has a negligible contribution to the left-hand side of (4.12) by the divisor bound, so it suffices to show that

$$\left| \sum_{x < n \leq x+H} (\alpha * \beta)(n) - \frac{H}{H_1} \sum_{x < n \leq x+H_1} (\alpha * \beta)(n) \right| \leq \frac{\eta}{4} H \log^{k-1} X$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_{A, \varepsilon, \eta, k}(X \log^{-A} X)$ . But this follows immediately from our type II major arc estimate (Lemma 3.5(i)), using Lemma 3.2 and partial summation to verify the assumption there.  $\square$

Utilizing Lemma 4.6, we can prove the following  $d_k$ -weighted (dual) Turán–Kubilius inequality in short intervals, which quickly leads to Lemma 4.5.

**Lemma 4.7** (Turán–Kubilius in short intervals with divisor function weight) *Let  $X \geq 3$ ,  $k \in \mathbb{N}$ ,  $A > 0$  and  $\varepsilon, \varepsilon_1, \varepsilon_2 \in (0, 1/10)$ . Also let  $X^\varepsilon \leq H \leq X^{1-\varepsilon}$  and suppose that  $\varepsilon_1 \leq \varepsilon_2 e^{-2k}$ . Then we have*

$$\sum_{x < n \leq x+H} d_k(n) \left| \sum_{X^{\varepsilon_1} < p \leq X^{\varepsilon_2}} 1_{p|n} - \sum_{X^{\varepsilon_1} < p \leq X^{\varepsilon_2}} \frac{k}{p} \right|$$

$$\leq 20\sqrt{k}H(\log^{k-1} X) \left(\log \frac{\varepsilon_2}{\varepsilon_1}\right)^{1/2} \tag{4.14}$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_{A,\varepsilon,\varepsilon_1,\varepsilon_2,k}(X \log^{-A} X)$ .

**Proof** We may assume that  $X$  is large enough in terms of  $A, \varepsilon, \varepsilon_1, \varepsilon_2, k$  and that  $A$  is large enough in terms of  $k$ .

For any  $Y \geq H' \geq 2$ , let  $\mathcal{E}(Y, H')$  be the set of  $x \in (Y, 2Y]$  for which

$$\left| \sum_{x < n \leq x+H'} d_k(n) - \frac{1}{(k-1)!} H' \log^{k-1} x \right| \geq \frac{1}{100k^2} \left(\log \frac{\varepsilon_2}{\varepsilon_1}\right)^{-1} H' \log^{k-1} Y.$$

Then, by Lemma 4.6 with  $q = 1$ , for any  $Y \geq 2$  and  $H' \in [Y^{\varepsilon/10}, Y]$  we have

$$\text{meas}(\mathcal{E}(Y, H')) \ll_{A,\varepsilon,\varepsilon_1,\varepsilon_2,k} Y \log^{-100A} Y. \tag{4.15}$$

Define

$$\begin{aligned} \mathcal{E} = \{x \in [X, 2X]: & \max_{\frac{1}{2} \leq M \leq H^{0.9}} \frac{1}{M} |\{r \in (M, 2M] \cap \mathbb{N}: \frac{x}{r} \in \mathcal{E}(X/r, H/r)\}| \\ & > \log^{-10A} X\}. \end{aligned}$$

By the union bound, Markov’s inequality and (4.15), we have

$$\begin{aligned} \text{meas}(\mathcal{E}) &\ll (\log X) \max_{M \in [\frac{1}{2}, H^{0.9}]} \frac{\log^{10A} X}{M} \\ &\times \sum_{M < r \leq 2M} \text{meas}(\{x \in [X, 2X]: x/r \in \mathcal{E}(X/r, H/r)\}) \\ &\ll (\log^{10A+1} X) \max_{r \in [1, H^{0.9}]} r \cdot \text{meas}\left(\mathcal{E}\left(\frac{X}{r}, \frac{H}{r}\right)\right) \\ &\ll_{A,\varepsilon,\varepsilon_1,\varepsilon_2,k} X \log^{-A} X. \end{aligned}$$

Now we work with  $x \in [X, 2X] \setminus \mathcal{E}$ . Let us first dispose of the large primes on the left hand side of (4.14). We let

$$\varepsilon'_2 = \frac{k\varepsilon_2}{\log \frac{\varepsilon_2}{\varepsilon_1}}.$$

Note that since  $\varepsilon_1 \leq \varepsilon_2 e^{-2k}$ , we have  $\varepsilon'_2 \leq \varepsilon_2$ . We have chosen  $\varepsilon'_2$  in such a way that (using the inequality  $\log t \leq t^{1/2}$  for  $t > 0$  on the second line)

$$\varepsilon'_2 = \frac{k\varepsilon_2}{\log \frac{\varepsilon_2}{\varepsilon_1}} \leq \frac{k}{10 \log \frac{\varepsilon_2}{\varepsilon_1}}, \tag{4.16}$$

$$\log \frac{\varepsilon_2}{\varepsilon_2'} = \log \frac{\log \frac{\varepsilon_2}{\varepsilon_1}}{k} \leq \frac{1}{k^{1/2}} \left( \log \frac{\varepsilon_2}{\varepsilon_1} \right)^{1/2}. \tag{4.17}$$

Now, using that  $x \in [X, 2X] \setminus \mathcal{E}$ , Mertens' theorem and (4.17) as well as Shiu's bound for  $p$  such that  $x/p \in \mathcal{E}(X/p, H/p)$ , we see that the contribution of primes  $p \in (X^{\varepsilon_2'}, X^{\varepsilon_2}]$  to the left hand side of (4.14) is at most

$$\begin{aligned} & 3k \sum_{X^{\varepsilon_2'} < p \leq X^{\varepsilon_2}} \frac{H}{p} \log^{k-1} X + O_\varepsilon \left( k \sum_{\substack{X^{\varepsilon_2'} < p \leq X^{\varepsilon_2} \\ x/p \in \mathcal{E}(X/p, H/p)}} \frac{H}{p} \log^{k-1} X \right) \\ & \leq 4kH \log^{k-1} X \log \frac{\varepsilon_2}{\varepsilon_2'} \leq 4\sqrt{k}H \log^{k-1} X \left( \log \frac{\varepsilon_2}{\varepsilon_1} \right)^{1/2}. \end{aligned}$$

Notice that we have to be careful when we apply Shiu's bound since the implied constant depends on  $\varepsilon$ , but this is not a problem when we save powers of  $\log X$  as we can assume that  $X$  is large in terms of  $\varepsilon$ .

Now we turn to the contribution of primes  $p \in (X^{\varepsilon_1}, X^{\varepsilon_2'})$  to (4.14). We apply the Cauchy–Schwarz inequality (using again that  $x \in [X, 2X] \setminus \mathcal{E}$ ) to reduce matters to proving

$$\sum_{x < n \leq x+H} d_k(n) \left| \sum_{X^{\varepsilon_1} < p \leq X^{\varepsilon_2'}} 1_{p|n} - \sum_{X^{\varepsilon_1} < p \leq X^{\varepsilon_2'}} \frac{k}{p} \right|^2 \leq 100kH (\log^{k-1} X) \log \frac{\varepsilon_2'}{\varepsilon_1}.$$

Opening the square, the claimed estimate becomes

$$\begin{aligned} & \sum_{X^{\varepsilon_1} < p_1, p_2 \leq X^{\varepsilon_2'}} \sum_{x < n \leq x+H} d_k(n) \left( 1_{p_1|n} - \frac{k}{p_1} \right) \left( 1_{p_2|n} - \frac{k}{p_2} \right) \\ & \leq 100kH (\log^{k-1} X) \log \frac{\varepsilon_2'}{\varepsilon_1}. \end{aligned}$$

For any prime  $p$  dividing  $n$ , we have  $d_k(n) = kd_k(n/p) + O(d_k(n)1_{p^2|n})$ . Using this together with the inequality  $(a - b)^2 \leq 2(a^2 + b^2)$  and the assumption  $x \in [X, 2X] \setminus \mathcal{E}$ , the contribution of  $p_1 = p_2$  to the sum above is crudely

$$\begin{aligned} & \leq 2 \sum_{X^{\varepsilon_1} < p \leq X^{\varepsilon_2'}} \sum_{x < n \leq x+H} d_k(n) \left( 1_{p|n} + \frac{k^2}{p^2} \right) \\ & = 2k \sum_{X^{\varepsilon_1} < p \leq X^{\varepsilon_2'}} \sum_{x/p < m \leq (x+H)/p} d_k(m) + O_k \left( HX^{-\varepsilon_1+o(1)} \right) \end{aligned}$$

$$\begin{aligned} &\leq 3kH(\log^{k-1} X) \sum_{X^{\varepsilon_1} < p \leq X^{\varepsilon'_2}} \frac{1}{p} + O_\varepsilon \left( k \sum_{\substack{X^{\varepsilon_1} < p \leq X^{\varepsilon'_2} \\ x/p \in \mathcal{E}(X/p, H/p)}} \frac{H}{p} \log^{k-1} X \right) \\ &\quad + O_k \left( H X^{-\varepsilon_1 + o(1)} \right) \\ &\leq 4kH(\log^{k-1} X) \log \frac{\varepsilon'_2}{\varepsilon_1}. \end{aligned}$$

In view of this and Mertens' theorem, it suffices to show that, for any  $P_1, P_2 \in [X^{\varepsilon_1}, X^{\varepsilon'_2}]$ , we have for all but  $\ll P_1 P_2 \log^{-A} X$  pairs of distinct primes  $p_1 \sim P_1, p_2 \sim P_2$  the estimate

$$\begin{aligned} &\sum_{x < n \leq x+H} d_k(n) \left( 1_{p_1|n} - \frac{k}{p_1} \right) \left( 1_{p_2|n} - \frac{k}{p_2} \right) \\ &\leq 95 \left( \log \frac{\varepsilon'_2}{\varepsilon_1} \right)^{-1} \frac{H}{p_1 p_2} \log^{k-1} X, \end{aligned} \tag{4.18}$$

say. Using again  $d_k(n) = kd_k(n/p) + O(d_k(n)1_{p^2|n})$  for a prime  $p | n$ , we see that the left-hand side of (4.18) is

$$\begin{aligned} &k^2 \sum_{x/(p_1 p_2) < m \leq (x+H)/(p_1 p_2)} d_k(m) - \frac{k^2}{p_1} \sum_{x/p_2 < m \leq (x+H)/p_2} d_k(m) \\ &\quad - \frac{k^2}{p_2} \sum_{x/p_1 < m \leq (x+H)/p_1} d_k(m) \\ &\quad + \frac{k^2}{p_1 p_2} \sum_{x < m \leq x+H} d_k(m) + O(HX^{o(1)-\varepsilon_1}). \end{aligned}$$

Since  $x \in [X, 2X] \setminus \mathcal{E}$ , for all but  $\ll P_1 P_2 \log^{-A} X$  pairs of distinct primes  $p_1 \sim P_1, p_2 \sim P_2$  the main term is

$$\begin{aligned} &\leq k^2 \frac{H}{(k-1)! p_1 p_2} \left( \log^{k-1} \frac{X}{p_1 p_2} - \log^{k-1} \frac{X}{p_2} - \log^{k-1} \frac{X}{p_1} + \log^{k-1} X \right) \\ &\quad + \left( \log \frac{\varepsilon'_2}{\varepsilon_1} \right)^{-1} \frac{H}{p_1 p_2} \log^{k-1} X. \end{aligned}$$

By the mean value theorem and (4.16), here

$$\begin{aligned} &\log^{k-1} \frac{X}{p_1 p_2} - \log^{k-1} \frac{X}{p_2} - \log^{k-1} \frac{X}{p_1} + \log^{k-1} X \\ &\leq 2(k-1) \log X^{\varepsilon'_2} \log^{k-2} X \end{aligned}$$

$$\leq 2k^2 \left( \log \frac{\varepsilon'_2}{\varepsilon_1} \right)^{-1} \log^{k-1} X,$$

and the desired bound follows since  $\frac{2k^2+2}{(k-1)!} \leq 86$  for all  $k \in \mathbb{N}$ . □

We are now ready to prove Lemma 4.5.

**Proof of Lemma 4.5** Let  $\kappa = \exp(-\eta^{-2})$ , and let  $c_g$  and  $Y_g$  be as in Lemma 4.5. Note that  $Y_g \asymp \eta^{-2}$ . Then, by Lemma 4.7, for any complex numbers  $\omega_n$  with  $|\omega_n| \leq 1$  we have

$$\begin{aligned} & \sum_{x < n \leq x+H} (g(n) - g^\sharp(n)) \omega_n \\ &= \frac{1}{Y_g} \sum_{x < n \leq x+H} (g(n) - g^\sharp(n)) \sum_{X^\kappa < p \leq X^{\varepsilon/10}} 1_{p|n} \omega_n + O(\eta H \log^{c_g} X) \end{aligned}$$

for all  $x \in [X, 2X]$  outside of an exceptional set of measure  $O_{A,\varepsilon,\eta,k}(X \log^{-A} X)$ , proving the first claim of the lemma.

From (3.34) and the divisor bound, we have

$$\sum_{x < n \leq x+H} g(n) \sum_{X^\kappa < p \leq X^{\varepsilon/10}} 1_{p|n} = \sum_{x < n \leq x+H} (a * b)(n) + O(HX^{-\kappa+o(1)}),$$

where  $a, b$  are divisor-bounded sequences and  $a$  is supported on  $[X^\kappa, X^{\varepsilon/10}]$ . On the other hand, from (3.36) we have

$$\sum_{x < n \leq x+H} d_k^\sharp(n) \sum_{X^\kappa < p \leq X^{\varepsilon/10}} 1_{p|n} = \sum_{j=1}^K \sum_{x < n \leq x+H} (a_j * b_j)(n) + O(HX^{-\kappa+o(1)}),$$

where  $K \ll 1$  and each  $a_j, b_j$  is divisor-bounded,  $a_j$  is supported on  $[1, X^{\varepsilon/5+\varepsilon/10}]$  for all  $1 \leq j \leq K$ , and  $b_j = \log^j n$ . This gives the second claim of the lemma. □

### 4.3 Deduction of Theorem 1.1

In this subsection we deduce Theorem 1.1 from Theorem 4.2. We present this deduction in a somewhat more general framework, in order to treat all the cases at once and ease potential applications to other sequences of arithmetic interest.

**Theorem 4.8** (Reduction to decorrelation with Archimedean characters) *Let  $X \geq 3$  and  $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$  for some  $0 \leq \theta < 1$  and let  $\varepsilon > 0$ . Let  $\kappa > 0$ ,  $A \geq 1$ , and  $\delta \in (0, \frac{1}{(\log X)^A})$ . Let  $G/\Gamma$  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  $1/\delta$ . Let  $f : \mathbb{N} \rightarrow \mathbb{R}$  be a function with  $|f(n)| \ll (\log n)^{O(1)} d_2(n)^{O(1)}$ , and suppose that one of the following holds.*

- (i)  $\theta \geq 1/3$  and  $f$  is a sum of  $O((\log X)^A)$  functions, each of which is a  $(\delta^{-1}, X^{\theta+\varepsilon/2})$  type I sum, a  $(\delta^{-1}, X^{(3\theta-1)/2+\varepsilon/2})$  type  $I_2$  sum, or a  $(\delta^{-1}, X^\kappa, X^{\theta+\varepsilon/2})$  type II sum.
- (ii)  $\theta = 0$  and  $f$  is a sum of  $O((\log X)^A)$  functions, each of which is a  $(\delta^{-1}, X^{\theta+\varepsilon/2})$  type I sum or a  $(\delta^{-1}, X^\kappa, X^{\theta+\varepsilon/2})$  type II sum.

Then we have

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H} f(n) \overline{F}(g(n)\Gamma) \right|^* \tag{4.19}$$

$$\ll_{A,d,D,\varepsilon,\kappa} \delta^{-O_{d,D,\kappa}(1)} \sup_{|t| \leq \delta^{-O_{d,D,\kappa}(1)}(X/H)^{d+1}} \left| \sum_{x < n \leq x+H} f(n)n^{it} \right|^* + \delta H$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_{A,d,D,\varepsilon,\kappa}(\delta X)$ .

**Proof of Theorem 1.1** We may assume that  $A > 0$  is large in terms of  $\varepsilon, d, D$  in Theorem 1.1. Now, parts (i)–(iii) of the theorem follow by applying Theorem 4.8(i) (with  $\delta = \log^{-A} X$  or  $\delta = X^{-ck,d,D,\varepsilon}$ ), Lemma 4.4, (4.6), and Theorem 3.1(i). Parts (iv)–(v), in turn, follow by applying Theorem 4.8(ii) (with  $\delta = \log^{-A} X$ ), Lemma 4.5, and Theorem 3.1(ii). □

**Proof of Theorem 4.8** Let  $X^{\theta+\varepsilon} \leq H \leq X^{1-\varepsilon}$  and  $\varepsilon > 0$ , and let  $A > 0$  be large in terms of  $d, D, \varepsilon, \kappa$ . For brevity, we allow all implied constants in  $O(1)$  notation to depend on  $d, D, \kappa$ , and in  $\gg$  notation to depend on  $A, d, D, \varepsilon, \kappa$ . We may assume that  $H$  is an integer.

We use induction on the dimension  $D$  of  $G/\Gamma$ . In view of the triangle inequality, we may assume that  $F$  has mean zero (by decomposing  $F = (F - \int F) + \int F$ ). In particular, this shows that the case  $D = 0$  of the theorem holds, so we now assume that  $D > 0$  and that the theorem has already been established for smaller values of  $D$ . In view of Proposition 2.11 and Lemma 2.18, we may assume that  $F$  oscillates with a central frequency  $\xi: Z(G) \rightarrow \mathbb{R}$  of Lipschitz norm at most  $\delta^{-O(1)}$ . If the center  $Z(G)$  has dimension larger than 1, or  $\xi$  vanishes, then the conclusion follows from induction hypothesis applied to  $G/\ker \xi$  (via Lemma 2.10). Henceforth we assume that  $G$  has one-dimensional center and that  $\xi$  is non-zero. (A zero-dimensional center is not possible since  $G$  is nilpotent and non-trivial.)

After these reductions (and adjusting the value of  $A$ ), we may assume for the sake of contradiction that the following holds: There exists a nilmanifold  $G/\Gamma$  of complexity at most  $1/\delta$  and  $F$  of Lipschitz norm at most  $1/\delta$  such that

$$\left| \sum_{x < n \leq x+H} f(n) F(g_x(n)\Gamma) \right|^* \gg \delta H \tag{4.20}$$

for all  $x$  in a subset  $E$  of  $[X, 2X]$  of measure  $\gg \delta X$  where, for each  $x \in E$ ,  $g_x: \mathbb{Z} \rightarrow G$  is a polynomial map. Thanks to Remark 4.3, we may assume that  $x \mapsto g_x$  is constant on intervals of the form  $[m, m + 1)$  with  $m \in \mathbb{N}$ .

Using the pigeonhole principle, for some type  $I$ , type  $I_2$  or type  $II$  sum  $h$  as in the theorem, one has the bound

$$\left| \sum_{x < n \leq x+H} h(n)F(g_x(n)\Gamma) \right|^* \gg \delta^2 H \tag{4.21}$$

for all  $x$  in a subset  $E'$  of  $E$  of measure  $\gg \delta^2 X$ .

Consider first the case when  $h$  is a  $(\delta^{-1}, X^\kappa, O(X^{\theta+\varepsilon/2}))$  type  $II$  sum and  $G$  is abelian, hence one-dimensional since  $G = Z(G)$ . Then we may identify  $G/\Gamma$  with the standard circle  $\mathbb{R}/\mathbb{Z}$  (increasing the Lipschitz constants for  $F, \xi$  by  $O(\delta^{-O(1)})$  if necessary) and  $\xi$  with an element of  $\mathbb{Z}$  of magnitude  $O(\delta^{-O(1)})$ , and we can write

$$F(y) = Ce(\xi y)$$

for some  $C = O(\delta^{-O(1)})$  and all  $y \in \mathbb{R}/\mathbb{Z}$ . We can write  $\xi \cdot g_x(n)\Gamma = P_x(n) \bmod 1$  for some polynomial  $P_x: \mathbb{Z} \rightarrow \mathbb{R}$  of degree at most  $d$ , thus by (4.20), (4.21) we have

$$\left| \sum_{x < n \leq x+H} f(n)e(-P_x(n)) \right|^* \gg \delta^{O(1)} H \tag{4.22}$$

and

$$\left| \sum_{x < n \leq x+H} h(n)e(-P_x(n)) \right|^* \gg \delta^{O(1)} H \tag{4.23}$$

for all  $x$  in  $E'$ . Theorem 4.2(iii) and (4.23) imply that there exists a real number  $T \ll \delta^{-O(1)}(X/H)^{d+1}$  and an integer  $1 \leq q \ll \delta^{-O(1)}$  such that

$$\|e(P_x(n))n^{-iT}\|_{\text{TV}((x,x+H] \cap \mathbb{Z}; q)} \ll \delta^{-O(1)}$$

for all  $x$  in a subset of  $E'$  of measure  $\gg \delta^{O(1)} X$ . By Lemma 2.2(ii) and (4.22), we thus obtain

$$\left| \sum_{x < n \leq x+H} f(n)n^{-iT} \right|^* \gg \delta^{O(1)} H$$

for all such  $x$ . But now (after adjusting the value of  $A$ ), from (4.20) we obtain the desired bound (4.19).

Now, in all the remaining cases,  $h$  is either a type  $I$  sum, a type  $I_2$  sum, or a type  $II$  sum (with the relevant parameters), and in the last case  $G$  is non-abelian with one-dimensional center. Hence, Theorem 4.2 and (4.21) imply that there exists a non-trivial horizontal character  $\eta: G \rightarrow \mathbb{R}$  of Lipschitz norm  $\delta^{-O(1)}$  such that

$$\|\eta \circ g_x \bmod 1\|_{C^\infty(x,x+H]} \ll \delta^{-O(1)} \tag{4.24}$$

for all  $x$  in a subset of  $E$  of measure at least  $\delta^{O(1)} X$ .

Now that we have (4.24), we can reduce the dimension and apply the induction hypothesis to contradict (4.20) and hence conclude the proof as follows. By (4.24) and Lemma 2.14, we have a decomposition  $g_x = \varepsilon_x g'_x \gamma_x$  for some  $\varepsilon_x, g'_x, \gamma_x \in \text{Poly}(\mathbb{Z} \rightarrow G)$  such that

- (i)  $\varepsilon_x$  is  $(\delta^{-O(1)}, (x, x + H])$ -smooth;
- (ii) There is a  $\delta^{-O(1)}$ -rational proper subnilmanifold  $G'_x / \Gamma'_x$  of  $G / \Gamma$  such that  $g'_x$  takes values in  $G'_x$ ; and
- (iii)  $\gamma_x$  is  $\delta^{-O(1)}$ -rational.

At present, the objects  $\varepsilon_x, G'_x, \Gamma'_x, \gamma_x$  in the above decomposition depend on  $x$ , which is undesirable. However, we can improve this situation by repeated application of the pigeonhole principle as follows (refining the set of  $x$  as necessary).

- Observe from Lemma 2.14 that  $G'_x$  is the kernel of  $\eta$  and is thus automatically independent of  $x$ , thus we may write  $G'_x = G'$ .
- Since  $G'_x / \Gamma'_x$  is a  $\delta^{-O(1)}$ -rational subnilmanifold of  $G / \Gamma$ , the generators of  $\Gamma'_x$  are  $\delta^{-O(1)}$ -rational in  $G$ , thus there are only  $O(\delta^{-O(1)})$  many possible choices for  $\Gamma'_x$ . Applying the pigeonhole principle, and refining  $x$  to a smaller subset of  $E$  of measure  $\delta^{O(1)}$ , we may assume that  $\Gamma'_x = \Gamma'$  is independent of  $x$ .
- Since  $\gamma_x$  is  $\delta^{-O(1)}$ -rational,  $\gamma_x \Gamma$  has a period  $1 \leq q_x \leq \delta^{-O(1)}$ . By the pigeonhole principle as before, we may refine the set of  $x$  and assume that  $q_x = q$  is independent of  $x$ .

In particular we now have

$$g_x \Gamma = \varepsilon_x g'_x \gamma_x \Gamma. \tag{4.25}$$

We can form a partition  $(x, x + H] = P_{x,1} \cup \dots \cup P_{x,r}$  for some  $1 \leq r \leq \delta^{-O(1)}$ , where each  $P_{x,i}$  is an arithmetic progression of modulus  $q$  and, for each  $x, d_G(\varepsilon_x(n), \varepsilon_x(n')) \leq \delta^4$  whenever  $n, n' \in P_{x,i}$  (which can be ensured by the smoothness of  $\varepsilon_x$  as long as  $|P_{x,i}| \leq \delta^C H$  for some sufficiently large constant  $C = C_{d,D,\kappa}$ ). By the triangle inequality for maximal sums (Lemma 2.2(i)), we have

$$\left| \sum_{x < n \leq x+H} f(n) F(g_x(n) \Gamma) \right|^* \leq \sum_{i=1}^r \left| \sum_{n \in P_{x,i}} f(n) F(g_x(n) \Gamma) \right|^*.$$

Thus, by the pigeonhole principle, there exists  $i$  such that

$$\left| \sum_{x < n \leq x+H} f(n) F(g_x(n) \Gamma) \right|^* \ll \delta^{-O(1)} \left| \sum_{n \in P_{x,i}} f(n) F(g_x(n) \Gamma) \right|^*.$$

for all  $x$  in a subset  $E' \subset E$  of measure at least  $\delta^{O(1)} X$ .

Fix such  $i$ . The function  $n \mapsto \gamma_x(n) \Gamma$  is constant on  $P_{x,i}$  and  $\delta^{-O(1)}$ -rational, hence we may write  $\gamma_x(n) \Gamma = \gamma_{x,i} \Gamma$  for all  $n \in P_{x,i}$  and some  $\gamma_{x,i} \in G$  which is  $\delta^{-O(1)}$ -rational. There are only  $O(\delta^{-O(1)})$  possibilities for  $\gamma_{x,i}$ , so by pigeonholing in  $x$  once again we may assume that  $\gamma_{x,i} = \gamma_i$  for some  $\delta^{-O(1)}$ -rational  $\gamma_i \in G$  independent of  $x$ .

On  $P_{x,i}$ , the function  $\varepsilon_x$  ranges in a  $\delta^4$ -ball in  $G$  at distance  $\delta^{-O(1)}$  to the origin. Since the ball of radius  $\delta^{-O(1)}$  around the origin can be covered by  $O(\delta^{-O(1)})$  balls of radius  $\delta^4$ , we may pigeonhole to conclude (after once again refining the set of  $x$ ) that there exists  $\varepsilon_{i,0}$  at distance  $O(\delta^{-O(1)})$  from the origin such that  $d_G(\varepsilon_x(n), \varepsilon_{i,0}) \ll \delta^4$  for all  $n \in P_{x,i}$  and all  $x$  in a subset of  $E$  of measure at least  $\delta^{O(1)}X$ .

Let  $g_{x,i} \in \text{Poly}(\mathbb{Z} \rightarrow G)$  be the polynomial sequence defined by

$$g_{x,i}(n) = \gamma_i^{-1} g'_x(n) \gamma_i,$$

which takes values in  $\gamma_i^{-1} G' \gamma_i$ . For each  $n \in P_{x,i}$  and  $x \in E$  we have from (4.25) and the right-invariance of  $d_G$  that

$$\begin{aligned} |F(g_x(n)\Gamma) - F(\varepsilon_{i,0}\gamma_i g_{x,i}(n)\Gamma)| &\leq \|F\|_{\text{Lip}} \cdot d_G(\varepsilon_x(n) g'_x(n) \gamma_i, \varepsilon_{i,0} g'_x(n) \gamma_i) \\ &= \|F\|_{\text{Lip}} \cdot d_G(\varepsilon_x(n), \varepsilon_{i,0}) \ll \delta^3. \end{aligned}$$

It follows from this and Lemma 2.18 that

$$\left| \sum_{x < n \leq x+H} f(n) F(g_x(n)\Gamma) \right|^* \leq \delta^{-O(1)} \left| \sum_{n \in P_{x,i}} f(n) F(\varepsilon_{i,0}\gamma_i g_{x,i}(n)\Gamma) \right|^* + O(\delta^2 H). \tag{4.26}$$

Note that since  $F$  is  $\delta^{-1}$ -Lipschitz, the function  $F(\varepsilon_{i,0}\gamma_i \cdot)$  is  $\delta^{-O(1)}$ -Lipschitz. Let  $B$  be large in terms of  $d, D, \kappa$ . Then, by Lemma 2.2(i) and the induction hypothesis, we have, for each  $i = 1, \dots, r$ ,

$$\begin{aligned} \left| \sum_{n \in P_{x,i}} f(n) F(\varepsilon_{i,0}\gamma_i g_{x,i}(n)\Gamma) \right|^* &\leq \sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G')} \left| \sum_{x < n \leq x+H} f(n) F(\varepsilon_{i,0}\gamma_i g(n)\Gamma) \right|^* \\ &\ll \delta^B H \end{aligned}$$

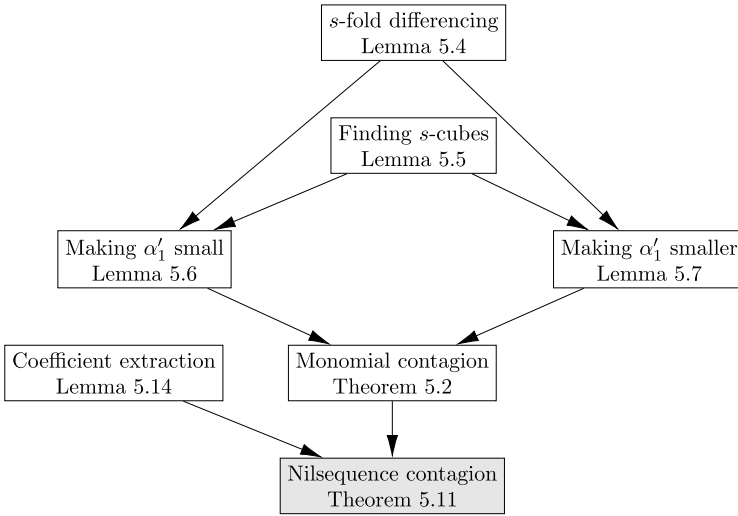
for  $x \in E'$  with  $E' \subset E$  having measure  $\gg \delta^{O(1)}X$ . Combining this with (4.26), we obtain

$$\left| \sum_{x < n \leq x+H} f(n) F(g(n)\Gamma) \right|^* \ll \delta^2 H$$

for  $x \in E'$ , contradicting our assumption (4.20). This completes the proof of Theorem 4.8. □

### 5 Contagion lemmas

In this section we develop the theory of ‘‘contagiousness’’, introduced recently by Walsh [42]. This will culminate in a ‘‘nilsequence contagion lemma’’, Theorem 5.11, which will be a key ingredient in the proof of the type II inverse theorems (Theorem 4.2(ii)–(iii)).



**Fig. 3** The proof of Theorem 5.11. This result will then be used to prove the type II inverse theorems (Theorem 4.2(ii)–(iii)); see Fig. 4

### 5.1 Walsh’s contagion lemma

Let  $P \geq 1$ , let  $S$  be a set of integers in  $[P, 2P]$ , and suppose that  $\alpha : S \rightarrow \mathbb{R}$  is a function obeying the bounds

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

for every  $n \in S$ , some real number  $\alpha_*$ , and some  $\varepsilon > 0$ . Then from the triangle inequality we have

$$\|n_1 \alpha(n_2) - n_2 \alpha(n_1)\|_{\mathbb{R}/\mathbb{Z}} \ll \varepsilon \tag{5.1}$$

for all  $n_1, n_2 \in S$ .

In the recent paper [42], Walsh established the following partial converse to this observation, encapsulating a phenomenon referred to in that paper as “contagiousness”:<sup>9</sup>

**Lemma 5.1** (Walsh’s contagion lemma) [42, Lemma 2.4] *Let  $\eta \in (0, 1/2)$ ,  $P \geq 1$ , and  $\varepsilon \in (0, c/P^2)$  for a sufficiently small absolute constant  $c$ . Let  $S$  be a set of primes in  $[P, 2P]$ , and let  $\alpha : S \rightarrow \mathbb{R}$  be a function satisfying*

$$\|p_1 \alpha(p_2) - p_2 \alpha(p_1)\|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon$$

<sup>9</sup>The name for the terminology relates to the fact, demonstrated by Lemma 5.1, that if we have the relation (5.1) for “many” pairs  $(n_1, n_2) \in S^2$ , then it holds for all  $(n_1, n_2) \in S'^2$  for a “large” subset  $S' \subseteq S$ , so the relations (5.1) are in a sense contagious.

for  $\geq \eta|S|^2$  pairs  $(p_1, p_2) \in S^2$ . Then there is a real number  $\alpha_*$  such that

$$\|\alpha(p) - \alpha_*p\|_{\mathbb{R}/\mathbb{Z}} \ll \varepsilon/P$$

for  $\gg_\eta |S|$  primes  $p \in S$ .

This lemma was then used in [42] to obtain a simplified proof of the local Fourier uniformity estimate

$$\int_X^{2X} \sup_{\alpha \in \mathbb{R}} \left| \sum_{x < n \leq x+H} \mu(n)e(-\alpha n) \right| dx = o(HX),$$

where  $H = X^\theta$  for fixed  $\theta > 0$ .

In order to adapt Lemma 5.1 to our current applications, we will extend and modify it in several ways, which are essential for our application to Theorem 4.2(ii)–(iii):

- (i) We will remove the restriction that  $S$  consists solely of primes.
- (ii) We will make the dependence on the parameter  $\eta$  polynomial in nature.
- (iii) We will relax the bound  $\varepsilon < c/P^2$  to  $\varepsilon < c/P^\kappa$  for arbitrarily small exponents  $\kappa > 0$ .
- (iv) We will replace the linear polynomials  $p_1, p_2, p$  in Lemma 5.1 with monomials  $p_1^s, p_2^s, p^s$  of arbitrary degree  $s$ .
- (v) We will then replace the frequency functions  $\alpha$  by nilsequences.

In addition, for technical reasons and to facilitate an induction argument, we will replace the frequency function  $\alpha$  with a pair  $\alpha_1, \alpha_2$  of frequency functions.

### 5.2 Monomial contagion

We begin by establishing a “monomial contagion lemma” that achieves the first four (i)–(iv) of the five objectives stated above.

**Theorem 5.2 (Monomial contagion lemma)** *Let  $s \in \mathbb{N}$  and  $\kappa > 0$ . There exists  $C_{s,\kappa} \geq 1$  such that the following holds. Let  $0 < \eta < 1/2$ ,  $P \geq \eta^{-C_{s,\kappa}}$  and  $0 < \varepsilon < \eta^{C_{s,\kappa}}/P^\kappa$ . Let  $S_1, S_2$  be subsets of  $[P, 2P] \cap \mathbb{Z}$  and let  $\alpha_1: S_1 \rightarrow \mathbb{R}, \alpha_2: S_2 \rightarrow \mathbb{R}$  be two functions satisfying*

$$\|n_1^s \alpha_2(n_2) - n_2^s \alpha_1(n_1)\|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon \tag{5.2}$$

for all  $(n_1, n_2)$  in a subset  $\mathcal{N}_0$  of  $S_1 \times S_2$  of cardinality at least  $\eta P^2$ . Then there is  $\alpha_* \in \mathbb{R}$  and an integer  $1 \leq q \ll \eta^{-O_{s,\kappa}(1)}$  such that, for  $\gg \eta^{O_{s,\kappa}(1)} P^2$  pairs  $(n_1, n_2) \in \mathcal{N}_0$ , one has

$$\|q(\alpha_i(n_i) - \alpha_* n_i^s)\|_{\mathbb{R}/\mathbb{Z}} \ll \eta^{-O_{s,\kappa}(1)} \varepsilon/P^s \tag{5.3}$$

for  $i = 1, 2$ .

In fact, in our proof the factor of  $q$  in (5.3) is only needed for  $i = 2$ , but we will not exploit this refinement of the theorem here. Because  $n_1, n_2$  are no longer restricted to be prime, the factor of  $q$  cannot be eliminated entirely; a simple counterexample is when  $S_1 = S_2 = [P, 2P] \cap \mathbb{Z}$ ,  $s = 1$ ,  $\alpha_1(n_1) = 0$ ,  $\alpha_2(n_2) = 1/2$ ,  $\eta \asymp 1$ ,  $P$  is large and  $\varepsilon$  is infinitesimal, and  $\mathcal{N}_0$  is the set of pairs  $(n_1, n_2) \in S_1 \times S_2$  with  $n_1, n_2 = 2 \pmod 4$ . In this example, there is no way to subtract off a common linear factor  $\alpha_* n_i$  from  $\alpha_i(n_i)$  for  $i = 1, 2$  to make them both close to zero for many  $(n_1, n_2) \in \mathcal{N}_0$ .

**Remark 5.3** Strictly speaking, Theorem 5.2 is not a strengthening of Lemma 5.1, both because of the introduction of the new quantity  $q$  discussed above, and also because the density parameter  $\eta$  is relative to the integers in  $[P, 2P]$  rather than the primes in  $[P, 2P]$ , thus in principle incurring additional losses of log  $P$ . However, these losses turn out to not be significant for our current application.

In the rest of this subsection we prove Theorem 5.2. Let  $s, \kappa$  be fixed, and assume  $C_{s,\kappa}$  is sufficiently large. We allow implied constants to depend on  $s, \kappa$ , but not on  $C_{s,\kappa}$ . Let  $\eta, \varepsilon, P, S_1, S_2, \alpha_1, \alpha_2, \mathcal{N}_0$  be as in the statement of the theorem. In all the lemmas in this subsection, the assumptions are as in Theorem 5.2.

We will make a number of modifications to the pair  $(\alpha_1, \alpha_2)$  in order to improve the estimates on the individual functions  $\alpha_1, \alpha_2$ . To keep track of these modifications, it is convenient to introduce the set  $\mathcal{A}$  of all pairs  $(\alpha'_1, \alpha'_2)$  of functions  $\alpha'_i: S \rightarrow \mathbb{R}$  which are “equivalent” to the original pair  $(\alpha_1, \alpha_2)$  in the sense that there exists a real number  $\alpha_*$  such that

$$\alpha'_i(n_i) = \alpha_i(n_i) - \alpha_* n_i^s \pmod 1$$

for all  $n_i \in S_i$  and  $i = 1, 2$ . Clearly  $(\alpha_1, \alpha_2)$  lies in  $\mathcal{A}$ . Also, if  $(\alpha'_1, \alpha'_2)$  lies in  $\mathcal{A}$ , and  $\beta$  is a real number, then the pair  $(\alpha''_1, \alpha''_2)$  defined by

$$\alpha''_i(n_i) := \{\alpha'_i(n_i) - \beta n_i^s\}$$

also lies in  $\mathcal{A}$  (recall that  $\{x\} \in (-1/2, 1/2]$  is the signed fractional part of  $x$ ). Finally, observe that if  $(\alpha'_1, \alpha'_2)$  lies in  $\mathcal{A}$ , we have

$$n_1^s \alpha'_2(n_2) - n_2^s \alpha'_1(n_1) = n_1^s \alpha_2(n_2) - n_2^s \alpha_1(n_1) \pmod 1$$

for any  $n_1 \in S_1, n_2 \in S_2$ , and hence from (5.2) we have

$$\|n_1^s \alpha'_2(n_2) - n_2^s \alpha'_1(n_1)\|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon \tag{5.4}$$

for all  $(n_1, n_2) \in \mathcal{N}_0$ .

Our strategy will be to locate pairs  $(\alpha'_1, \alpha'_2) \in \mathcal{A}$  for which one has good bounds on the magnitude of  $\alpha'_1$ ; our first attempt to do so will only locate such a pair obeying a relatively weak bound

$$|\alpha'_1(n_1)| \leq 2^s \varepsilon$$

for many  $n_1$ , but by an iterative process we will eventually be able to locate a pair obeying the stronger bound

$$|\alpha'_1(n_1)| \ll \frac{\eta^{-O(1)}}{P^s} \varepsilon$$

for many  $n_1$ . Once we have obtained this control on  $\alpha'_1(n_1)$ , we will be able to obtain similar control on  $\alpha'_2(n_2)$  for many  $n_2$  (after multiplying by a small integer  $q$ ), at which point we can establish the theorem.

We turn to the details. To gain control on  $\alpha_1$  we will “differentiate” the inequality (5.2) (or (5.4))  $s$  times in the  $n_2$  variable in order to mostly eliminate the  $n_2^s$  term. More precisely, we can take  $s$ -fold differences via the following lemma (where, as per the convention above, the assumptions are as in Theorem 5.2).

**Lemma 5.4** (Taking  $s$ -fold differences) *Let  $(\alpha'_1, \alpha'_2) \in \mathcal{A}$ , and let  $\mathcal{N}$  be a subset of  $\mathcal{N}_0$ . Let  $n_{2,0}$  be an integer, and let  $\vec{h} = (h_1, \dots, h_s)$  be a tuple of integers, with the property that*

$$(n_1, n_{2,0} + \vec{\omega} \cdot \vec{h}) \in \mathcal{N}_0 \tag{5.5}$$

whenever  $(n_1, n_{2,0}) \in \mathcal{N}$  and  $\vec{\omega} \in \{0, 1\}^s$ , with  $\cdot$  denoting the usual inner product. Then one has

$$\|h\alpha'_1(n_1) - \beta n_1^s\|_{\mathbb{R}/\mathbb{Z}} \leq 2^s \varepsilon$$

whenever  $(n_1, n_{2,0}) \in \mathcal{N}$ , where  $h$  is the integer

$$h := s!h_1 \cdots h_s$$

and  $\beta$  is the real number

$$\beta := \sum_{\vec{\omega} \in \{0,1\}^s} (-1)^{|\vec{\omega}|} \alpha'_2(n_{2,0} + \vec{\omega} \cdot \vec{h}). \tag{5.6}$$

**Proof** By (5.5) and (5.4), we have

$$\|n_1^s \alpha'_2(n_{2,0} + \vec{\omega} \cdot \vec{h}) - (n_{2,0} + \vec{\omega} \cdot \vec{h})^s \alpha'_1(n_1)\|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon$$

whenever  $(n_1, n_{2,0}) \in \mathcal{N}$  and  $\vec{\omega} \in \{0, 1\}^s$ . Hence by the triangle inequality and (5.6)

$$\left\| \beta n_1^s - \sum_{\vec{\omega} \in \{0,1\}^s} (-1)^{|\vec{\omega}|} (n_{2,0} + \vec{\omega} \cdot \vec{h})^s \alpha'_1(n_1) \right\|_{\mathbb{R}/\mathbb{Z}} \leq 2^s \varepsilon.$$

From binomial expansion of  $(n_{2,0} + \vec{\omega} \cdot \vec{h})^s$  (paying particular attention to the  $\omega_1 \cdots \omega_s$  terms) we observe the identity

$$\sum_{\vec{\omega} \in \{0,1\}^s} (-1)^{|\vec{\omega}|} (n_{2,0} + \vec{\omega} \cdot \vec{h})^s = s!h_1 \cdots h_s = h$$

and the claim follows. □

In order to use the above lemma, we need to be able to ensure a plentiful supply of pairs  $(n_1, n_{2,0})$  obeying (5.5), with a value of  $h$  that we can “tune” to be of a specified size. To achieve this we establish the following version of the Hilbert cube lemma [14], [37, Exercise 1.3.2].

**Lemma 5.5** (Finding combinatorial cubes in dense sets) *Let  $s \geq 1$  be fixed and let  $\eta > 0$ . Let  $\mathcal{N}$  be a subset of  $\mathcal{N}_0$  of cardinality  $\gg \eta^{O(1)} P^2$ , and let  $1 \leq H \leq P$ . Then either  $H \leq \eta^{-O(1)}$ , or else (if the constant  $C_{s,\kappa}$  is large enough depending on the implied constants in the preceding hypothesis) there exists a tuple  $\vec{h} = (h_1, \dots, h_s)$  of positive integers  $\eta^{O(1)} H \ll h_1, \dots, h_s \leq H$  such that one has the containment*

$$(n_1, n_2 + \vec{\omega} \cdot \vec{h})_{\vec{\omega} \in \{0,1\}^s} \subset \mathcal{N}^{(0,1)^s} \tag{5.7}$$

for  $\gg \eta^{O(1)} P^2$  choices of pairs  $(n_1, n_2) \in \mathcal{N}_0$ , where  $\cdot$  denotes the usual dot product.

**Proof** Let  $C$  be a large constant (depending on  $s, \kappa$ , and the implied constants in the hypotheses) to be chosen later. We may assume that  $H \geq \eta^{-C}$ , since otherwise the conclusion follows trivially. Given a tuple  $(f_{\vec{\omega}})_{\vec{\omega} \in \{0,1\}^s}$  of functions  $f_{\vec{\omega}}: \mathbb{Z}^2 \rightarrow \mathbb{R}$  supported on a finite set, we define the Gowers box inner product

$$\langle (f_{\vec{\omega}})_{\vec{\omega} \in \{0,1\}^s} \rangle_{\square^s} := \sum_{n_1, n_2 \in \mathbb{Z}} \sum_{\vec{h}^0, \vec{h}^1 \in [1, H]^s} \prod_{\vec{\omega} \in \{0,1\}^s} f_{\vec{\omega}}(n_1, n_2 + h_1^{\omega_1} + \dots + h_s^{\omega_s})$$

where  $\vec{\omega} = (\omega_1, \dots, \omega_s)$  and  $\vec{h}^0 = (h_1^0, \dots, h_s^0)$ ,  $\vec{h}^1 = (h_1^1, \dots, h_s^1)$  are understood to have integer components. We then define the associated Gowers box norm

$$\|f\|_{\square^s} := \langle (f)_{\vec{\omega} \in \{0,1\}^s} \rangle_{\square^s}^{1/2^s}.$$

From the Gowers–Cauchy–Schwarz inequality (see e.g., [5, Lemma B.2]) we have

$$|\langle (f_{\vec{\omega}})_{\vec{\omega} \in \{0,1\}^s} \rangle_{\square^s}| \leq \prod_{\vec{\omega}' \in \{0,1\}^s} \|f_{\vec{\omega}'}\|_{\square^s}. \tag{5.8}$$

Let  $\mathcal{Q} \subset \mathbb{N}$  be the set of pairs  $(n_1, n_2) \in \mathbb{Z}^2$  with  $n_1 \in [P, 2P]$  and  $|n_2| \leq (s+2)P$ . Then a routine computation gives

$$\|1_{\mathcal{Q}}\|_{\square^s} \ll (H^{2s} P^2)^{1/2^s}.$$

We apply (5.8) with  $f_{0,\dots,0} := 1_{\mathcal{N}}$  and  $f_{\vec{\omega}} := 1_{\mathcal{Q}}$  for all other  $\vec{\omega} \in \{0, 1\}^s$ , to obtain

$$\langle (f_{\vec{\omega}})_{\vec{\omega} \in \{0,1\}^s} \rangle_{\square^s} \ll \|1_{\mathcal{N}}\|_{\square^s} (H^{2s} P^2)^{1-1/2^s}.$$

On the other hand, if  $(n_1, n_2 + h_1^0 + \dots + h_s^0) \in \mathcal{N}$  for some  $n_1, n_2 \in \mathbb{Z}$  and  $\vec{h}^0, \vec{h}^1 \in [1, H]^s$ , then  $(n_1, n_2 + h_1^{\omega_1} + \dots + h_s^{\omega_s}) \in \mathcal{Q}$  for all  $\vec{\omega} \in \{0, 1\}^s$ . Thus we can expand

$$\begin{aligned} \langle (f_{\vec{\omega}})_{\vec{\omega} \in \{0,1\}^s} \rangle_{\square^s} &= \sum_{n_1, n_2 \in \mathbb{Z}} \sum_{\vec{h}^0, \vec{h}^1 \in [1, H]^s} 1_{\mathcal{N}}(n_1, n_2 + h_1^0 + \dots + h_s^0) \\ &= \sum_{n_1, n_2 \in \mathbb{Z}} \sum_{\vec{h}^0, \vec{h}^1 \in [1, H]^s} 1_{\mathcal{N}}(n_1, n_2) \end{aligned}$$

$$\begin{aligned} &\gg H^{2s} |\mathcal{N}| \\ &\gg \eta^{O(1)} H^{2s} P^2. \end{aligned}$$

Combining these inequalities, we conclude that

$$\|1_{\mathcal{N}}\|_{\square^s}^{2^s} \gg \eta^{O(1)} H^{2s} P^2$$

and thus

$$\sum_{n_1, n_2 \in \mathbb{Z}} \sum_{\vec{h}^0, \vec{h}^1 \in [1, H]^s} \prod_{\vec{\omega} \in \{0, 1\}^s} 1_{\mathcal{N}}(n_1, n_2 + h_1^{\omega_1} + \dots + h_s^{\omega_s}) \gg \eta^{O(1)} H^{2s} P^2.$$

By swapping coefficients of  $\vec{h}^0$  and  $\vec{h}^1$  as necessary and using symmetry, we may reduce to the contribution in which  $h_i^1 \geq h_i^0$  for all  $i = 1, \dots, s$ . Replacing  $n_2$  by  $n_2 + h_1^0 + \dots + h_s^0$  and writing  $\vec{h}^1 - \vec{h}^0 = \vec{h}$ , we conclude

$$\sum_{n_1, n_2 \in \mathbb{Z}} \sum_{\vec{h} \in [0, H]^s} \prod_{\vec{\omega} \in \{0, 1\}^s} 1_{\mathcal{N}}(n_1, n_2 + \vec{\omega} \cdot \vec{h}) \gg \eta^{O(1)} H^s P^2.$$

The contribution of those  $\vec{h}$  with at least one coefficient less than  $\eta^C H$  can be easily bounded by  $O(\eta^C H^s P^2)$ . Thus for  $C$  large enough we have

$$\sum_{n_1, n_2 \in \mathbb{Z}} \sum_{\vec{h} \in [\eta^C H, H]^s} \prod_{\vec{\omega} \in \{0, 1\}^s} 1_{\mathcal{N}}(n_1, n_2 + \vec{\omega} \cdot \vec{h}) \gg \eta^{O(1)} H^s P^2.$$

By the pigeonhole principle, there thus exists  $\vec{h} \in [\eta^C H, H]^s$  such that

$$\sum_{n_1, n_2 \in \mathbb{Z}} \prod_{\vec{\omega} \in \{0, 1\}^s} 1_{\mathcal{N}}(n_1, n_2 + \vec{\omega} \cdot \vec{h}) \gg \eta^{O(1)} P^2$$

and the claim follows. □

As a first application of the above two lemmas, we locate a pair  $(\alpha'_1, \alpha'_2) \in \mathcal{A}$  with  $\alpha'_1$  somewhat small.

**Lemma 5.6** (Making  $\alpha'_1(n)$  somewhat small) *There exists  $(\alpha'_1, \alpha'_2) \in \mathcal{A}$  such that*

$$|\alpha'_1(n_1)| \leq 2^s \varepsilon$$

for  $\gg \eta^{O(1)} P^2$  pairs  $(n_1, n_2) \in \mathcal{N}_0$ .

**Proof** Applying Lemma 5.5 with  $H = \eta^{-C}$  for some sufficiently large constant  $C$ , we can find a tuple  $\vec{h} = (h_1, \dots, h_s)$  of positive integers  $h_1, \dots, h_s \ll \eta^{-O(1)}$  and a subset  $\mathcal{N}$  of  $\mathcal{N}_0$  of cardinality  $\gg \eta^{O(1)} P^2$  such that

$$(n_1, n_2 + \vec{\omega} \cdot \vec{h}) \in \mathcal{N}_0$$

for all  $(n_1, n_2) \in \mathcal{N}$  and  $\vec{\omega} \in \{0, 1\}^s$ .

By pruning all sparse columns from  $\mathcal{N}$ , we may assume that whenever  $(n_1, n_2) \in \mathcal{N}$ , then one also has  $(n_1, n'_2) \in \mathcal{N}$  for  $\gg \eta^{O(1)}P$  choices of  $n'_2 \in S_2$ .

By the pigeonhole principle, we may find  $n_2^0 \in S_2$  such that  $(n_1, n_2^0) \in \mathcal{N}$  for  $\gg \eta^{O(1)}P$  choices of  $n_1 \in S_1$ . Applying Lemma 5.4, we conclude that

$$\|h\alpha_1(n_1) - \beta n_1^s\|_{\mathbb{R}/\mathbb{Z}} \leq 2^s \varepsilon \tag{5.9}$$

whenever  $(n_1, n_2^0) \in \mathcal{N}$ , for some real number  $\beta$  (not depending on  $n_1$ ) and with

$$h := s!h_1 \cdots h_s.$$

In particular  $1 \leq h \ll \eta^{-O(1)}$ . From (5.9) we then see that whenever  $(n_1, n_2^0) \in \mathcal{N}$ , we have

$$\left\| \alpha_1(n_1) - \frac{\beta}{h} n_1^s - \frac{a}{h} \right\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{2^s \varepsilon}{h} \tag{5.10}$$

for some integer  $1 \leq a \leq h$ . By the pigeonhole principle, we can then find a single  $1 \leq a \leq h$  and a subset  $S'_1$  of  $S_1$  of size  $\gg \eta^{O(1)}P$  such that for all  $n_1 \in S'_1$ , one has  $(n_1, n_2^0) \in \mathcal{N}$ , and (5.10) holds for this fixed choice of  $a$ .

For technical reasons we need to analyze how  $n_1$  shares primes in common with  $h$ . Let  $\mathcal{H}$  be the multiplicative semigroup generated by the primes dividing  $h$ . We observe that

$$\begin{aligned} \sum_{q \in \mathcal{H} \cap \mathbb{N}} \frac{1}{q^{1/2}} &= \prod_{p|h} \left(1 - \frac{1}{p^{1/2}}\right)^{-1} \ll \exp\left(\sum_{p|h} \frac{2}{p^{1/2}}\right) \\ &\ll \exp\left(\sum_{p|h} \log p\right) \ll \exp(\log h) \ll \eta^{-O(1)}. \end{aligned}$$

In particular, for any  $Q \geq 1$ , we have

$$\sum_{\substack{q \in \mathcal{H} \cap \mathbb{N} \\ q \geq Q}} \frac{1}{q} \ll \eta^{-O(1)} Q^{-1/2}.$$

From this and the union bound, we see that if  $1 \leq Q \leq P$ , then the number of those  $n_1 \in S'_1$  that are divisible by some  $q \in \mathcal{H} \cap \mathbb{N}$  with  $q \geq Q$  is at most  $O(\eta^{-O(1)}Q^{-1/2}P)$ . Thus, for a suitable choice of  $Q = \eta^{-O(1)}$ , such  $n_1$  only occupy at most half (say) of the set  $S'_1$ . By the pigeonhole principle, we may thus find  $q \in \mathcal{H} \cap \mathbb{N}$  with  $1 \leq q < Q$  and a subset  $S''_1$  of  $S'_1$  of cardinality  $\gg \eta^{O(1)}P$  with the property that for all  $n_1 \in S''_1$ ,  $q$  is the largest element of  $\mathcal{H} \cap \mathbb{N}$  that divides  $n_1$ , thus  $n_1 = qn'_1$  where  $n'_1$  is coprime to  $h$ . By further application of the pigeonhole principle, we may also restrict  $n'_1$  to a single primitive residue class  $n'_1 = b \pmod h$  of  $h$ . Letting  $k$  be a positive integer that solves  $kb^s \equiv a \pmod h$ , we then have

$$\frac{k}{q^s h} n_1^s = \frac{k(n'_1)^s}{h} = \frac{a}{h} \pmod 1$$

for all  $n_1 \in S'_1$ . If we then define

$$\beta' := \frac{\beta}{h} + \frac{k}{q^s h}$$

we conclude from (5.10) that

$$\|\alpha_1(n_1) - \beta' n_1^s\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{2^s \varepsilon}{h}$$

for all  $n_1 \in S'_1$ . Because we have previously pruned all sparse columns from  $\mathcal{N}$ , we conclude that

$$\|\alpha_1(n_1) - \beta' n_1^s\|_{\mathbb{R}/\mathbb{Z}} \leq \frac{2^s \varepsilon}{h}$$

for  $\gg \eta^{O(1)} P^2$  of the pairs  $(n_1, n_2) \in \mathcal{N}_0$ . Setting  $\alpha'_i(n_i) := \{\alpha_i(n_i) - \beta' n_i^s\}$  for  $i = 1, 2$ , we obtain the claim.  $\square$

Next, we show that if we have already found a pair  $(\alpha'_1, \alpha'_2) \in \mathcal{A}$  in which  $\alpha'_1$  is often of some small size  $O(\mu)$ , we can find a new pair  $(\alpha''_1, \alpha''_2)$  in which  $\alpha''_1$  is often as small as  $O\left(\eta^{-O(1)} \varepsilon \left(\mu + \frac{1}{P^s}\right)\right)$ .

**Lemma 5.7** (Making  $\alpha'_1(n)$  even smaller) *Let  $\mu > 0$ , and suppose that there is  $(\alpha'_1, \alpha'_2) \in \mathcal{A}$  such that*

$$|\alpha'_1(n_1)| \leq \mu \tag{5.11}$$

*for  $\gg \eta^{O(1)} P^2$  of the pairs  $(n_1, n_2) \in \mathcal{N}_0$ . Then (if  $C_{s,\kappa}$  is sufficiently large depending on the implied constants in the hypothesis) there exists  $(\alpha''_1, \alpha''_2) \in \mathcal{A}$  such that*

$$|\alpha''_1(n_1)| \ll \eta^{-O(1)} \varepsilon \left(\mu + \frac{1}{P^s}\right)$$

*for  $\gg \eta^{O(1)} P^2$  pairs  $(n_1, n_2) \in \mathcal{N}_0$ .*

**Proof** Let  $C_*$  be a large constant depending only on  $s, \kappa$  to be chosen later; until this constant is selected, we do *not* allow implied constants to depend on  $C_*$ . The constant  $C_{s,\kappa}$  will be large in terms of  $C_*$ . We may assume that  $\mu \leq \eta^{C_*}$ , since otherwise the claim follows from Lemma 5.6.

Applying Lemma 5.5 with  $H = \eta^{C_*/(2s)} \min(\mu^{-1/s}, P)$ , we can find a tuple  $\vec{h} = (h_1, \dots, h_s)$  of positive integers  $\eta^{O(1)} H \ll h_1, \dots, h_s \leq H$  and a subset  $\mathcal{N}$  of  $\mathcal{N}_0$  of cardinality  $\gg \eta^{O(1)} P^2$  such that

$$(n_1, n_2 + \vec{\omega} \cdot \vec{h}) \in \mathcal{N}_0$$

for all  $(n_1, n_2) \in \mathcal{N}$  and  $\vec{\omega} \in \{0, 1\}^s$ . By pruning as in the proof of Lemma 5.6, we may assume that whenever  $(n_1, n_2) \in \mathcal{N}$ , then one also has  $(n_1, n'_2) \in \mathcal{N}$  for  $\gg \eta^{O(1)} P$  choices of  $n'_2 \in S_2$ .

By the pigeonhole principle, we may find  $n_2^0 \in S_2$  such that  $(n_1, n_2^0) \in \mathcal{N}$  for  $\gg \eta^{O(1)}P$  choices of  $n_1 \in S_1$ . Applying Lemma 5.4 as before, we have

$$\|h\alpha'_1(n_1) - \beta n_1^s\|_{\mathbb{R}/\mathbb{Z}} \leq 2^s \varepsilon \tag{5.12}$$

whenever  $(n_1, n_2^0) \in \mathcal{N}$ , where  $\beta$  is a real number and  $h$  is a positive integer of size  $\eta^{O(1)}Y \ll h \ll Y$ , where  $Y := (\eta^{C_*/(2s)} \min(\mu^{-1/s}, P))^s = \eta^{C_*/2} \min(\mu^{-1}, P^s)$ .

From (5.11) we have

$$|h\alpha'_1(n_1)| \ll \eta^{C_*/2} \tag{5.13}$$

and hence

$$\|\beta n_1^s\|_{\mathbb{R}/\mathbb{Z}} \ll \eta^{C_*/2}$$

for  $\gg \eta^{O(1)}P$  choices of  $n_1 \in S_1$ . Applying Lemma 2.3, we may find a positive integer  $1 \leq q \ll \eta^{-O(1)}$  such that

$$\|q\beta\|_{\mathbb{R}/\mathbb{Z}} \ll \eta^{C_*/2 - O(1)} / P^s. \tag{5.14}$$

From (5.12) we have

$$\|qh\alpha'_1(n_1) - \{q\beta\}n_1^s\|_{\mathbb{R}/\mathbb{Z}} \ll \eta^{-O(1)}\varepsilon \tag{5.15}$$

whenever  $(n_1, n_2^0) \in \mathcal{N}$ . On the other hand from (5.13) and (5.14) we have

$$|qh\alpha'_1(n_1) - \{q\beta\}n_1^s| \ll \eta^{C_*/2 - O(1)}.$$

For  $C_*$  large enough, this implies that  $qh\alpha'_1(n_1) - \{q\beta\}n_1^s$  has magnitude less than  $1/2$ , and so from (5.15) we now have

$$|qh\alpha'_1(n_1) - \{q\beta\}n_1^s| \ll \eta^{-O(1)}\varepsilon$$

whenever  $(n_1, n_2^0) \in \mathcal{N}$ . Thus

$$\left| \alpha'_1(n_1) - \frac{\{q\beta\}}{qh} n_1^s \right| \ll \eta^{-O(1)} \frac{\varepsilon}{qh} \ll \eta^{-C_*/2 - O(1)} \varepsilon \left( \mu + \frac{1}{P^s} \right).$$

Setting

$$\alpha''_i(n_i) := \alpha'_i(n_i) - \frac{\{q\beta\}}{qh} n_i^s$$

for  $i = 1, 2$ , we obtain the claim. □

If we apply Lemma 5.6 followed by  $j$  applications of Lemma 5.7, we see that for any fixed  $j$  (depending on  $s, \kappa$ ) we can find  $(\alpha'_1, \alpha'_2) \in \mathcal{A}$  such that

$$|\alpha'_1(n_1)| \ll \eta^{-O(1)} \varepsilon \left( \varepsilon^j + \frac{1}{P^s} \right)$$

for  $\gg \eta^{O(1)} P^2$  pairs  $(n_1, n_2) \in \mathcal{N}_0$ . Since by hypothesis  $0 < \varepsilon < \eta^{C_{s,\kappa}} / P^\kappa$ , we can choose  $j$  large enough depending on  $\kappa$  so that the  $\varepsilon^j \leq \frac{1}{P^s}$ , and thus we now have

$$|\alpha'_1(n_1)| \ll \frac{\eta^{-O(1)}}{P^s} \varepsilon \tag{5.16}$$

for all pairs  $(n_1, n_2)$  in a subset  $\mathcal{N}$  of  $\mathcal{N}_0$  of size  $\gg \eta^{O(1)} P^2$ .

Having obtained good control on  $\alpha'_1$ , we turn to  $\alpha'_2$ . From (5.4) and (5.16) we have

$$\|n_1^s \alpha'_2(n_2)\|_{\mathbb{R}/\mathbb{Z}} \ll \eta^{-O(1)} \varepsilon \tag{5.17}$$

for all  $(n_1, n_2) \in \mathcal{N}$ . Thus we can find a subset  $S'_2$  of  $S_2$  of cardinality  $\gg \eta^{O(1)} P$  such that for every  $n_2 \in S'_2$ , (5.17) holds for  $\gg \eta^{O(1)} P$  choices of  $n_1$  with  $(n_1, n_2) \in \mathcal{N}$ . Applying Lemma 2.3, we thus see that for each  $n_2 \in S'_2$  there is an integer  $1 \leq q \ll \eta^{-O(1)}$  such that

$$\|q \alpha'_2(n_2)\|_{\mathbb{R}/\mathbb{Z}} \ll \eta^{-O(1)} \varepsilon / P^s.$$

By pigeonholing and refining  $S'_2$  as necessary, we may assume without loss of generality that  $q$  is independent of  $n_2$ . We conclude that for  $\gg \eta^{O(1)} P^2$  pairs  $(n_1, n_2) \in \mathcal{N}_0$ , we have

$$\|q \alpha'_i(n_i)\|_{\mathbb{R}/\mathbb{Z}} \ll \eta^{-O(1)} \varepsilon / P^s$$

for  $i = 1, 2$ , and Theorem 5.2 follows.

### 5.3 Nilsequence contagion

In this subsection we bootstrap the monomial contagion lemma (Theorem 5.2) to prove a nilsequence contagion lemma.

Given a filtered nilmanifold  $G/\Gamma$ , we introduce the following relation between polynomial sequences that are equivalent up to smooth and rational components.

**Definition 5.8** Let  $G/\Gamma$  be a filtered nilmanifold,  $I \subset \mathbb{R}$  be an interval with  $|I| \geq 1$ , and let  $\eta \in (0, 1)$ . Let  $g, g' \in \text{Poly}(\mathbb{Z} \rightarrow G)$ . We write  $g \sim_{I,\eta} g'$  if

$$g = \varepsilon g' \gamma,$$

for some  $\varepsilon, \gamma \in \text{Poly}(\mathbb{Z} \rightarrow G)$ , with  $\varepsilon$  being  $(\eta^{-1}, I)$ -smooth and  $\gamma$  being  $\eta^{-1}$ -rational.

**Remark 5.9** Note that the relation  $\sim_{I,\eta}$  depends also on the choice of the filtered nilmanifold  $G/\Gamma$ . However, as the choice of the filtered nilmanifold will always be clear from context, we omit this data from the notation.

Note that while the relation  $\sim_{I,\eta}$  is not quite an equivalence relation, it is an equivalence relation up to polynomial losses in  $\eta$ . It is also scalable with respect to the interval  $I$ . These facts are summarized in the following lemma.

**Lemma 5.10** (Basic properties of the  $\sim_{I,\eta}$  relation) *Let  $B, d, D \geq 0$ . Let  $\eta \in (0, 1/2)$ , and let  $G/\Gamma$  be a filtered nilmanifold of dimension  $D$ , degree  $d$  and complexity  $\leq \eta^{-B}$ . Let  $I \subset \mathbb{R}$  be an interval of length  $\geq 1$ , and let  $g_1, g_2, g_3 \in \text{Poly}(\mathbb{Z} \rightarrow G)$ .*

(i) *We have*

$$\begin{aligned} g_1 &\sim_{I,\eta} g_1, \\ g_1 \sim_{I,\eta} g_2 &\implies g_2 \sim_{I,\eta} O_{B,d,D(1)} g_1, \\ g_1 \sim_{I,\eta} g_2 \quad \text{and} \quad g_2 \sim_{I,\eta} g_3 &\implies g_1 \sim_{I,\eta} O_{B,d,D(1)} g_3. \end{aligned}$$

(ii) *Let  $I_1, I_2, I_3 \subset \mathbb{R}$  be intervals of length at least  $d + 1$  with  $I_1 \subset I_2 \subset I_3$  and  $|I_3| \leq \eta^{-B}|I_2|$ . Then*

$$\begin{aligned} g_1 \sim_{I_2,\eta} g_2 &\implies g_1 \sim_{I_1,\eta} g_2, \\ g_1 \sim_{I_2,\eta} g_2 &\implies g_1 \sim_{I_3,\eta} O_{B,d,D(1)} g_2. \end{aligned}$$

**Proof** We allow implied constants to depend on  $B, d, D$ .

The first claim in part (i) is trivial. The second part follows from the fact that if  $\varepsilon$  is  $(\eta^{-1}, I)$ -smooth and  $\gamma$  is  $\eta^{-1}$ -rational, then by Lemma 2.13 also  $\varepsilon^{-1}$  is  $(\eta^{-O(1)}, I)$ -smooth and  $\gamma^{-1}$  is  $\eta^{-O(1)}$ -rational. The third part follows by noting that if  $g_1 = \varepsilon_1 g_2 \gamma_1$  and  $g_2 = \varepsilon_2 g_3 \gamma_2$  with  $\varepsilon_i$  being  $(\eta^{-1}, I)$ -smooth and  $\gamma_i$  being  $\eta^{-1}$ -rational for  $i = 1, 2$ , then  $g_1 = \varepsilon_1 \varepsilon_2 g_3 \gamma_2 \gamma_1$  and by Lemma 2.13 we know that  $\varepsilon_1 \varepsilon_2$  is  $(\eta^{-O(1)}, I)$ -smooth and  $\gamma_2 \gamma_1$  is  $\eta^{-O(1)}$ -rational.

The first claim in part (ii) is trivial. We proceed to prove the second part. It suffices to show that if  $\varepsilon \in \text{Poly}(\mathbb{Z} \rightarrow G)$  is  $(\eta^{-1}, I_2)$ -smooth, then it is  $(\eta^{-O(1)}, I_3)$ -smooth. This amounts to showing that for all  $n \in I_3$  we have  $d_G(\varepsilon(n), 1_G) \leq \eta^{-O(1)}$  and

$$d_G(\varepsilon(n), \varepsilon(n - 1)) \leq \frac{\eta^{-O(1)}}{|I_3|}. \tag{5.18}$$

Note that if we have (5.18), and if  $n_0 \in I_2$  is chosen arbitrarily, then by repeated application of the triangle inequality, for any  $n \in I_3$ , we have

$$d_G(\varepsilon(n), 1_G) \leq d_G(\varepsilon(n_0), 1_G) + |n - n_0| \frac{\eta^{-O(1)}}{|I_3|} \ll \eta^{-O(1)}$$

by the fact that  $\varepsilon$  is  $(\eta^{-1}, I_2)$ -smooth. Therefore, it suffices to prove (5.18).

Let  $V = \log G$ . Then  $V$  is a vector space of dimension  $D = \dim G$ , so it is isomorphic to  $\mathbb{R}^D$ . Equip  $V$  with a Euclidean norm  $\|\cdot\|$ . Consider the function  $P(n) = \log(\varepsilon(n)\varepsilon(n - 1)^{-1})$ . The sequence  $n \mapsto \varepsilon(n)\varepsilon(n - 1)^{-1}$  is a polynomial map from  $\mathbb{Z}$  to  $G$ , hence by the Baker–Campbell–Hausdorff formula,  $P$  is a polynomial of degree at most  $d$  from  $\mathbb{Z}$  to  $\log G \cong \mathbb{R}^D$ . By hypothesis,  $\|P(n)\| = O(\eta^{-O(1)}/|I_2|)$  for all  $n \in I_2$ , hence by Lagrange interpolation (and the assumption  $|I_2| \geq d + 1$ ) we have  $\|P(n)\| = O(\eta^{-O(1)}/|I_3|)$  for all  $n \in I_3$ , and (5.18) follows.  $\square$

With the above notation, we are ready to state the nilsequence contagion lemma.

**Theorem 5.11** (Nilsequence contagion lemma) *Let  $\kappa \in (0, 1)$  and  $\eta \in (0, 1/2)$ . Let  $G/\Gamma$  be a filtered nilmanifold of dimension  $D$ , degree  $d$  and complexity at most  $1/\eta$ . Let  $C$  be a constant that is sufficiently large depending on  $\kappa, D, d$ . Let  $P \geq \eta^{-C}$  and let  $I$  be an interval with  $|I| \geq \eta^{-C} P^{2+\kappa}$ . Let  $S_1, S_2$  be subsets of  $[P, 2P]$ , and suppose that we have polynomial maps  $g_n, g'_n: \mathbb{Z} \rightarrow G$  for  $n \in S_1, n' \in S_2$  such that*

$$g_n(n'\cdot) \sim \frac{1}{nn'} I, \eta g'_n(n\cdot)$$

for  $\geq \eta P^2$  pairs  $(n, n') \in S_1 \times S_2$ . Then there is a polynomial map  $g_*: \mathbb{Z} \rightarrow G$  such that

$$g_n \sim \frac{1}{n} I, \eta^{O_{\kappa, D, d}(1)} g_*(n\cdot)$$

for  $\gg \eta^{O_{\kappa, D, d}(1)} P$  values of  $n \in S_1$  and

$$g'_n \sim \frac{1}{n'} I, \eta^{O_{\kappa, D, d}(1)} g_*(n'\cdot)$$

for  $\gg \eta^{O_{\kappa, D, d}(1)} P$  values of  $n' \in S_2$ .

It is instructive to see why Theorem 5.11 contains Theorem 5.2 as a special case, so we pause to outline this implication. Let the assumptions be as in Theorem 5.2, and allow implied constants to depend on  $s, \kappa$ . Apply Theorem 5.11 with  $I = [\frac{1}{2}\varepsilon^{-1/s} P^2, \varepsilon^{-1/s} P^2]$  and with  $G/\Gamma = \mathbb{R}/\mathbb{Z}$ , equipped with a filtration of degree  $s$ . Take  $g_n(n') = \alpha_1(n)(n')^s, g'_n(n) = \alpha_2(n')n^s$ . Then for  $(n, n') \in N_0$  we have  $g_n(n'\cdot) \sim \frac{1}{nn'} I, \eta g'_n(n\cdot)$ , since by (5.2) for  $y \in \frac{1}{nn'} I$  the function  $\epsilon(y) := g_n(n'y) - g'_n(ny)$  satisfies

$$\|\epsilon(y) - \epsilon(y - 1)\|_{\mathbb{R}/\mathbb{Z}} \ll \|(\alpha_1(n)(n')^s - \alpha_2(n')n^s)\|_{\mathbb{R}/\mathbb{Z}} y^{s-1} \ll \frac{1}{|\frac{1}{nn'} I|} \ll \varepsilon^{1/s},$$

where we used the triangle inequality estimate  $\|\beta y^s - \beta(y - 1)^s\|_{\mathbb{R}/\mathbb{Z}} \ll \|\beta\|_{\mathbb{R}/\mathbb{Z}} y^{s-1}$  for  $\beta \in \mathbb{R}$ . Theorem 5.11 now tells us that there exists a polynomial  $g_*$  of degree at most  $s$  such that  $g_n \sim \frac{1}{n} I, \eta^{O(1)} g_*(n\cdot)$  for  $n \in S_1^*$ , where  $S_1^* \subset S_1$  has cardinality  $\gg \eta^{O(1)} P$ . Hence, there exists an integer  $1 \leq q \ll \eta^{-O(1)}$  such that the polynomial  $y \mapsto g_n(qy) - g_*(nqy)$  is  $\eta^{-O(1)}$ -smooth on  $\frac{1}{qn} I$ . Now, denoting by  $\alpha_*$  the degree  $s$  coefficient of  $g_*$ , the highest degree coefficient of this polynomial is  $q^s \alpha_1(n) - \alpha_* q^s n^s$ , so we in particular have

$$\|q^s (\alpha_1(n) - \alpha_* n^s)\|_{\mathbb{R}/\mathbb{Z}} \ll \eta^{-O(1)} \left(\frac{|I|}{qn}\right)^{-s} \ll \eta^{-O(1)} \varepsilon/P^s$$

for  $n \in S_1^*$ , as desired. The same argument gives the analogous conclusion for  $\alpha_2$ .

We now prove Theorem 5.11. Let  $\kappa, \eta, G/\Gamma, d, D, C, P, I$  be as in this theorem; we allow implied constants to depend on  $\kappa, D, d$ . The strategy is to inductively apply Theorem 5.2 to certain ‘‘coefficients’’ of the  $g_n, g'_n$ , until all of the coefficients of  $g_n, g'_n$  are related to a single polynomial map  $g_*$ . These coefficients will be indexed by

a pair  $(j, k)$  of natural numbers, and are related to a certain way to express the group  $\text{Poly}(\mathbb{Z} \rightarrow G)$  as a tower of abelian extensions (i.e., to describe  $\text{Poly}(\mathbb{Z} \rightarrow G)$  as a solvable<sup>10</sup> group).

In order to define these coefficients properly, we will need to introduce additional filtration structures on  $G$ . Namely, we define additional filtrations  $G_{[j,k],\bullet} = (G_{[j,k],i})_{i \geq 0}$  for  $j, k \geq 0$  by

$$G_{[j,k],i} := G_{j+i+1_{i < k}}.$$

One can check that these are indeed filtrations. Furthermore we have that  $G_{[j,k],i}$  is monotone decreasing with respect to lexicographical ordering on  $[j, k]$ , thus

$$G_{[j',k'],i} \leq G_{[j,k],i}$$

whenever  $[j', k'] \geq [j, k]$ , in the sense that either  $j' > j$  or  $j' = j$  and  $k' > k$ . Also we have

$$[G_{[j_1,k_1],i_1}, G_{[j_2,k_2],i_2}] \leq G_{[j_1+j_2,k_1+k_2],i_1+i_2}$$

for any  $i_1, i_2, j_1, j_2, k_1, k_2 \geq 0$ . Finally we have the identities

$$G_{\bullet} = G_{[0,0],\bullet} = G_{[0,1],\bullet}$$

and

$$G_{[j,k],\bullet} = G_{[j+1,0],\bullet}$$

when  $j + k > d$ .

If we then define  $\text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]}$  to be the group of polynomial maps  $g$  from  $\mathbb{Z}$  to the filtered group  $(G, G_{[j,k],\bullet})$ , then  $\text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]}$  is monotone decreasing with respect to  $[j, k]$  in the sense that

$$\text{Poly}(\mathbb{Z} \rightarrow G)_{[j',k']} \leq \text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]}$$

whenever  $[j', k'] \geq [j, k]$ , and also we have the filtration property

$$[\text{Poly}(\mathbb{Z} \rightarrow G)_{[j_1,k_1]}, \text{Poly}(\mathbb{Z} \rightarrow G)_{[j_2,k_2]}] \leq \text{Poly}(\mathbb{Z} \rightarrow G)_{[j_1+j_2,k_1+k_2]}$$

for any  $[j_1, k_1], [j_2, k_2]$ ; in particular, the quotient groups  $\text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]} / \text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k+1]}$  are all abelian. Finally we have the identities

$$\text{Poly}(\mathbb{Z} \rightarrow G) = \text{Poly}(\mathbb{Z} \rightarrow G)_{[0,0]} = \text{Poly}(\mathbb{Z} \rightarrow G)_{[0,1]} \tag{5.19}$$

and

$$\text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]} = \text{Poly}(\mathbb{Z} \rightarrow G)_{[j+1,0]} \tag{5.20}$$

<sup>10</sup>In fact  $\text{Poly}(\mathbb{Z} \rightarrow G)$  is actually a nilpotent group, but for this argument the solvable structure is more relevant.

when  $j + k > d$ . We refer to elements of  $\text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]}$  as  $[j, k]$ -polynomial maps; informally, these are polynomial maps in which a certain number of lower order coefficients vanish, leaving only higher order terms.

An element  $g$  of  $\text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]}$  can be extended by [29, Lemma 4.2] to a polynomial map from  $\mathbb{R}$  to  $G_{[j,k],\bullet}$ , thus

$$g(n) = \exp\left(\sum_{\ell=0}^d X_\ell n^\ell\right)$$

for all  $n \in \mathbb{Z}$  and some coefficients  $X_\ell \in \log G_{[j,k],\ell}$ . By the Baker–Campbell–Hausdorff formula and induction, for every  $m \in \{-1, 0, \dots, d\}$  we can obtain representations

$$g(n) = g'_0(g'_1)^n \dots (g'_m)^{n^m} \exp\left(\sum_{\ell=m+1}^d X_{\ell,m} n^\ell\right)$$

for all  $n \in \mathbb{Z}$  and some coefficients  $g'_\ell \in G_{[j,k],\ell}$  and  $X_{\ell,m} \in \log G_{[j,k],\ell}$ . Setting  $m = d$  and relabeling, we conclude that we have the representation

$$g(n) = \tilde{g}_{j+1} \tilde{g}_{j+2}^n \dots \tilde{g}_{j+k}^{n^{k-1}} g_{j+k}^{n^k} \dots g_{j+d}^{n^d}$$

for all  $n \in \mathbb{Z}$ , where  $\tilde{g}_i \in G_i$  for  $j + 1 \leq i \leq j + k$  and  $g_i \in G_i$  for  $j + k \leq i \leq j + d$ , in particular  $g$  is equal to the monomial  $n \mapsto g_{j+k}^{n^k}$  modulo  $\text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k+1]}$ . We define the  $[j, k]$ -frequency map

$$\log_{[j,k]} : \text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]} \rightarrow \log G_{j+k} / \log G_{j+k+1}$$

by the formula

$$\log_{[j,k]} g := \log g_{j+k} \text{ mod } \log G_{j+k+1}.$$

One can verify that this is a surjective group homomorphism. One can think of  $\log_{[j,k]} g$  as the “lowest order” or “most important” coefficient of the  $[j, k]$ -polynomial map  $g$ .

**Example 5.12** The following table for  $d = 2$  shows the form of  $n \mapsto g(n)$  for  $g \in \text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]}$ , as well as the value of  $\log_{[j,k]} g$ , for various choices of  $[j, k]$ , where for each  $i$ ,  $g_i, \tilde{g}_i$  denote arbitrary elements of  $G_i$  (see Table 1).

Observe that we have a subnormal series

$$\begin{aligned} &\text{Poly}(\mathbb{Z} \rightarrow G)_{[0,0]} = \text{Poly}(\mathbb{Z} \rightarrow G)_{[0,1]} \geq \text{Poly}(\mathbb{Z} \rightarrow G)_{[0,2]} \geq \text{Poly}(\mathbb{Z} \rightarrow G)_{[0,3]} \\ &= \text{Poly}(\mathbb{Z} \rightarrow G)_{[1,0]} \geq \text{Poly}(\mathbb{Z} \rightarrow G)_{[1,1]} \geq \text{Poly}(\mathbb{Z} \rightarrow G)_{[1,2]} \\ &= \text{Poly}(\mathbb{Z} \rightarrow G)_{[2,0]} \geq \text{Poly}(\mathbb{Z} \rightarrow G)_{[2,1]} \\ &= \text{Poly}(\mathbb{Z} \rightarrow G)_{[3,0]} \end{aligned}$$

**Table 1** Elements  $g$  of  $\text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]}$

$[j, k]$	$g(n)$	$\log_{[j,k]} g$
[0, 0]	$g_0 g_1^n g_2^{n^2}$	0
[0, 1]	$\tilde{g}_1 g_1^n g_2^{n^2}$	$\log g_1 \pmod{\log G_2}$
[0, 2]	$\tilde{g}_1 \tilde{g}_2^n g_2^{n^2}$	$\log g_2$
[0, 3]	$\tilde{g}_1 \tilde{g}_2^n$	0
[1, 0]	$g_1 g_2^n$	$\log g_1 \pmod{\log G_2}$
[1, 1]	$\tilde{g}_2 g_2^n$	$\log g_2$
[1, 2]	$\tilde{g}_2$	0
[2, 0]	$g_2$	$\log g_2$
[2, 1]	1	0
[3, 0]	1	0

that starts at  $\text{Poly}(\mathbb{Z} \rightarrow G) = \text{Poly}(\mathbb{Z} \rightarrow G)_{[0,0]}$  and ends at the trivial group  $\{1\} = \text{Poly}(\mathbb{Z} \rightarrow G)_{[3,0]}$ ; the maps  $\log_{[j,k]}$  are essentially the quotient maps between adjacent groups in this subnormal series. In particular, if  $g \in \text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]}$  and  $[j', k']$  is the successor to  $[j, k]$  in the lexicographical ordering, then  $g$  lies in the next group  $\text{Poly}(\mathbb{Z} \rightarrow G)_{[j',k']}$  in the subnormal series if and only if the coefficient  $\log_{[j,k]} g$  vanishes.

**Example 5.13** In the case that  $G$  is the Heisenberg group generated by two generators  $e_1, e_2$  with central commutator  $[e_1, e_2]$ , then we have  $g(n)\Gamma = e_1^{\{P_1(n)\}} e_2^{\{P_2(n)\}} \times [e_1, e_2]^{\{P_{12}(n)\}}\Gamma$  for some bracket polynomials  $P_1(n), P_2(n), P_{12}(n)$ ; for instance, if

$$g(n) = e_1^{\alpha n + \kappa} e_2^{\beta n + \sigma} [e_1, e_2]^{\gamma n^2 + \delta n + \epsilon}$$

for various real numbers  $\alpha, \beta, \gamma, \delta, \epsilon, \sigma, \kappa$  then we can take  $P_1(n) := \alpha n + \kappa, P_2(n) := \beta n + \sigma$ , and

$$P_{12}(n) := \lfloor \alpha n + \kappa \rfloor (\beta n + \sigma) + \gamma n^2 + \delta n + \epsilon.$$

The above table then specializes as follows (see Table 2).

On each row of this table,  $\log_{[j,k]} g$  informally captures the ‘‘leading coefficients’’ of the triple  $(P_1(n), P_2(n), P_{12}(n))$  of bracket polynomials, and the kernel of this map produces the slightly smaller class of bracket polynomials in the row below. If one specializes to the case  $\alpha = \beta = \kappa = \sigma = 0$  one recovers the classical situation of a quadratic polynomial  $\gamma n^2 + \delta n + \epsilon$  and its coefficients  $\gamma, \delta, \epsilon$ , with the degree  $k$  coefficient being extracted by the map  $\log_{[d-k,k]}$ .

For any interval  $I$  of length  $\geq 1$ , any  $\eta \in (0, 1/2)$  and any  $[j, k]$ , we define a relation

$$g \sim_{I,[j,k],\eta} g'$$

**Table 2** Specialization of Table 1 to  $g(n)\Gamma = e_1^{\{P_1(n)\}} e_2^{\{P_2(n)\}} [e_1, e_2]^{\{P_{12}(n)\}} \Gamma$ 

$[j, k]$	$P_1(n)$	$P_2(n)$	$P_{12}(n)$	$\log_{[j,k]} g$
$[0, 0]$	$\alpha n + \kappa$	$\beta n + \sigma$	$\lfloor \alpha n + \kappa \rfloor (\beta n + \sigma) + \gamma n^2 + \delta n + \epsilon$	0
$[0, 1]$	$\alpha n + \kappa$	$\beta n + \sigma$	$\lfloor \alpha n + \kappa \rfloor (\beta n + \sigma) + \gamma n^2 + \delta n + \epsilon$	$(\alpha, \beta)$
$[0, 2]$	$\kappa$	$\sigma$	$\lfloor \kappa \rfloor \sigma + \gamma n^2 + \delta n + \epsilon$	$\gamma$
$[0, 3]$	$\kappa$	$\sigma$	$\lfloor \kappa \rfloor \sigma + \delta n + \epsilon$	0
$[1, 0]$	$\kappa$	$\sigma$	$\lfloor \kappa \rfloor \sigma + \delta n + \epsilon$	$(\kappa, \sigma)$
$[1, 1]$	0	0	$\delta n + \epsilon$	$\delta$
$[1, 2]$	0	0	$\epsilon$	0
$[2, 0]$	0	0	$\epsilon$	$\epsilon$
$[2, 1]$	0	0	0	0
$[3, 0]$	0	0	0	0

on polynomials  $g, g' : \mathbb{Z} \rightarrow G$  if one has a factorisation

$$g = \varepsilon g' \gamma g_{[j,k]}$$

where  $\varepsilon$  is  $\eta^{-1}$ -smooth on  $I$ ,  $\gamma$  is  $\eta^{-1}$ -rational, and  $g_{[j,k]} \in \text{Poly}(\mathbb{Z} \rightarrow G)_{[j,k]}$ , that is to say  $g$  is equal to  $\varepsilon g' \gamma$  modulo  $[j, k]$ -polynomial maps. From Lemma 2.13, we see that  $\sim_{I,[j,k],\eta}$  is an equivalence relation (up to polynomial losses in  $\eta$ ). Note that the relation  $\sim_{I,[d+1,0],\eta}$  corresponds to the relation  $\sim_{I,\eta}$  defined in Definition 5.8.

For  $i \geq 0$ , define the  $i$ -torus  $T_i$  to be  $T_i := \log G_i / \log(G_{i+1}\Gamma_i)$ ;  $T_1$  is the horizontal torus when  $G_2 = [G_1, G_1]$  and  $T_d$  is the vertical torus, while  $T_i$  is trivial for  $i = 0$  or  $i > d$ . Given an element  $\xi_i$  of  $\log G_i / \log G_{i+1}$ , we define  $\|\xi_i\|_{T_i}$  to be the distance to the origin of the projection of  $\xi_i$  to  $T_i$ .

The relation  $\sim_{I,[j,k+1],\eta}$  implies a relation on the  $[j, k]$ -frequencies of two  $[j, k]$ -polynomial maps  $g, g'$ , even if the relation is conjugated by an arbitrary additional polynomial map  $g_*$ .

**Lemma 5.14** *Suppose that  $g_* : \mathbb{Z} \rightarrow G$  is a polynomial map and  $g, g' : \mathbb{Z} \rightarrow G$  are  $[j, k]$ -polynomial maps, and suppose that*

$$g_* g \sim_{I,[j,k+1],\eta} g_* g'$$

for some interval  $I$  and some  $\eta \in (0, 1/2)$ . Then we have

$$\|q(\log_{[j,k]}(g) - \log_{[j,k]}(g'))\|_{T_{j+k}} \ll \eta^{-O(1)} |I|^{-j-k}$$

for some integer  $1 \leq q \ll \eta^{-O(1)}$ .

**Proof** We can assume that  $|I| \geq C\eta^{-C}$  for some large constant  $C$  since the claim is trivial otherwise. We can also assume  $k + j > 0$  since the conclusion is trivial for  $k = j = 0$ .

By hypothesis, we have

$$g_* g = \varepsilon g_* g' \gamma \text{ mod Poly}(\mathbb{Z} \rightarrow G)_{[j,k+1]}$$

for some  $(\eta^{-1}, I)$ -smooth  $\varepsilon \in \text{Poly}(\mathbb{Z} \rightarrow G)$  and  $\eta^{-1}$ -rational  $\gamma \in \text{Poly}(\mathbb{Z} \rightarrow G)$ . Thus

$$g = [g_*^{-1}, \varepsilon] \varepsilon g' \gamma \text{ mod Poly}(\mathbb{Z} \rightarrow G)_{[j, k+1]}, \tag{5.21}$$

where  $[a, b] := aba^{-1}b^{-1}$ . We now claim inductively that for all  $[0, 1] \leq [j', k'] \leq [j, k]$ , one has

$$\varepsilon, \gamma \in \text{Poly}(\mathbb{Z} \rightarrow G)_{[j', k']}.$$

This is already established for  $[j', k'] = [0, 1]$  by (5.19). Suppose it is true for some  $[0, 1] \leq [j', k'] < [j, k]$  (so in particular  $j + k \geq 1$ ). Then we have

$$g = \varepsilon g' \gamma \text{ mod Poly}(\mathbb{Z} \rightarrow G)_{[j', k'+1]}$$

and then on applying  $\log_{[j', k']}$  (which annihilates  $g, g'$ ) we have

$$0 = \log_{[j', k]}(\varepsilon) + \log_{[j', k]}(\gamma).$$

Since  $\varepsilon$  is  $\eta^{-O(1)}$ -smooth on  $I$ ,  $\log \eta$  takes values in a ball of radius  $O(\eta^{-O(1)})$  on  $I$ , so by the Lagrange interpolation theorem, the degree  $d$  coefficient of  $\log \varepsilon$  has size  $O(\eta^{-O(1)}|I|^{-d})$  for all  $d$ . In particular,  $\log_{[j', k]}(\varepsilon)$  is of size  $O(\eta^{-O(1)}|I|^{-j'-k'})$ . Since  $\gamma$  is  $\eta^{-O(1)}$ -rational,  $\log_{[j', k]}(\gamma)$  is  $\eta^{-O(1)}$ -rational also. Recalling that  $|I| \geq C\eta^{-C}$ , the only way these two statements are consistent are if  $\log_{[j', k]}(\varepsilon) = \log_{[j', k]}(\gamma) = 0$ . This implies that  $\varepsilon, \gamma \in \text{Poly}(\mathbb{Z} \rightarrow G)_{[j', k'+1]}$ . Inducting (using (5.20)) we obtain the claim. In particular,  $\varepsilon, \gamma \in \text{Poly}(\mathbb{Z} \rightarrow G)_{[j, k]}$ . Applying  $\log_{[j, k]}$  to (5.21) we now get

$$\log_{[j, k]} g = \log_{[j, k]} \varepsilon + \log_{[j, k]} g' + \log_{[j, k]} \gamma.$$

Since  $\log_{[j, k]}(\varepsilon) = O(\eta^{-O(1)}|I|^{-j-k})$  and  $\log_{[j, k]}(\gamma)$  is  $\eta^{-O(1)}$ -rational, we obtain the claim. □

Now we can prove Theorem 5.11.

**Proof of Theorem 5.11** Initialize  $\mathcal{N} = S_1, \mathcal{N}' = S_2$ . Then we have

$$g_n(n') \sim \frac{1}{n\eta} I_{, \eta^{O(1)}} g'_n(n') \tag{5.22}$$

for  $\gg \eta^{O(1)} P^2$  pairs  $(n, n') \in \mathcal{N} \times \mathcal{N}'$  (in fact, the  $\eta^{O(1)}$  term above could be replaced with  $\eta$ ).

We claim inductively that for  $[0, 1] \leq [j, k] \leq [d + 1, 0]$ , we can refine  $\mathcal{N}, \mathcal{N}'$  by a factor of  $\eta^{O(1)}$  such that the relation (5.22) still holds for  $\gg \eta^{O(1)} P^2$  pairs, and there is a polynomial map  $g_* : \mathbb{Z} \rightarrow G$  such that

$$g_n \sim \frac{1}{n} I_{, [j, k], \eta^{O(1)}} g_*(n),$$

$$g'_n \sim \frac{1}{n'} I_{, [j, k], \eta^{O(1)}} g_*(n')$$

for all  $n \in \mathcal{N}, n' \in \mathcal{N}'$ ; setting  $[j, k] = [d + 1, 0]$  will give the claim.

For  $[j, k] = [0, 1]$  this follows by setting  $g_*$  to be the identity and using (5.19). By (5.20) and induction,<sup>11</sup> it thus suffices to show that if the claim holds for some  $[0, 1] \leq [j, k] < [d + 1, 0]$  then it also holds for  $[j, k + 1]$ . By hypothesis we have

$$g_n \sim \frac{1}{n} I_{[j, k+1], \eta}^{O(1)} g_*(n \cdot) \tilde{g}_n$$

$$g'_{n'} \sim \frac{1}{n'} I_{[j, k+1], \eta}^{O(1)} g_*(n' \cdot) \tilde{g}'_{n'}$$

for some  $[j, k]$ -polynomials  $\tilde{g}_n, \tilde{g}'_{n'}$ . By (5.22) we then have

$$g_*(nn' \cdot) \tilde{g}_n(n' \cdot) \sim \frac{1}{nn'} I_{[j, k+1], \eta}^{O(1)} g_*(nn' \cdot) \tilde{g}'_{n'}(n \cdot)$$

for  $\gg \eta^{O(1)} P^2$  pairs  $n \in \mathcal{N}, n' \in \mathcal{N}'$ . Applying Lemma 5.14, we conclude that

$$\|q(\log_{[j, k]}(\tilde{g}_n(n' \cdot)) - \log_{[j, k]}(\tilde{g}'_{n'}(n \cdot)))\|_{T_{j+k}} \ll \eta^{-O(1)} (|I|/P^2)^{-j-k}$$

for some integer  $1 \leq q \ll \eta^{-O(1)}$ , which we can pigeonhole to be independent of  $n, n'$ . From Taylor expansion we have

$$\log_{[j, k]} g(n \cdot) = n^{j+k} \log_{[j, k]} g$$

and hence

$$\|q((n')^{j+k} \log_{[j, k]}(\tilde{g}_n) - n^{j+k} \log_{[j, k]}(\tilde{g}'_{n'}))\|_{T_{j+k}} \ll \eta^{-O(1)} (|I|/P^2)^{-j-k}.$$

For  $k + j > 0$ , we may apply Theorem 5.2 once for each coordinate of the torus  $T_{j+k}$ , and modify  $q$  as necessary, to refine  $\mathcal{N}, \mathcal{N}'$  so that

$$\|q(\log_{[j, k]}(\tilde{g}_n) - \beta n^{j+k})\|_{T_{j+k}} \ll \eta^{-O(1)} (|I|/P)^{-j-k}$$

and

$$\|q(\log_{[j, k]}(\tilde{g}'_{n'}) - \beta (n')^{j+k})\|_{T_{j+k}} \ll \eta^{-O(1)} (|I|/P)^{-j-k}$$

for all  $n \in \mathcal{N}, n' \in \mathcal{N}'$  and some  $\beta \in \log G_{j+k} / \log G_{j+k+1}$ , while still keeping (5.22) for  $\gg \eta^{O(1)} P^2$  pairs. In the  $k = 0$  case we can trivially achieve this just by setting  $\beta = 0$ . From this we may factor

$$g_n \sim \frac{1}{n} I_{[j, k+1], \eta}^{O(1)} g_*(n \cdot) g_{[j, k]}^{(n \cdot)^{j+k}}$$

and

$$g'_{n'} \sim \frac{1}{n'} I_{[j, k+1], \eta}^{O(1)} g_*(n' \cdot) g_{[j, k]}^{(n' \cdot)^{j+k}},$$

where  $g_{[j, k]}$  is any element of  $G_{j+k}$  with  $\log_{[j, k]} g_{[j, k]} = \beta$ . This allows us to close the induction, replacing  $g_*$  with  $g_* g_{[j, k]}^{(\cdot)^{j+k}}$ . This completes the proof of Theorem 5.11. □

---

<sup>11</sup>Note that the induction proceeds in a “staircase” fashion: from case  $[0, 1]$  we can eventually get to case  $[0, d + 1]$ , which by (5.20) is the same as case  $[1, 0]$ , and then repeating the same procedure we eventually get to case  $[1, d + 1]$ , which by (5.20) is the same as case  $[2, 0]$ , and so on.

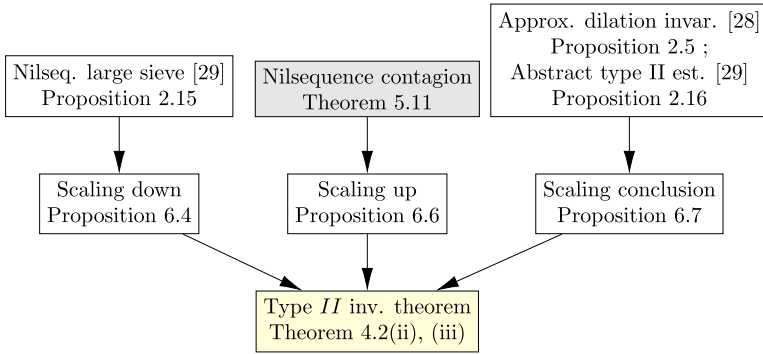


Fig. 4 Proof of the type II inverse theorem, Theorem 4.2(ii), (iii). For the proof of Theorem 5.11, see Fig. 3

### 6 Proof of type II case

In this section we establish the type II cases (ii), (iii) of Theorem 4.2, using a version of Walsh’s contagion argument [42].

Recall the definition of the relation  $\sim_{I,\eta}$  from Definition 5.8. The following two definitions will also be important for expressing the main argument in this section.

**Definition 6.1** (Configurations of points and polynomial sequences) Let  $I$  be an interval and  $\sigma > 0$ . We say that a collection  $\mathcal{J} \subset I \times \text{Poly}(\mathbb{Z} \rightarrow G)$  is a  $(\sigma, H)$ -configuration in  $I$ , if  $\mathcal{J}$  is of the form  $\{(x, g_x) : x \in \mathcal{X}\}$ , where  $|\mathcal{X}| \geq \sigma|I|/H$ ,  $g_x \in \text{Poly}(\mathbb{Z} \rightarrow G)$ , and the points of  $\mathcal{X} \subset I$  are  $H$ -separated (that is,  $x_1, x_2 \in \mathcal{X}$  and  $x_1 \neq x_2$  imply  $|x_1 - x_2| \geq H$ ).

Informally, a  $(\sigma, H)$ -configuration abstracts (and discretizes) the notion of having correlations with nilsequences  $F(g_x(n)\Gamma)$  on a large family of intervals  $(x, x + H]$  in  $I$ ; compare with the definition of an  $(X, H)$ -family of intervals in [28, Sect. 2] for linear phases.

**Definition 6.2** (Upwards scaling) Let  $C \geq 1$ ,  $\eta, \sigma \in (0, 1/2)$ , and  $A, H, Y \geq 1$ . Let  $\mathcal{J}_0$  be a  $(\sigma, H)$ -configuration in  $[Y, CY]$ , and let  $\mathcal{J}_1$  be a  $(\sigma, AH)$ -configuration in  $[AY, CAY]$ .

We write

$$\mathcal{J}_0 \longrightarrow_{\eta, C} \mathcal{J}_1$$

if for each  $(y, g_y) \in \mathcal{J}_1$  we have

$$\sum_{A < a \leq CA} \sum_{(x, g_x) \in \mathcal{J}_0} I_{a, C, AH, \sigma}(y, x) \geq \eta A$$

where  $I_{a, C, AH, \sigma}(y, x)$  is the indicator function

$$I_{a, C, AH, \sigma}(y, x) := 1_{|y-ax| \leq CAH} 1_{g_y(a) \sim \frac{1}{a}(y, y+AH), \sigma} g_x. \tag{6.1}$$

Note that the relation  $\mathcal{J}_0 \xrightarrow{\eta, C} \mathcal{J}_1$  gets weaker if  $\eta$  is decreased or  $C$  is increased. Informally, the relation  $\mathcal{J}_0 \xrightarrow{\eta, C} \mathcal{J}_1$  indicates that the nilsequences associated to the intervals  $(y, y + AH]$  in  $\mathcal{J}_1$  are essentially rescaled versions of the nilsequences associated to the intervals  $(x, x + H]$  in  $\mathcal{J}_0$ . There are multiple ways to connect these intervals by rescaling; the idea of the contagion argument is to use many existing rescaling relationships at one pair of scales to create additional scaling relationships at a larger pair of scales.

The next lemma shows that the notion of upwards scaling is “transitive”.

**Lemma 6.3** (Transitivity of upwards scaling) *Let  $X \geq H \geq A_2 \geq A_1 \geq 3$ . Let  $\sigma \in (0, \frac{1}{2\log(A_1A_2)})$  and  $C \in [1, \sigma^{-1}]$ . Let  $G/\Gamma$  be a filtered nilmanifold of degree and dimension  $O(1)$  and complexity  $\leq \sigma^{-1}$ . Let  $\mathcal{J}_0$  be a  $(\sigma, H)$ -configuration in  $[Y, CY]$ , let  $\mathcal{J}_1$  be a  $(\sigma, A_1H)$ -configuration in  $[A_1Y, CA_1Y]$ , and let  $\mathcal{J}_2$  be a  $(\sigma, A_1A_2H)$ -configuration in  $[A_1A_2Y, CA_1A_2Y]$ . Suppose that*

$$\mathcal{J}_0 \xrightarrow{\sigma, C} \mathcal{J}_1 \quad \text{and} \quad \mathcal{J}_1 \xrightarrow{\sigma, C} \mathcal{J}_2.$$

Then we have

$$\mathcal{J}_0 \xrightarrow{\sigma^{O(1)}, 2C^2} \mathcal{J}_2. \tag{6.2}$$

**Proof** From Definition 6.2, we see that for every  $(z, g_z) \in \mathcal{J}_2$  we have

$$\sum_{A_2 < a_2 \leq CA_2} \sum_{(y, g_y) \in \mathcal{J}_1} I_{a_2, C, A_1A_2H, \sigma}(z, y) \geq \sigma A_2$$

and for every  $(y, g_y) \in \mathcal{J}_1$  we have

$$\sum_{A_1 < a_1 \leq CA_1} \sum_{(x, g_x) \in \mathcal{J}_0} I_{a_1, C, A_1H, \sigma}(y, x) \geq \sigma A_1.$$

Combining the previous two estimates, for every  $(z, g_z) \in \mathcal{J}_2$  we obtain

$$\sum_{\substack{A_1 < a_1 \leq CA_1 \\ A_2 < a_2 \leq CA_2}} \sum_{(y, g_y) \in \mathcal{J}_1} I_{a_2, C, A_1A_2H, \sigma}(z, y) \sum_{(x, g_x) \in \mathcal{J}_0} I_{a_1, C, A_1H, \sigma}(y, x) \geq \sigma^2 A_1 A_2. \tag{6.3}$$

Note that for every non-zero term in (6.3) we have  $|z - a_2y| \leq CA_1A_2H$  and  $|y - a_1x| \leq CA_1H$ , so we must have  $|z - a_1a_2x| \leq 2C^2A_1A_2H$ . Applying Lemma 5.10(ii), this implies that for every non-zero term in (6.3) we have

$$g_z(a_1a_2 \cdot) \sim \frac{1}{a_1a_2} (z, z + A_1A_2H)_{1, \sigma} g_y(a_1 \cdot) \quad \text{and} \quad g_y(a_1 \cdot) \sim \frac{1}{a_1a_2} (z, z + A_1A_2H)_{1, \sigma^{O(1)}} g_x \cdot,$$

and by the transitivity of the relation  $\sim_{I, \eta}$  (Lemma 5.10(i)), these give

$$g_z(a_1a_2 \cdot) \sim \frac{1}{a_1a_2} (z, z + A_1A_2H)_{1, \sigma^{O(1)}} g_x \cdot$$

In other words,

$$I_{a_1 a_2, 2C^2, A_1 A_2 H, \sigma^{o(1)}}(z, x) = 1.$$

Hence, we obtain

$$\sum_{\substack{A_1 < a_1 \leq C A_1 \\ A_2 < a_2 \leq C A_2}} \sum_{(x, g_x) \in \mathcal{J}_0} I_{a_1 a_2, 2C^2, A_1 A_2 H, \sigma^{o(1)}}(z, x) \sum_{\substack{(y, g_y) \in \mathcal{J}_1 \\ |y - a_1 x| \leq C A_1 H}} 1 \geq \sigma^2 A_1 A_2.$$

Since the sum over  $y$  is bounded by  $\ll C$ , we may make the substitution  $a = a_1 a_2$  and conclude that for every  $(z, g_z) \in \mathcal{J}_2$  we have

$$\sum_{A_1 A_2 < a \leq C^2 A_1 A_2} \sum_{(x, g_x) \in \mathcal{J}_0} d_2(a) I_{a, 2C^2, A_1 A_2 H, \sigma^{o(1)}}(z, x) \gg \frac{\sigma^2}{C} A_1 A_2. \tag{6.4}$$

Applying to (6.4) the Cauchy–Schwarz inequality and the standard divisor sum bound

$$\sum_{a \leq X} d_2(a)^2 \ll X \log^3 X,$$

for all  $(z, g_z) \in \mathcal{J}_2$  we obtain

$$\begin{aligned} \sum_{A_1 A_2 < a \leq C^2 A_1 A_2} \left( \sum_{(x, g_x) \in \mathcal{J}_0} I_{a, 2C^2, A_1 A_2 H, \sigma^{o(1)}}(z, x) \right)^2 &\gg \frac{\sigma^4}{C^4} \frac{A_1 A_2}{\log^3(C^2 A_1 A_2)} \\ &\gg \frac{\sigma^8}{C^4} A_1 A_2. \end{aligned}$$

Since the sum inside the square has at most  $O(C^2)$  non-zero terms, we conclude that

$$\sum_{A_1 A_2 < a \leq C^2 A_1 A_2} \sum_{(x, g_x) \in \mathcal{J}_0} I_{a_1 a_2, 2C^2, A_1 A_2 H, \sigma^{o(1)}}(z, x) \gg \frac{\sigma^8}{C^6} A_1 A_2.$$

The claim (6.2) now follows. □

The following proposition importantly tells us that either Theorem 4.2 holds, or the map  $x \mapsto g_x$  can be extended to a smaller scale.

**Proposition 6.4 (Scaling down)** *Let  $d, D \geq 1, \varepsilon > 0$ , and let  $X \geq H \geq X^\varepsilon \geq 3$  and  $\delta \in (0, \frac{1}{\log X})$ . Let  $G/\Gamma$  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  $f$  is a  $(\delta, A_{11}^-, A_{11}^+)$  type II sum for some  $A_{11}^+ \geq A_{11}^- \geq X^\varepsilon$ . Suppose that*

$$\left| \sum_{x < n \leq x+H} f(n) F(g_x(n)\Gamma) \right|^* \geq \delta H \tag{6.5}$$

for all  $x$  in a subset  $E$  of  $[X, 2X]$  of measure at least  $\delta X$  where, for each  $x \in E$ ,  $g_x : \mathbb{Z} \rightarrow G$  is a polynomial map. Then at least one of the following holds.

(i) We have

$$H \ll_{d,D,\varepsilon} \delta^{-O_{d,D,\varepsilon}(1)} A_{II}^+.$$

(ii) There exists a non-trivial horizontal character  $\eta : G \rightarrow \mathbb{R}$  having Lipschitz norm  $O_{d,D,\varepsilon}(\delta^{-O_{d,D,\varepsilon}(1)})$  such that

$$\|\eta \circ g_x\|_{C^\infty(x,x+H)} \ll_{d,D,\varepsilon} \delta^{-O_{d,D,\varepsilon}(1)}$$

for all  $x$  in a subset of  $E$  of measure  $\gg_{d,D,\varepsilon} \delta^{O_{d,D,\varepsilon}(1)} X$ .

(iii) There exists  $A \in [A_{II}^-, A_{II}^+]$ , a  $(\delta^{O_{d,D,\varepsilon}(1)}, H)$ -configuration  $\mathcal{J}_0$  in  $[X, 2X]$  and a  $(\delta^{O_{d,D,\varepsilon}(1)}, H/A)$ -configuration  $\mathcal{J}_{-1}$  in  $[X/(2A), 2X/A]$ , such that  $\mathcal{J}_0 = \{(x, g_x) : x \in E'\}$  for some set  $E' \subset E$  and

$$\mathcal{J}_{-1} \xrightarrow{\delta^{O_{d,D,\varepsilon}(1)}, O_{d,D,\varepsilon}(1)} \mathcal{J}_0. \tag{6.6}$$

**Remark 6.5** Note that in the case of Theorem 4.2(iii) (where  $F(g_x(n)) = e(P_x(n))$  for some polynomial  $P_x : \mathbb{Z} \rightarrow \mathbb{R}$ ), if conclusion (ii) of Proposition 6.4 holds, then  $\eta(u) = qu$  for some integer  $1 \leq |q| \ll \delta^{-O_{d,\varepsilon}(1)}$ , so

$$\|qP_x \pmod 1\|_{C^\infty(x,x+H)} \ll_{d,\varepsilon} \delta^{-O_{d,\varepsilon}(1)}.$$

If  $\alpha_{x,j}$  denotes the degree  $j$  coefficient of  $P_x(x + \cdot)$ , this implies that

$$\|q\alpha_{x,j}\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O_{d,\varepsilon}(1)} / H^j$$

for all  $1 \leq j \leq d$ . This in turn implies that

$$\|e(P_x(\cdot))\|_{\text{TV}((x,x+H) \cap \mathbb{Z}; q)} \ll_{d,\varepsilon} \delta^{-O_{d,\varepsilon}(1)}.$$

Hence, in the proof of Theorem 4.2(ii)–(iii) we may assume that conclusion (iii) above holds.

**Proof** We let all implied constants to depend on  $d, D, \varepsilon$ . As in Remark 4.3, we may assume that  $x \mapsto g_x$  is constant on intervals of the form  $[m, m + 1)$  with  $m \in \mathbb{N}$ .

We induct on the dimension  $D$  of  $G/\Gamma$ . In view of Proposition 2.11 and Lemma 2.18, we may assume that  $F$  oscillates with a central frequency  $\xi : Z(G) \rightarrow \mathbb{R}$  of Lipschitz norm at most  $\delta^{-O(1)}$ . If the center  $Z(G)$  has dimension larger than 1, or  $\xi$  vanishes, then the conclusion follows from induction hypothesis applied to  $G/\ker \xi$  (via Lemma 2.10). Henceforth we assume that  $G$  has one-dimensional center and that  $\xi$  is non-zero.

Let  $x \in E$ . By (6.5), Lemma 2.2(i) and the assumption that  $f = \alpha * \beta$  is a  $(\delta, A_{II}^-, A_{II}^+)$  type II sum, we have

$$\sum_{A_{II}^- \leq a \leq A_{II}^+} |\alpha(a)| \left| \sum_{x/a < b \leq (x+H)/a} \beta(b) F(g_x(ab)\Gamma) \right|^* \geq \delta H.$$

By the pigeonhole principle and the assumption  $\delta \leq 1/\log X$ , we conclude that there is some  $A \in [A_{II}^-, A_{II}^+]$  such that

$$\sum_{A < a \leq 2A} |\alpha(a)| \left| \sum_{x/a < b \leq x/a + H/A} \beta(b) F(g_x(ab)\Gamma) \right|^* \gg \delta^2 H. \tag{6.7}$$

Now, let  $\mathcal{B}$  be the set of  $x \in [X, 2X]$  for which we have

$$\sum_{A < a \leq 2A} |\alpha(a)| 1_{|\alpha(a)| \geq \delta^{-10}} \left| \sum_{x/a < b \leq x/a + H/A} \beta(b) F(g_x(ab)\Gamma) \right|^* \geq \delta^3 H.$$

Since  $F$  is Lipschitz of norm  $\leq \delta^{-1}$ ,  $|F|$  is pointwise bounded by  $\delta^{-1}$ . Then, by Markov’s inequality, the Cauchy–Schwarz inequality, and the assumed  $L^2$  bounds on  $\alpha$  and  $\beta$ , we have

$$\begin{aligned} \delta^3 H \text{meas}(\mathcal{B}) &\leq \delta^{-1} \int_X^{2X} \sum_{A < a \leq 2A} |\alpha(a)| 1_{|\alpha(a)| \geq \delta^{-10}} \sum_{x/a < b \leq x/a + H/A} |\beta(b)| dx \\ &\ll \delta^{-1} H \sum_{A < a \leq 2A} |\alpha(a)| 1_{|\alpha(a)| \geq \delta^{-10}} \sum_{b \leq (2X+H)/A} |\beta(b)| \\ &\ll \delta^9 H \sum_{A < a \leq 2A} |\alpha(a)|^2 \sum_{b \leq (2X+H)/A} |\beta(b)| \\ &\ll \delta^{9-3/2} H X, \end{aligned}$$

so  $\text{meas}(\mathcal{B}) \ll \delta^{9-3/2-3} X \ll \delta^4 X$ .

From now on, let  $x \in E^* := E \setminus \mathcal{B}$ ; the set  $E^*$  has measure  $\gg \delta X$ . From (6.7) we conclude that for  $x \in E^*$  we have

$$\sum_{A < a \leq 2A} |\alpha(a)| 1_{|\alpha(a)| \leq \delta^{-10}} \left| \sum_{x/a < b \leq x/a + H/A} \beta(b) F(g_x(ab)\Gamma) \right|^* \gg \delta^2 H,$$

so

$$\sum_{A < a \leq 2A} \left| \sum_{x/a < b \leq x/a + H/A} \beta(b) F(g_x(ab)\Gamma) \right|^* \gg \delta^{12} H. \tag{6.8}$$

Let

$$S(x, a) := \left| \sum_{x/a < b \leq x/a + H/A} \beta(b) F(g_x(ab)\Gamma) \right|^*,$$

and let  $\mathcal{B}^*$  be the set of  $x \in E^*$  for which

$$\sum_{A < a \leq 2A} S(x, a) 1_{S(x,a) \geq \delta^{-20} \frac{H}{A}} \geq \delta^{13} H.$$

Then, by the Cauchy–Schwarz inequality and the pointwise bound on  $|F|$ , we have

$$\begin{aligned} \delta^{13} H \text{meas}(\mathcal{B}^*) &\leq \delta^{20} \frac{A}{H} \int_X \sum_{A < a \leq 2A} S(x, a)^2 dx \\ &\ll \delta^{18} \int_X \sum_{A < a \leq 2A} \sum_{x/a < b \leq x/a + H/A} |\beta(b)|^2 dx. \end{aligned}$$

Exchanging the order of integration and summation and using the assumed  $L^2$  bound on  $\beta$ , we see that this is  $\ll \delta^{17} H X$ . Hence  $\text{meas}(\mathcal{B}^*) \ll \delta^4 X$ .

From now on, let  $x \in E^{**} := E^* \setminus \mathcal{B}^*$ ; the set  $E^{**}$  has measure  $\gg \delta X$ . From (6.8) we conclude that for  $x \in E^{**}$  we have

$$\sum_{A < a \leq 2A} S(x, a) 1_{S(x,a) \leq \delta^{-20} \frac{H}{A}} \gg \delta^{12} H.$$

By Markov’s inequality, we deduce that there is a set  $\mathcal{S} \subset ((X, 2X] \times (A, 2A]) \cap \mathbb{N}^2$  with  $|\mathcal{S}| \gg \delta^{33} AX$  such that for all  $(x, a) \in \mathcal{S}$  we have

$$S(x, a) = \left| \sum_{x/a < b \leq x/a + H/A} \beta(b) F(g_x(ab)\Gamma) \right|^* \gg \delta^{32} \frac{H}{A}. \tag{6.9}$$

Let us say that  $(x, a) \in \mathcal{S}$  is *good* if  $x \in E^{**}$  and

$$\sum_{x/a < b \leq x/a + \delta^{100} H/A} |\beta(b)| \leq \delta^{50} \frac{H}{A} \quad \text{and} \quad \sum_{x/a < b \leq x/a + 2H/A} |\beta(b)|^2 \leq \delta^{-50} \frac{H}{A}. \tag{6.10}$$

The number of  $(x, a) \in \mathcal{S}$  that are not good is by Markov’s inequality

$$\begin{aligned} &\ll \delta^{-50} \frac{A}{H} \sum_{A < a \leq 2A} \sum_{X < x \leq 2X} \sum_{x/a < b \leq x/a + \delta^{100} H/A} |\beta(b)| \\ &\quad + \delta^{50} \frac{A}{H} \sum_{A < a \leq 2A} \sum_{X < x \leq 2X} \sum_{x/a < b \leq x/a + 2H/A} |\beta(b)|^2 \\ &\ll \delta^{100-50} A^2 \sum_{b \leq (2X+H)/A} |\beta(b)| + \delta^{50} A^2 \sum_{b \leq (2X+2H)/A} |\beta(b)|^2 \ll \delta^{49} AX. \end{aligned}$$

Thus, there are  $\gg \delta^{33} AX$  good pairs  $(x, a) \in \mathcal{S}$ .

We need to perform some discretization in the possible summation lengths. To this end, for  $z > 0$ , let

$$Q(z) := \min \left( [z, \infty) \cap \left( \delta^{100} \frac{H}{A} \mathbb{Z} \right) \right).$$

Now, by (6.9) and (6.10), for any good  $(x, a) \in \mathcal{S}$  we have

$$\left| \sum_{Q(x/a) < b \leq Q(x/a) + H/A} \beta(b) F(g_x(ab) \Gamma) \right|^* \gg \delta^{32} \frac{H}{A}. \tag{6.11}$$

Let  $C_1$  be a large constant, and let  $C_2$  be large in terms of  $C_1$ . We may assume that  $H/A \geq \delta^{-C_2}$ , since otherwise conclusion (i) of the proposition holds. Since for any good  $(x, a) \in \mathcal{S}$  we have (6.11), we may apply the nilsequence large sieve (Proposition 2.15) to (6.11) to deduce that there exists a collection  $(\tilde{g}_{Q(x/a), j})_{1 \leq j \leq J}$  of elements of  $\text{Poly}(\mathbb{Z} \rightarrow G)$  such that  $J \ll \delta^{-O(1)}$  and such that at least one of the following holds for any good  $(x, a) \in \mathcal{S}$ :

- (a) There is a non-trivial horizontal character  $\eta_{x,a}: G \rightarrow \mathbb{R}$  of Lipschitz norm  $\leq \delta^{-C_1}$  such that

$$\|\eta_{x,a} \circ g_x(a \cdot) \pmod{1}\|_{C^\infty(Q(x/a), Q(x/a) + H/A)} \leq \delta^{-C_1}.$$

- (b) For some  $1 \leq j \leq J$  (depending on  $(x, a)$ ) we have

$$g_x(a \cdot) \sim_{(Q(x/a), Q(x/a) + H/A), \delta^{C_1}} \tilde{g}_{Q(x/a), j} \cdot$$

We now split into two cases.

**Case 1: (a) holds for at least half of the good pairs in  $\mathcal{S}$ .**

By (2.3), whenever  $(x, a)$  has property (a) we have

$$\|\eta_{x,a} \circ g_x(a \cdot) \pmod{1}\|_{C^\infty(x/a, (x+H)/a)} \ll \delta^{-C_1}. \tag{6.12}$$

Now, by the pigeonhole principle there is some non-trivial horizontal character  $\eta': G \rightarrow \mathbb{R}$  of Lipschitz norm  $\leq \delta^{-C_1}$  and some set of integers in  $(X, 2X]$  of size  $\gg \delta^{C_1 + O(1)} X$  such that for each such  $x$  there exist  $\gg \delta^{C_1 + O(1)} A$  choices of  $a \in (A, 2A] \cap \mathbb{N}$  for which (6.12) holds with  $\eta_{x,a} = \eta'$ . By Corollary 2.4 (and the assumption  $H/A \geq \delta^{-C_2}$ ), we deduce that for some integer  $1 \leq |q| \ll \delta^{-O_{C_1}(1)}$  the horizontal character  $\eta'' = q\eta'$  satisfies

$$\|\eta'' \circ g_x \pmod{1}\|_{C^\infty(x, x+H)} \ll \delta^{-O_{C_1}(1)}$$

for  $\gg \delta^{C_1 + O(1)} X$  integers  $x \in (X, 2X]$ . Conclusion (ii) follows, since  $x \mapsto g_x$  is constant on intervals of the form  $[m, m + 1)$  with  $m \in \mathbb{N}$ .

**Case 2: (b) holds for at least half of the good pairs in  $\mathcal{S}$ .**

By Lemma 5.10(ii), for any pair  $(x, a)$  satisfying (b), for some  $1 \leq j \leq J$  we have

$$g_x(a \cdot) \sim_{\frac{1}{a}(x, x+H), \delta^{O(C_1)}} \tilde{g}_{Q(x/a), j} \cdot$$

By the pigeonhole principle, there exists some  $1 \leq j_0 \leq J$  such that for  $\gg \delta^{C_1+O(1)} AX$  pairs  $(x, a) \in \mathcal{S}$  we have

$$g_x(a \cdot) \sim_{\frac{1}{a}(x, x+H], \delta^{O(C_1)}} \tilde{g}_Q(x/a), j_0. \tag{6.13}$$

For  $t \in [0, 1)$ , let

$$\mathcal{Y}_t := \left( \frac{H}{A} \mathbb{Z} + t \frac{H}{A} \right) \cap \left[ \frac{X}{2A}, \frac{2X}{A} \right].$$

Then  $\mathcal{Y}_t$  is  $H/A$ -separated and has size  $\gg \frac{X}{H}$ . By the pigeonhole principle, there is some  $t_0 \in [0, 1)$  such that the set

$$\{(x, a) \in \mathcal{S} : (x, a) \text{ satisfies (6.13) and } Q(x/a) = Q(y) \text{ for some } y \in \mathcal{Y}_{t_0}\}$$

has size  $\gg \delta^{C_1+O(1)} AX$ . By the pigeonhole principle again, this implies that there exists an  $H$ -separated set  $E' \subset (X, 2X] \cap \mathbb{N}$  of size  $\gg \delta^{C_1+O(1)} X/H$  such that for every  $x \in E'$  there exist  $\gg \delta^{C_1+O(1)} A$  choices of  $a \in (A, 2A] \cap \mathbb{N}$  for which  $(x, a)$  satisfies (6.13) and  $Q(x/a) = Q(y)$  for some unique  $y \in \mathcal{Y}_{t_0}$ .

For  $y \in \mathcal{Y}_{t_0}$ , define  $\tilde{g}_y := \tilde{g}_{Q(y), j_0}$ . Then for every  $x \in E'$  there exist  $\gg \delta^{C_1+O(1)} A$  choices of  $a \in (A, 2A] \cap \mathbb{N}$  such that for some  $y \in \mathcal{Y}_{t_0}$  we have  $Q(x/a) = Q(y)$  (so that in particular  $|y - x/a| \leq \delta^{100} H/A$ ) and

$$g_x(a \cdot) \sim_{\frac{1}{a}(x, x+H], \delta^{O(C_1)}} \tilde{g}_y.$$

Taking  $\mathcal{J}_{-1} = \{(y, \tilde{g}_y) : y \in \mathcal{Y}_{t_0}\}$  and  $\mathcal{J}_0 = \{(x, g_x) : x \in E'\}$ , we now have the desired relation (6.6). □

Using the improved contagiousness result (Theorem 5.11), we have the following result that extends [42, Proposition 3.7] from the linear phase case to nilsequences and also, crucially, works with smaller  $H$ .<sup>12</sup> This step crucially allows us to “scale up” the map  $x \mapsto g_x$  of Proposition 6.4 to a larger scale where it is easier to analyze.

**Proposition 6.6 (Scaling up)** *Let  $Y \geq H \geq A \geq 2$ ,  $0 < \sigma < \frac{1}{2 \log Y}$ , and  $2 \leq C \leq \sigma^{-1}$ . Let  $G/\Gamma$  be a filtered nilmanifold of degree  $d$ , dimension  $D$  and complexity  $\leq \sigma^{-1}$ . Let  $\mathcal{J}_0$  be a  $(\sigma, H)$ -configuration in  $[Y, CY]$ , and let  $\mathcal{J}_1$  be a  $(\sigma, AH)$ -configuration in  $[AY, CAY]$  such that*

$$\mathcal{J}_0 \longrightarrow_{\sigma, C} \mathcal{J}_1.$$

Then, if

$$H \geq \sigma^{-K} A, \text{ and } A \geq \sigma^{-K} \tag{6.14}$$

---

<sup>12</sup>Writing [42, Proposition 3.7] in the notation of Proposition 6.6, the condition (6.14) would be replaced with a more restrictive condition of the type  $H \geq \sigma^{-K} A^2$ .

for a large enough constant  $K$  (depending on  $d, D$ ), there exists a  $(\sigma^{O(1)}, A^2H)$ -configuration  $\mathcal{J}_2$  in  $[A^2Y, 2CA^2Y]$  such that

$$\mathcal{J}_1 \longrightarrow_{\sigma^{O_d, D(1)}, O_{d, D}(C^2)} \mathcal{J}_2.$$

**Proof** We allow implied constants to depend on  $d, D$ . By Definition 6.2, we have

$$\sum_{(y, g_y) \in \mathcal{J}_1} \sum_{A < a \leq CA} \sum_{(x, g_x) \in \mathcal{J}_0} I_{a, C, AH, \sigma}(y, x) \gg \sigma \frac{Y}{H} \cdot \sigma A.$$

On using the Cauchy–Schwarz inequality, this implies that

$$\begin{aligned} \sum_{(x, g_x) \in \mathcal{J}_0} \left( \sum_{(y, g_y) \in \mathcal{J}_1} \sum_{A < a \leq CA} I_{a, C, AH, \sigma}(y, x) \right)^2 &\gg \frac{(\sigma^2 AY/H)^2}{|\mathcal{J}_0|} \\ &\gg \sigma^4 C^{-1} A^2 \frac{Y}{H}. \end{aligned}$$

Opening the square, we obtain

$$\sum_{(x, g_x) \in \mathcal{J}_0} \sum_{\substack{(y_1, g_{y_1}) \in \mathcal{J}_1 \\ (y_2, g_{y_2}) \in \mathcal{J}_1}} \sum_{\substack{A < a_1 \leq CA \\ A < a_2 \leq CA}} I_{a_1, C, AH, \sigma}(y_2, x) I_{a_2, C, AH, \sigma}(y_1, x) \gg \sigma^4 C^{-1} A^2 \frac{Y}{H}.$$

Note that for every non-zero term in this sum we have  $|y_2 - a_1x| \leq CAH$  and  $|y_1 - a_2x| \leq CAH$ , hence  $|a_1y_1 - a_2y_2| \leq 2C^2A^2H$ . In particular, there exists  $z \in C^2A^2H\mathbb{Z}$  with  $z \in [A^2Y, 2C^2A^2Y]$  such that  $|z - a_1y_1| \leq 2C^2A^2H$  and  $|z - a_2y_2| \leq 2C^2A^2H$ . We conclude that

$$\begin{aligned} &\sum_{(x, g_x) \in \mathcal{J}_0} \sum_{\substack{(y_1, g_{y_1}) \in \mathcal{J}_1 \\ (y_2, g_{y_2}) \in \mathcal{J}_1}} \sum_{\substack{A < a_1 \leq CA \\ A < a_2 \leq CA}} \sum_{\substack{z \in [A^2Y, 2C^2A^2Y] \\ z \in C^2A^2H\mathbb{Z}}} \mathbf{1}_{|z - a_1y_1|, |z - a_2y_2| \leq 2C^2A^2H} \\ &\quad \times I_{a_1, C, AH, \sigma}(y_2, x) I_{a_2, C, AH, \sigma}(y_1, x) \\ &\gg \sigma^4 C^{-1} A^2 \frac{Y}{H}. \end{aligned}$$

Moving the  $z$  sum outside and applying the pigeonhole principle, we see that there exists a  $C^2A^2H$ -separated set  $\mathcal{J}_2^*$  in  $[A^2Y, 2C^2A^2Y]$  of size  $\gg \sigma^{O(1)}Y/H$  such that for each  $z \in \mathcal{J}_2^*$  there is a set  $\mathcal{Q}_z$  of  $\gg \sigma^{O(1)}A^2$  quadruples  $(a_1, a_2, (y_1, g_{y_1}), (y_2, g_{y_2})) \in ([A, CA] \cap \mathbb{N})^2 \times \mathcal{J}_1^2$  such that

$$|z - a_1y_1|, |z - a_2y_2| \leq 2C^2A^2H$$

and

$$I_{a_1, C, AH, \sigma}(y_2, x) = I_{a_2, C, AH, \sigma}(y_1, x) = 1$$

for some  $(x, g_x) \in \mathcal{J}_0$ . For fixed  $z$ , Each  $a_1$  is associated to  $O(\sigma^{-O(1)})$  many choices of  $y_1$ , so by pruning we may assume that  $y_1$  is determined by  $a_1$ ; similarly we may assume  $y_2$  is determined by  $a_2$ . By further pruning, we may assume that whenever  $(a_1, a_2, (y_1, g_{y_1}), (y_2, g_{y_2})) \in \mathcal{Q}_z$  holds for some  $(a_1, a_2)$ , it holds for  $\gg \sigma^{O(1)} A^2$  such tuples. By Lemma 5.10(ii), we now obtain

$$g_{y_1}(a_2 \cdot) \sim \frac{1}{a_1 a_2} (z, z + A^2 H)_{\sigma^{O(1)}} g_x,$$

and similarly

$$g_{y_2}(a_1 \cdot) \sim \frac{1}{a_1 a_2} (z, z + A^2 H)_{\sigma^{O(1)}} g_x,$$

and hence by Lemma 5.10(i) for  $(a_1, a_2, (y_1, g_{y_1}), (y_2, g_{y_2})) \in \mathcal{Q}_z$  we have

$$g_{y_1}(a_2 \cdot) \sim \frac{1}{a_1 a_2} (z, z + A^2 H)_{\sigma^{O(1)}} g_{y_2}(a_1 \cdot).$$

Let  $z \in \mathcal{J}_2^*$ , and let  $(y, g_y) \in \mathcal{J}_1$  be the last coordinate of some element of  $\mathcal{Q}_z$ . Since  $H/A \geq \sigma^{-K}$  and  $K$  is a large enough constant, from the nilsequence contagion result (Theorem 5.11) and Lemma 5.10(i) we conclude that there exists  $g_z \in \text{Poly}(\mathbb{Z} \rightarrow G)$  and  $\gg \sigma^{O(1)} A$  values of  $a \in [A, CA] \cap \mathbb{N}$  such that

$$g_z(a \cdot) \sim \frac{1}{a} (z, z + A^2 H)_{\sigma^{O(1)}} g_y.$$

We select such a  $g_z$  for each  $z \in \mathcal{J}_2^*$ , and set  $\mathcal{J}_2 = \{(z, g_z) : z \in \mathcal{J}_2^*\}$ . The desired claim follows. □

We are going to apply Proposition 6.6 iteratively. The next proposition tells us how we can obtain information on  $x \mapsto g_x$  in the original scale after passing to a suitably large scale.

**Proposition 6.7** (Conclusion of repeated scaling) *Let  $X \geq H \geq 2$  and  $M \geq X$ . Let  $0 < \sigma < \frac{1}{2 \log M}$  and  $2 \leq C \leq \sigma^{-1}$ . Let  $G/\Gamma$  be a filtered nilmanifold of degree  $d$ , dimension  $D$  and complexity  $\leq \sigma^{-1}$ . Suppose that there exists a  $(\sigma, H)$ -configuration  $\mathcal{J}_0 = \{(x, g_x) : x \in E'\}$  in  $[X, 2X]$  and a  $(\sigma, MH)$ -configuration  $\mathcal{J}_1$  in  $[MX, CMX]$  such that*

$$\mathcal{J}_0 \xrightarrow{\sigma, C} \mathcal{J}_1.$$

*Then, if  $H \geq \sigma^{-K}$  for a large enough constant  $K$  (in terms of  $d, D$ ), the following hold.*

- (i) *If  $G/\Gamma = \mathbb{R}/\mathbb{Z}$  and  $g_x = P_x$  for some polynomials  $P_x : \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$  of degree at most  $d$ , then there exists a real number  $|T| \ll \sigma^{-O_{d,D,C}(1)} (X/H)^{d+1}$  and a positive integer  $1 \leq q' \ll \sigma^{-O_{d,D,C}(1)}$  such that*

$$\|e(P_x(n))n^{-iT}\|_{\text{TV}((x, x+H)\cap\mathbb{Z}; q')} \ll \sigma^{-O_{d,D,C}(1)}.$$

*for all  $x$  in a subset of  $E'$  of cardinality  $\gg \sigma^{O_{d,D,C}(1)} X/H$ .*

(ii) If  $G$  is non-abelian with one-dimensional center, then there exists a non-trivial horizontal character  $\eta: G \rightarrow \mathbb{R}$  of Lipschitz norm  $O(\sigma^{-O_{d,D,C}(1)})$  such that

$$\|\eta \circ g_x\|_{C^\infty(x,x+H)} \ll \sigma^{-O_{d,D,C}(1)}$$

for all  $x$  in a subset of  $E'$  of cardinality  $\gg \sigma^{O_{d,D,C}(1)} X/H$ .

**Proof** We allow implied constants to depend on  $d, D, C$ . We arbitrarily fix an element  $(z, g_z)$  of  $\mathcal{J}_1$ . By assumption, we have

$$\sum_{(x, g_x) \in \mathcal{J}_0} \sum_{M < m \leq CM} I_{m,C,MH,\sigma}(z, x) \geq \sigma M.$$

Applying the pigeonhole principle and the fact that  $|\mathcal{J}_0| \leq \frac{X}{H}$ , we see that there is a  $(\sigma/(2C), H)$ -configuration  $\mathcal{J}'_0 \subset \mathcal{J}_0$  such that for every  $(x, g_x) \in \mathcal{J}'_0$  we have

$$\sum_{M < m \leq CM} I_{m,C,MH,\sigma}(z, x) \gg \sigma \frac{MH}{X}.$$

Let  $\mathcal{M}_x$  be the set of integers  $M < m \leq CM$  for which the summand is positive. Then  $|\mathcal{M}_x| \gg \sigma MH/X$ . Note that

$$\mathcal{M}_x \subset \left[ \frac{z}{x} - \frac{CMH}{X}, \frac{z}{x} + \frac{CMH}{X} \right] \tag{6.15}$$

and that for all  $m \in \mathcal{M}_x$  we have

$$g_z(m \cdot) \sim \frac{1}{m} {}_{(z, z+MH), \sigma} g_x. \tag{6.16}$$

Using Lemma 5.10(i)–(ii) and the fact that  $|\frac{z}{m_i} - \frac{z}{(m_1+m_2)/2}| \ll CH$  for  $i = 1, 2$  and  $m_i \in \mathcal{M}_x$ , for any  $x$  that is the first coordinate of some element of  $\mathcal{J}'_0$  and any  $(m_1, m_2) \in \mathcal{M}_x^2$  we have

$$g_z(m_1 \cdot) \sim \frac{2}{m_1+m_2} {}_{(z, z+MH), \sigma^{O(1)}} g_z(m_2 \cdot).$$

For any  $M' \geq 1$ , set  $J_{M'} := ((1 - \frac{10MH}{z}) \frac{z}{M'}, (1 + \frac{10MH}{z}) \frac{z}{M'})$ . In view of the pigeonhole principle, Lemma 5.10(ii) and (6.15) and (6.16), we can find a  $10CMH/X$ -separated set  $\mathcal{M} \subset [M, CM]$  (which is the union of  $\mathcal{M}_x$  for some  $x \in \mathcal{J}'_0$ ) of cardinality  $\gg \sigma^{O(1)} X/H$  such that, denoting  $I_{M'} = (M', (1 + \frac{H}{X})M')$ , for each  $M' \in \mathcal{M}$  there exist  $\gg \sigma^{O(1)} |I_{M'}|^2$  pairs  $(m_1, m_2) \in I_{M'}^2$  for which

$$g_z(m_1 \cdot) \sim_{J_{M'}, \sigma^{O(1)}} g_z(m_2 \cdot), \tag{6.17}$$

and additionally

$$g_z(m_1 \cdot) \sim \frac{1}{m_1} {}_{(z, z+MH), \sigma} g_x \tag{6.18}$$

for some  $x$  that is the first coordinate of an element of  $\mathcal{J}'_0$  with  $|z/x - m_1| \leq 2CMH/X$ . Unwrapping the definition of the relation  $\sim_{I,\eta}$ , from (6.17) we conclude that for each  $M' \in \mathcal{M}$  there exist  $\gg \sigma^{O(1)}|I_{M'}|^2$  pairs  $(m_1, m_2) \in I_{M'}^2$  for which there exists a factorization

$$g_z(m_1 \cdot) = \varepsilon_{m_1, m_2} g_z(m_2 \cdot) \gamma_{m_1, m_2}, \tag{6.19}$$

where  $\varepsilon_{m_1, m_2}$  is  $(\sigma^{-O(1)}, J_{M'})$ -smooth and  $\gamma_{m_1, m_2}$  is  $\sigma^{-O(1)}$ -rational. The argument now splits into two cases.

**Proof of part (ii).** Applying Proposition 2.16 (with  $(z, M, MH)$  in place of  $(X, A, H)$ ) and using the fact that  $G$  is non-abelian, we conclude that either

$$MH \ll \sigma^{-O(1)} \max(M, X) \tag{6.20}$$

or there exists a non-trivial horizontal character  $\eta: G \rightarrow \mathbb{R}$  of Lipschitz norm at most  $O(\sigma^{-O(1)})$  such that

$$\|\eta \circ g_z\|_{C^\infty(z, z+MH]} \ll \sigma^{-O(1)}. \tag{6.21}$$

The estimate (6.20) cannot hold if  $K$  is chosen large enough in the statement of the proposition. Hence we have (6.21). Recalling that (6.18) holds for any  $m_1 \in \mathcal{M}$ , by Lemma 5.10(ii) we see that for any  $m \in \mathcal{M}$  there is some  $x$  that is the first coordinate of an element of  $\mathcal{J}'_0$  with  $|z/x - m| \leq 2CMH/X$  such that

$$\|\eta \circ g_x\|_{C^\infty(x, x+H]} \ll \sigma^{-O(1)}. \tag{6.22}$$

Since  $\mathcal{M}$  has cardinality  $\gg \sigma^{O(1)}X/H$  and is  $10CMH/X$ -separated, each  $m \in \mathcal{M}$  corresponds to a different  $x$ , so there must exist  $\gg \sigma^{O(1)}X/H$  elements  $x \in E'$  for which (6.22) holds. Now part (ii) follows.

**Proof of part (i).** Now we have  $G/\Gamma = \mathbb{R}/\mathbb{Z}$ . Then by (6.19) for any  $M' \in \mathcal{M}$  there exist  $\gg \sigma^{O(1)}|I_{M'}|^2$  pairs  $(m_1, m_2) \in I_{M'}^2$  such that

$$\|q_{m_1, m_2, z}(g_z(m_2 \cdot) - g_z(m_1 \cdot)) \bmod 1\|_{C^\infty(J_{M'})} \ll \sigma^{-O(1)},$$

for some integer  $1 \leq q_{m_1, m_2, z} \leq \sigma^{-O(1)}$ . Pigeonholing in the possible choices of  $q_{m_1, m_2, z}$ , we can assume that this quantity is independent of  $m_1, m_2, z$ . Now Proposition 2.5 (with  $(z, M, MH)$  in place of  $(X, A, H)$ ) implies that there exists some integer  $1 \leq q \leq \sigma^{-O(1)}$  and some real number  $T \ll (X/H)^{d+1} \sigma^{-O(1)}$  such that

$$\|q \cdot g_z - T \log(\cdot)\|_{C^d(z, z+MH]} \ll \sigma^{-O(1)}. \tag{6.23}$$

Let  $m \in \mathcal{M}$ . Dilating (6.23) by  $m$  (using (2.4) and the fact that  $T \log(\cdot)$  is close to a polynomial of degree  $d$  by Taylor expansion), we obtain

$$\|q \cdot g_z(m \cdot) - T \log(\cdot)\|_{C^d(\frac{z}{m}, \frac{z}{m} + \frac{MH}{m})} \ll \sigma^{-O(1)}.$$

On the other hand, from (6.18) there is some  $x$  that is the first coordinate of some element of  $\mathcal{J}'_0$  with  $|z/x - m| \leq 2CMH/X$  for which we have

$$\|q'_{z, m, x} \cdot g_z(m \cdot) - q'_{z, m, x} \cdot g_x\|_{C^d(\frac{z}{m}, \frac{z}{m} + \frac{MH}{m})} \ll \sigma^{-O(1)}$$

for some integers  $1 \leq q'_{z,m,x} \ll \sigma^{-O(1)}$ . Hence by the triangle inequality and (2.3) (and the fact that  $T \log(\cdot)$  is close to a polynomial of degree  $d$ ) we have

$$\|qq'_{z,m,x} \cdot g_x - q'_{z,m,x} T \log(\cdot)\|_{C^d(x,x+H)} \ll \sigma^{-O(1)}.$$

By the pigeonhole principle, we may thus find a natural number  $1 \leq q'' \ll \sigma^{-O(1)}$  such that

$$\|qq'' \cdot g_x - q'' T \log(\cdot)\|_{C^d(x,x+H)} \ll \sigma^{-O(1)}. \tag{6.24}$$

Since  $\mathcal{M}$  is  $10CMH/X$ -separated, each  $m \in \mathcal{M}$  with  $|z/x - m| \leq 2CMH/X$  corresponds to a different choice of  $x$ , so the above holds for all  $x$  in a subset of  $E'$  of cardinality  $\gg \sigma^{O(1)}X/H$ .

By Taylor expansion, we can find a polynomial  $Q_x$  of degree  $d' = O(1)$  such that

$$\|Q_x - \frac{T}{q} \log(\cdot)\|_{C^d(x,x+H)} \ll 1 \tag{6.25}$$

and thus by (6.24), (6.25) and the triangle inequality

$$\|qq''(g_x - Q_x)\|_{C^d(x,x+H)} \ll \sigma^{-O(1)}.$$

This implies that

$$qq''(g_x(n) - Q_x(n)) = \sum_{j=0}^{\max(d,d')} \alpha_j \binom{n - \lfloor x \rfloor}{j} \pmod 1$$

for some coefficients  $\alpha_j$  of size  $O(\sigma^{-O(1)}/H^j)$ . One can then calculate

$$\|e(g_x - Q_x)\|_{\text{TV}((x,x+H]; \max(d,d')!qq'')} \ll \sigma^{-O(1)}$$

and thus by (6.25) and the triangle inequality

$$\|e(g_x - \frac{T}{q} \log(\cdot))\|_{\text{TV}((x,x+H]; \max(d,d')!qq'')} \ll \sigma^{-O(1)}.$$

This completes the proof of part (i) after some relabeling. □

We are now ready to conclude the proof of our type II estimates by combining the results of this section.

**Proof of Theorem 4.2(ii)–(iii)** Let the notation and assumptions be as in Theorem 4.2(ii)–(iii). Let  $B$  be a large enough constant (in terms of  $d, D, \varepsilon$ ), and let us say that  $x$  is *bad* if there is no non-trivial horizontal character  $\eta$  of Lipschitz norm  $\leq \delta^{-B}$  such that  $\|\eta \circ g_x \pmod 1\|_{C^\infty(x,x+H)} \leq \delta^{-B}$ . We may assume that all  $x \in E$  are bad (since if the set  $E'$  of non-bad  $x$  has measure  $\geq \delta X/2$ , say, we are done, and otherwise we may consider the set  $E \setminus E'$ , which has measure  $\geq \delta X/2$ ).

We apply Proposition 6.4. If conclusion (i) or (ii) of that proposition holds, we are done. Suppose then that conclusion (iii) holds. Then there exists  $A \in [A_{II}^-, A_{II}^+]$ , a  $(\delta^{O_{d,D,\varepsilon}(1)}, H)$ -configuration  $\mathcal{J}_0$  in  $[X, 2X]$  and a  $(\delta^{O_{d,D,\varepsilon}(1)}, H/A)$ -configuration  $\mathcal{J}_{-1}$  in  $[X/(2A), 2X/A]$ , such that  $\mathcal{J}_0 = \{(x, g_x) : x \in E'\}$  for some set  $E' \subset E$  and

$$\mathcal{J}_{-1} \longrightarrow_{\delta^{O_{d,D,\varepsilon}(1)}, O_{d,D,\varepsilon}(1)} \mathcal{J}_0.$$

Let  $K \in \mathbb{N}$  be the least integer such that  $A^K \geq X$ . Note that as  $A_{II}^- \geq X^\varepsilon$  we have  $K \ll_\varepsilon 1$ . Applying Proposition 6.6  $K$  times, and noting that for any given constant  $C'$  we may assume that  $H \geq \delta^{-C'} A$ , we see that for all  $2 \leq k \leq K$  there is some  $C_k = O_{d,D,\varepsilon}(1)$  and some  $(\delta^{O_{d,D,\varepsilon}(1)}, HA^k)$ -configuration  $\mathcal{J}_k$  in  $[A^k X/2, C_k A^k X/2]$  such that we have

$$\mathcal{J}_{k-1} \longrightarrow_{\delta^{O_{d,D,\varepsilon}(1)}, C_k} \mathcal{J}_k.$$

By applying Lemma 6.3 repeatedly, this implies that

$$\mathcal{J}_0 \longrightarrow_{\delta^{O_{d,D,\varepsilon}(1)}, C'_K} \mathcal{J}_K$$

for some  $C'_K = O_{d,D,\varepsilon}(1)$ . But now by Proposition 6.7 (and the assumption that  $x$  is bad for all  $x \in E$ ), we obtain one of the conclusions of Theorem 4.2(ii)–(iii).  $\square$

### 7 Controlling the Gowers uniformity norms

In order to deduce our Gowers uniformity result in almost all short intervals (Theorem 1.3) 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 restricted to short intervals.<sup>13</sup> In the case of long intervals, the existence of pseudorandom majorants for these functions follows from earlier works [5, 31]. In the case of short intervals  $(X, X + H]$ , the existence of these pseudorandom majorants is established in the prequel [30] for  $\Lambda$  when  $H \geq X^\varepsilon$ , and for  $d_k$  when  $H \geq X^{1/5+\varepsilon}$ . The main purposes of this section are to show that the majorant for  $d_k$  constructed in [30, Sect. 9] is pseudorandom in  $(x, x + H]$  for almost all  $x \in [X, 2X]$  in the larger regime  $H \geq X^\varepsilon$  and to deduce Theorem 1.3 from this. For this deduction in the case of  $d_k$ , we will need a logarithm-free type  $I$  estimate given as Proposition 7.6.

We begin by recalling from [30] the relevant definition of pseudorandomness.

**Definition 7.1** (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  $(D, \eta)$ -pseudorandom at location  $x$  and scale  $H$  if the function  $v_x(n) := v(x + n)$  satisfies the following. Let

<sup>13</sup>In some applications of Gowers norm theory, the recent quasipolynomial inverse theorem of Leng–Sah–Sawhney [18] has been able to dispose of the need for pseudorandom majorants, when combined with strong dimension-uniform bounds on the nilsequence side. As discussed in Remark 1.4, we do not know how to obtain good dimension dependencies in our main theorems, and therefore we need to construct pseudorandom majorants.

$d, t \leq D$  and  $\psi_1, \dots, \psi_t$  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  $|\dot{\psi}_i| \leq D$  and  $|\psi_i(0)| \leq DH$ , and with  $\dot{\psi}_i$  and  $\dot{\psi}_j$  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 7.2** Note that if a function  $v: \mathbb{Z} \rightarrow \mathbb{R}_{\geq 0}$  is  $(D, \eta)$ -pseudorandom at location  $x$  and scale  $H$ , then in particular

$$\|v_x - 1\|_{U^s(x, x+H)} \ll_D \eta^{1/2^s}.$$

for every  $s \leq D$ .

We then state the inverse theorem for unbounded functions that we are going to use.

**Proposition 7.3** (An inverse theorem for pseudorandomly bounded functions) [30, Proposition 9.4] *Let  $s \in \mathbb{N}$  and  $\eta \in (0, 1)$ . Let  $I$  be an interval of length  $\geq 2$ . 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  $G/\Gamma$  and any Lipschitz function  $F: G/\Gamma \rightarrow \mathbb{C}$ , we have*

$$\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \frac{1}{|I|} \sum_{n \in I} f(n) \overline{F}(g(n)\Gamma) \right| \ll_{\|F\|_{\text{Lip}}, G/\Gamma} \eta.$$

Then we have the Gowers uniformity estimate

$$\|f\|_{U^s(I)} = o_{s, \eta \rightarrow 0}(1).$$

The following lemma from [30] will be important for us.

**Lemma 7.4** (Pseudorandomness over long intervals implies pseudorandomness over short intervals) [30, Lemma 9.5] *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 in terms of  $C$ . Let  $v$  be  $(D, \eta)$ -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  $\lambda_n$  such that*

$$v(n) = \sum_{\substack{d|An+B \\ d \leq X^{\varepsilon/(2D)}}} \lambda_d \quad \text{for } n \notin \mathcal{S}, \tag{7.1}$$

$$|\lambda_n| \leq (\log X)^k d(n)^k \quad \text{for all } n,$$

$$|v(n)| \leq (\log X)^k d(An + B)^k \quad \text{for } n \in \mathcal{S}.$$

Also suppose that  $\mathcal{S}$  is small in the sense that

$$|\mathcal{S} \cap [y - 2DH, y + 2DH]| \ll H/(\log X)^{4C} \quad \text{for } y \in \{0, X\}.$$

Then  $v$  is  $(D, 2\eta)$ -pseudorandom at location  $X$  and scale  $H$ .

We are now ready to state and prove the existence of 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  $W = \prod_{p \leq w} p$ .

**Lemma 7.5** (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 enough growing function of  $X$ , and denote  $W = \prod_{p \leq w} p$ . Also let  $w \leq \tilde{w} \leq \exp((\log X)^{1/10})$  and  $\tilde{W} = \prod_{p \leq \tilde{w}} p$ .*

(i) *There exists a constant  $C_0 \geq 1$  such that each of the functions*

$$\frac{\varphi(W)}{W} \Lambda(Wn + b)/C_0, \quad \frac{\varphi(W)}{W} \Lambda_{\tilde{w}}(Wn + b)$$

*for  $1 \leq b \leq W$  with  $(b, W) = 1$ , is majorized on  $(X, X + H]$  by a  $(D, \eta)$ -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  $(D, \eta)$ -pseudorandom at location  $X$  and scale  $H$ .*

(ii) *Let  $W'$  be such that  $W \mid W' \mid W^{\lfloor w \rfloor}$ . There exists a constant  $C_k \geq 1$  such that, for all but at most  $XH^{-1/2}(\log X)^{O(1)}$  values of  $x \in [X, 2X]$ , the functions*

$$\begin{aligned} &(\log X) \frac{\varphi(W)}{W} \prod_{w \leq p \leq X} \left(1 + \frac{k}{p}\right)^{-1} d_k(W'n + b)/C_k, \\ &(\log X) \frac{\varphi(W)}{W} \prod_{w \leq p \leq X} \left(1 + \frac{k}{p}\right)^{-1} d_k^\sharp(W'n + b)/C_k \end{aligned}$$

*for  $1 \leq b \leq W'$  with  $(b, W') = 1$ , are majorized on  $(x, x + H]$  by a  $(D, \eta)$ -pseudorandom function at location  $x$  and scale  $H$  for some  $\eta = o_{w \rightarrow \infty}(1)$ .*

**Proof** Part (i) is precisely [30, Lemma 9.6(1)]. For part (ii), we follow the construction in the proof of [30, Lemma 9.6(2)]. From the definition of  $d_k^\sharp$ , it follows that  $d_k^\sharp(n) \ll_k d_k(n)$  for all  $1 \leq n \leq 2X$  (see [30, (3.14)]), so by Lemma 7.4 it suffices to show that, for some  $C'_k \geq 1$  and all  $x \in [X, 2X]$  outside an admissible exceptional set, 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 majorized by a function that is  $(D, o_{w \rightarrow \infty}(1))$ -pseudorandom at location 0 and scale  $H$ , and which is of the form (7.1) outside an exceptional set  $\mathcal{S}$  satisfying the size bounds

$$\begin{aligned}
 |\mathcal{S} \cap [x - 2DH, x + 2DH]| &\ll H/(\log X)^{4C}, \\
 |\mathcal{S} \cap [-2DH, 2DH]| &\ll H/(\log X)^{4C}
 \end{aligned}
 \tag{7.2}$$

for some large enough  $C \geq 1$ .

We claim that this follows from [30, Lemma 9.6]. Indeed, by that lemma, the function  $h$  is majorized by a  $(D, o_{X \rightarrow \infty}(1))$ -pseudorandom function at location 0 and scale  $H$ . The fact that the majorant is of the form (7.1) follows from the arguments in [30, proof of Lemma 9.6], with the exceptional set  $\mathcal{S}$  being defined as

$$\mathcal{S} := \left\{ n \leq 2DX : \exists p : v_p(n) \geq \max \left\{ 2, 8C \frac{\log \log X}{\log p} \right\} \text{ or } \prod_{p \leq X^{1/(\log \log X)^3}} p^{v_p(n)} \geq X^{\gamma/\log \log X} \right\}$$

for some constant  $\gamma > 0$ . Also from [30, proof of Lemma 9.6], we have the second estimate in (7.2). The remaining task now is to prove the first estimate in (7.2).

We now use Rankin’s trick to bound  $|\mathcal{S} \cap [x - 2DH, x + 2DH]|$  for  $x \in [X, 2X]$ . Denoting  $v = X^{1/(\log \log X)^3}$ , the number of  $n \in [x - 2DH, x + 2DH]$  that satisfy  $\prod_{p \leq X^{1/(\log \log X)^3}} p^{v_p(n)} \geq X^{\gamma/\log \log X}$  is (as in [30])

$$\begin{aligned}
 &\ll \sum_{\substack{ab \in [x-2DH, x+2DH] \\ p|a \implies p > v \\ p|b \implies p \leq v \\ b \geq X^{\gamma/\log \log X}}} 1 \leq \sum_{\substack{ab \in [x-2DH, x+2DH] \\ p|a \implies p > v \\ p|b \implies p \leq v}} \left( \frac{b}{X^{\gamma/\log \log X}} \right)^{\frac{10C(\log \log X)^2}{\gamma \log X}} \\
 &\ll \frac{1}{(\log X)^{10C}} \sum_{n \in [x-2DH, x+2DH]} g(n),
 \end{aligned}
 \tag{7.3}$$

where  $g$  is the completely multiplicative function satisfying

$$g(p) = \begin{cases} 1, & \text{if } p > v, \\ p^{\frac{10C(\log \log X)^2}{\gamma \log X}}, & \text{if } p \leq v. \end{cases}$$

By Shiu’s bound (Lemma 2.18), we conclude that (7.3) is  $\ll H(\log X)^{-4C}$ .

We are left with bounding the number of  $n \in [x - 2DH, x + 2DH]$  satisfying  $v_p(n) \geq \max\{2, 8C \frac{\log \log X}{\log p}\}$  for some  $p$ . This number is

$$\ll \sum_{p < (\log X)^{4C}} Hp^{-8C(\log \log X)/(\log p)} + \sum_{(\log X)^{4C} \leq p \leq H^{1/2}} \frac{H}{p^2}$$

$$\begin{aligned}
 &+ \sum_{H^{1/2} < p \leq (2X)^{1/2}} |\{n \in [x - 2DH, x + 2DH]: p^2 | n\}| \\
 &\ll H(\log X)^{-4C} + \sum_{H^{1/2} < p \leq (2X)^{1/2}} |\{n \in [x - 2DH, x + 2DH]: p^2 | n\}|. \tag{7.4}
 \end{aligned}$$

Define  $\mathcal{E}$  to be the set of  $x \in [X, 2X]$  such that the sum over  $p$  in (7.4) is  $\geq H(\log X)^{-4C}$ . Observe that

$$\begin{aligned}
 |\mathcal{E}|H(\log X)^{-4C} &\leq \sum_{\substack{X \leq x \leq 2X \\ x \in \mathbb{N}}} \sum_{H^{1/2} < p \leq (2X)^{1/2}} |\{n \in [x - 2DH, x + 2DH]: p^2 | n\}| \\
 &\ll H \sum_{H^{1/2} < p \leq (2X)^{1/2}} \sum_{X-2DH \leq n \leq 2X+2DH} \frac{1}{p^2 |n|} \\
 &\ll XH \sum_{H^{1/2} < p \leq (2X)^{1/2}} \frac{1}{p^2} \ll XH^{1/2},
 \end{aligned}$$

so we have  $|\mathcal{E}| \ll XH^{-1/2}(\log X)^{4C}$ . We conclude that the desired property (7.2) holds for all  $x \in [X, 2X]$  outside the sufficiently small exceptional set  $\mathcal{E}$ .  $\square$

### 7.1 Gowers uniformity of the von Mangoldt and Möbius functions

We are now ready to prove Theorem 1.3(i)–(ii).

**Proof of Theorem 1.3(i)–(ii)** (i) Let  $1 \leq a \leq W$  with  $(a, W) = 1$  (where, as in the theorem,  $W = \prod_{p \leq w} p$  with  $w$  tending to infinity slowly with  $X$ ). Set  $f = \frac{\varphi(W)}{W}(\Lambda - \Lambda^\sharp)$ . By Theorem 1.1(ii), there is an exceptional set  $\mathcal{E}$  of measure  $\ll_A X(\log X)^{-A}$  such that for all  $x \in [X(\log X)^{-A/2}, 2X] \setminus \mathcal{E}$ ,  $X^{1/3+\varepsilon/2} \leq H \leq X^{1-\varepsilon}$  and for any  $F, G/\Gamma$  as in that theorem, we have

$$\begin{aligned}
 &\sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H} f(Wn+a)\overline{F}(g(n)\Gamma) \right| \\
 &= \sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{\substack{Wx+a < n \leq W(x+H)+a \\ n \equiv a \pmod{W}}} f(n)\overline{F}\left(g\left(\frac{n-a}{W}\right)\Gamma\right) \right| \\
 &\ll_A H/\log^A X,
 \end{aligned}$$

since there exists a polynomial sequence  $\tilde{g}: \mathbb{Z} \rightarrow G$  such that  $\tilde{g}(n) = g((n-a)/W)$  for all  $n \equiv a \pmod{W}$ . Using Lemma 7.5(i), we can find functions  $v_j$  for  $j \in \{1, 2\}$  that are  $(2s, o_{w \rightarrow \infty}(1))$ -pseudorandom at location  $x$  and scale  $H$  for all  $x \in [X(\log X)^{-A/2}, 2X] \setminus \mathcal{E}$ , and such that  $v_1(n)$  and  $v_2(n)$  majorize

$$\frac{\varphi(W)}{W}\Lambda(Wn+a)/C_0, \quad \text{and} \quad v_2(n) = \frac{\varphi(W)}{W}\Lambda^\sharp(Wn+a),$$

respectively, where  $C_0$  is a suitably chosen constant. In particular, by the triangle inequality for the Gowers norms, we have

$$\begin{aligned} \left\| \frac{\nu_1 + \nu_2}{2} - 1 \right\|_{U^{2s}(x, x+H)} &\leq \frac{1}{2} (\|\nu_1 - 1\|_{U^{2s}(x, x+H)} + \|\nu_2 - 1\|_{U^{2s}(x, x+H)}) \\ &= o_{w \rightarrow \infty}(1). \end{aligned}$$

Since  $|f(Wn + a)| \ll \nu_1(n) + \nu_2(n)$ , the inverse theorem for the Gowers norms (Proposition 7.3) tells us that

$$\begin{aligned} \left\| \frac{\varphi(W)}{W} (\Lambda(W \cdot + a) - \Lambda^\sharp(W \cdot + a)) \right\|_{U^s(x, x+H)} &= \|f(W \cdot + a)\|_{U^s(x, x+H)} \\ &= o_{w \rightarrow \infty}(1). \end{aligned} \tag{7.5}$$

On the other hand, by Lemma 7.5(i), Remark 7.2, and the definition of  $\Lambda^\sharp$  we also have for  $x \in [X(\log X)^{-A/2}, 2X] \setminus \mathcal{E}$  that

$$\left\| \frac{\varphi(W)}{W} \Lambda^\sharp(W \cdot + a) - 1 \right\|_{U^s(x, x+H)} = o_{w \rightarrow \infty}(1). \tag{7.6}$$

By (7.5), (7.6) and the triangle inequality for the Gowers norms, we obtain

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

By [30, Lemma 9.8], this implies that also

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

holds for all  $x \in [X(\log X)^{-A/2}, 2X] \setminus \mathcal{E}$ .

(ii) This part follows from the same argument as part (i), taking  $f = \mu$  and  $W = 1$  and applying Theorem 1.1(iv) (and now Lemma 7.5 is not needed).  $\square$

### 7.2 Gowers uniformity of the divisor functions

For the proof of Theorem 1.3(iii), we shall need a version of the type  $I$  inverse theorem that does not lose any logarithms. Note that Theorem 4.2(i) does lose logarithms since we assume there that  $\delta < 1/\log X$ , so that losing powers of  $\delta$  corresponds to losing powers of logarithm).

**Proposition 7.6** (A log-free type  $I$  estimate) *Let  $d, D, k \geq 1$ ,  $X \geq H \geq 2$ , and  $\delta \in (0, 1/2)$ . Let  $1 \leq M \leq \delta^K H$  for some large enough  $K = K(d, D)$ , and let  $G/\Gamma$  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*

$$\left| \sum_{X < n \leq X+H} f(n) F(g(n)\Gamma) \right|^* \geq \delta H (\log X)^{k-1},$$

where  $g : \mathbb{Z} \rightarrow G$  is a polynomial map. Suppose that  $f = \alpha * \beta$ , where  $|\alpha(m)| \leq d_{k-1}(m)$ ,  $\alpha$  is supported on  $(M, 2M]$ , and  $\|\beta\|_{\text{TV}(\mathbb{N})} \leq \frac{1}{\delta}$ . Then we have

$$\left| \sum_{X < n \leq X+H} f(n) \right|^* \gg_{d,D,k} \delta^{O_{d,D,k}(1)} H(\log X)^{k-1}. \tag{7.7}$$

**Proof** In what follows, we let  $d, D, k \geq 1$  be fixed; we allow all implied constants to depend on these parameters. Let  $A \geq 1$  be large in terms of these fixed parameters. By the factorization theorem for nilsequences [7, Theorem 1.19], there exist a rational subgroup  $G' \subset G$  and  $\delta^{-1} \leq Q \leq \delta^{-O_A(1)}$  such that we can factorize  $g = \varepsilon g' \gamma$ , where the polynomial sequences  $\varepsilon, g' \gamma : \mathbb{Z} \rightarrow G$  have the following properties:

- $\varepsilon$  is  $(Q, [X, X + H])$ -smooth;
- $g'$  is totally  $Q^{-A}$ -equidistributed in  $G'/(G' \cap \Gamma)$ ;
- $\gamma$  is  $Q$ -rational, with  $(\gamma(n)\Gamma)_{n \in \mathbb{Z}}$  having period at most  $Q$ .

Moreover,  $G'/(G' \cap \Gamma)$  has a Mal'cev basis in which each element is a  $Q$ -rational combination of elements of the Mal'cev basis of  $G$ .

By splitting the interval  $[X, X + H]$  into arithmetic progressions of length  $\asymp H/Q^2$  and modulus equal to the period of  $\gamma$ , we conclude that for some  $Q^{O(1)}$ -rational elements  $\varepsilon_0, \gamma_0 \in G'$  we have

$$\left| \sum_{X < n \leq X+H} f(n) F(\varepsilon_0 g'(n) \gamma_0 \Gamma) \right|^* \gg \delta^{O(1)} H(\log X)^{k-1},$$

Let  $G'' = \gamma_0^{-1} G' \gamma_0$ ,  $\Gamma'' = \Gamma \cap G''$ ,  $g''(n) = \gamma_0^{-1} g'(n) \gamma_0$ , and let  $\tilde{F} : G''/\Gamma'' \rightarrow \mathbb{C}$  be given by  $\tilde{F}(x) = F(\varepsilon_0 \gamma_0 x)$ . Then we have

$$\left| \sum_{X < n \leq X+H} f(n) \tilde{F}(g''(n) \Gamma) \right|^* \gg \delta^{O(1)} H(\log X)^{k-1}. \tag{7.8}$$

By subtracting off the mean of  $\tilde{F}$ , we conclude that either (7.7) holds, or else (7.8) holds with  $\tilde{F}$  having mean 0. We may assume that the latter holds. Performing some nil-algebra as in [6, Proof of Claim in Sect. 2], we see that there is a Mal'cev basis  $\mathcal{Y}$  for  $G''/\Gamma''$ , which is a  $Q^{O(1)}$ -rational combination of the elements of the Mal'cev basis of  $G$ , such that  $g''$  is totally  $Q^{-cA+O(1)}$ -equidistributed on  $G''/\Gamma''$  with respect to the metric induced by  $\mathcal{Y}$  for some  $c > 0$  depending only on the fixed parameters, and additionally  $\tilde{F}$  has Lipschitz norm  $\ll \delta^{-O(1)}$ .

From the assumption on  $f$  and the triangle inequality, we have

$$\sum_{M < m \leq 2M} d_{k-1}(m) \left| \sum_{X/m < r \leq (X+H)/m} \beta(r) \tilde{F}(g''(rm) \Gamma) \right|^* \gg \delta^{O(1)} H(\log X)^{k-1}.$$

Since  $\|\beta\|_{\text{TV}(\mathbb{N})} \leq \frac{1}{\delta}$ , invoking Lemma 2.2(ii) we obtain

$$\sum_{M_1 < m_1 \leq 2M_1} d_{k-1}(m) \left| \sum_{X/m < r \leq (X+H)/m} \tilde{F}(g''(rm)\Gamma) \right|^* \gg \delta^{O(1)} H (\log X)^{k-1}.$$

By the pigeonhole principle, this implies that for some  $M_1, \dots, M_{k-1} \geq 1$  with  $M_1 \cdots M_{k-1} \asymp M$  we have

$$\sum_{M_1 < m_1 \leq 2M_1} \cdots \sum_{M_{k-1} < m_{k-1} \leq 2M_{k-1}} \left| \sum_{X/(m_1 \cdots m_{k-1}) < r \leq (X+H)/(m_1 \cdots m_{k-1})} \tilde{F}(g''(rm_1 \cdots m_{k-1})\Gamma) \right|^* \gg \delta^{O(1)} H.$$

Since  $|F| \leq 1/\delta$  by the assumption that  $F$  has Lipschitz norm at most  $1/\delta$ , from the pigeonhole principle we conclude that there exists a set  $\mathcal{M}$  of  $\gg \delta^{O(1)} M$  integer tuples  $(m_1, \dots, m_{k-1}) \in (M_1, 2M_1] \times \cdots \times (M_{k-1}, 2M_{k-1}]$  such that

$$\left| \sum_{X/(m_1 \cdots m_{k-1}) < r \leq (X+H)/(m_1 \cdots m_{k-1})} \tilde{F}(g''(rm_1 \cdots m_{k-1})\Gamma) \right|^* \gg \delta^{O(1)} \frac{H}{M_1 \cdots M_{k-1}}.$$

From the quantitative Leibman theorem (Lemma 2.9), we deduce that for  $(m_1, \dots, m_{k-1}) \in \mathcal{M}$  there exists some non-trivial horizontal character  $\eta: G'' \rightarrow \mathbb{R}$  (depending on the tuple) of Lipschitz norm  $\ll \delta^{-O(1)}$  such that

$$\|(\eta \circ g'')(m_1 \cdots m_{k-1} \cdot) \bmod 1\|_{C^\infty(\frac{X}{m_1 \cdots m_{k-1}}, \frac{X+H}{m_1 \cdots m_{k-1}})} \ll \delta^{-O(1)}. \tag{7.9}$$

By pigeonholing, we can find some such horizontal character  $\eta$  such that (7.9) holds for  $\gg \delta^{O(1)} M$  integer tuples  $(m_1, \dots, m_{k-1}) \in (M_1, 2M_1] \times \cdots \times (M_{k-1}, 2M_{k-1}]$ . Let  $\mathcal{M}'$  be the set formed by these tuples. Since  $\eta \circ g''$  is a polynomial of degree at most  $d$ , we can Taylor expand it as

$$(\eta \circ g'')(n) = \alpha_d(n - X)^d + \cdots + \alpha_1(n - X) + \alpha_0 \bmod 1$$

for some  $\alpha_j \in \mathbb{R}/\mathbb{Z}$ . Therefore, for any  $m \geq 1$  we have

$$(\eta \circ g'')(mn) = \sum_{j=0}^d \alpha_j m^j \left(n - \frac{X}{m}\right)^j \bmod 1.$$

Now (7.9) and (2.2) give for any  $(m_1, \dots, m_{k-1}) \in \mathcal{M}'$  the bound

$$\max_{1 \leq j \leq d} \left(\frac{H}{M_1 \cdots M_{k-1}}\right)^j \|\alpha_j(m_1 \cdots m_{k-1})^j\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}. \tag{7.10}$$

In particular, there exist  $\gg \delta^{O(1)} M_1$  integers  $m_1 \in (M_1, 2M_1]$  for which (7.10) holds for  $\gg \delta^{O(1)} M_2 \cdots M_{k-1}$  integer tuples  $(m_2, \dots, m_{k-1}) \in (M_2, 2M_2] \times \cdots \times$

$(M_{k-1}, 2M_{k-1}]$ . By Vinogradov’s lemma (Lemma 2.3), this implies that for some integer  $1 \leq q_1 \ll \delta^{-O(1)}$  we have

$$\max_{1 \leq j \leq d} \left( \frac{H}{M_2 \cdots M_{k-1}} \right)^j \|q_1 \alpha_j(m_2 \cdots m_{k-1})^j\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}$$

for  $\gg \delta^{O(1)} M_2 \cdots M_{k-1}$  integer tuples  $(m_2, \dots, m_{k-1}) \in (M_2, 2M_2] \times \cdots \times (M_{k-1}, 2M_{k-1}]$  (by pigeonholing we can take  $q_1$  to be independent of the tuple). Continuing inductively, we see that for some positive integer  $q \ll \delta^{-O(1)}$  we have

$$\max_{1 \leq j \leq d} H^j \|q \alpha_j\|_{\mathbb{R}/\mathbb{Z}} \ll \delta^{-O(1)}.$$

But by the definition of the smoothness norms, this implies that

$$\|(q\eta) \circ g'' \bmod 1\|_{C^\infty(X, X+H)} \ll \delta^{-O(1)}.$$

The horizontal character  $q\eta$  has Lipschitz norm  $\ll \delta^{-O(1)}$ . As noted above,  $g''$  is totally  $Q^{-cA+O(1)}$ -equidistributed on  $G''/\Gamma''$ . But now we obtain a contradiction from the converse to the quantitative Leibman theorem [3, Lemma 5.3] and the assumption that  $A$  was large enough. □

We are now ready to prove Theorem 1.3(iii).

**Proof of Theorem 1.3(iii)** We apply Theorem 1.1(v). Following the same argument as in the proof of Theorem 1.3(i), but with

$$f(n) = \left( \frac{W}{\varphi(W)} \right)^{k-1} (\log^{1-k} X)(d_k(n) - d_k^\sharp(n))$$

and using Lemma 7.5(ii) in place of Lemma 7.5(i), we see that there is some exceptional set  $\mathcal{E}$  of measure  $\ll_A X/\log^A X$  such that for all  $x \in [X(\log X)^{-A/2}, 2X] \setminus \mathcal{E}$  we have

$$\left\| \left( \frac{W}{\varphi(W)} \right)^{k-1} (\log^{1-k} X)(d_k(W' \cdot + a) - d_k^\sharp(W' \cdot + a)) \right\|_{U^s(x, x+H)} = o_{w \rightarrow \infty}(1). \tag{7.11}$$

Arguing as in [30, Proof of Theorem 1.5] this implies that

$$\|d_k - d_k^\sharp\|_{U^s(x, x+H)} = o_{X \rightarrow \infty}(\log^{k-1} X)$$

for all  $x \in [X, 2X] \setminus \mathcal{E}'$ , where  $\mathcal{E}' \supset \mathcal{E}$  is another exceptional set of measure  $\ll_A X/\log^A X$ . We still need to show that the second estimate in (1.10) holds outside a small exceptional set.

In view of (7.11) and the triangle inequality for the Gowers norms, we are left with showing that

$$\left\| \left( \frac{W}{\varphi(W)} \right)^{k-1} (\log^{1-k} X) d_k^\sharp(W' \cdot + a) - \frac{1}{(k-1)!} \right\|_{U^s(x, x+H)} = o_{w \rightarrow \infty}(1) \tag{7.12}$$

for all  $x \in [X, 2X] \setminus \mathcal{E}''$ , where  $\mathcal{E}'' \supset \mathcal{E}'$  is another exceptional set of measure  $\ll_A X/\log^A X$ .

By Lemma 7.5(ii) and Proposition 7.3, it suffices to show that, for any filtered  $(s - 1)$ -step nilmanifold  $G/\Gamma$  and any Lipschitz function  $F: G/\Gamma \rightarrow \mathbb{C}$ , we have

$$\begin{aligned} & \sup_{g \in \text{Poly}(\mathbb{Z} \rightarrow G)} \left| \sum_{x < n \leq x+H} \left( \left( \frac{W}{\varphi(W)} \right)^{k-1} (\log^{1-k} X) d_k^\sharp(W'n + a) - \frac{1}{(k-1)!} \right) \right. \\ & \quad \left. \times \overline{F}(g(n)\Gamma) \right| \\ & = o_{w \rightarrow \infty; \|F\|_{\text{Lip}; G/\Gamma}(H)} \end{aligned}$$

for all  $x \in [X, 2X]$  outside of an exceptional set of measure  $\ll_A X/\log^A X$ .

By Lemma 2.19(ii), the function inside the sum is a sum of  $O(1)$  terms of the form  $\alpha_j * \beta_j$ , where  $\|\beta_j\|_{\text{TV}(\mathbb{N})} \ll 1$  and  $|\alpha_j(n)| \leq d_{k-1}(n) 1_{[1, X^{\epsilon/5}]}(n)$ . Hence, from Proposition 7.6 we conclude that (7.12) holds provided that

$$\left| \sum_{x < n \leq x+H} \left( \left( \frac{W}{\varphi(W)} \right)^{k-1} (\log^{1-k} X) d_k^\sharp(W'n + a) - \frac{1}{(k-1)!} \right) \right|^* = o_{w \rightarrow \infty}(H).$$

From (7.11) with  $s = 1$ , it follows that

$$\begin{aligned} & \left| \sum_{x < n \leq x+H} \left( \left( \frac{W}{\varphi(W)} \right)^{k-1} (\log^{1-k} X) (d_k(W'n + a) - d_k^\sharp(W'n + a)) \right) \right|^* \\ & = o_{X \rightarrow \infty}(H). \end{aligned}$$

for all  $x \in [X, 2X] \setminus \mathcal{E}$ . Hence, it suffices to show that

$$\left| \sum_{x < n \leq x+H} \left( \left( \frac{W}{\varphi(W)} \right)^{k-1} (\log^{1-k} X) d_k(W'n + a) - \frac{1}{(k-1)!} \right) \right|^* = o_{X \rightarrow \infty}(H)$$

for all  $x \in [X, 2X]$  outside of an exceptional set of measure  $\ll X/\log^A X$ . But since  $W' \ll \log \log X$  (say) this follows from Lemma 4.6.  $\square$

## 8 Applications

### 8.1 The polynomial phase case

**Proof of Corollary 1.2** Part (i) of the corollary is immediate from Theorem 1.1(i). From Theorem 1.1(ii), we either have the conclusion of part (ii) or else

$$\left| \sum_{x < n \leq x+H} \Lambda^\sharp(n) e(P_x(n)) \right| \geq \frac{H}{\log^A X}$$

for  $x \in E$ , with  $E$  a subset of  $[X, 2X]$  of measure  $\gg_{A,d,\varepsilon} X(\log X)^{-A}$ . By Lemma 2.19(i), this implies that for some sequence  $a: \mathbb{N} \rightarrow \mathbb{C}$  supported on  $[1, X^{\varepsilon/5}]$  and satisfying  $|a(n)| \ll \log X$  we have

$$\left| \sum_{x < n \leq x+H} (a * 1)(n) e(P_x(n)) \right| \geq \frac{H}{2 \log^A X}$$

for all  $x \in E$ . Applying Theorem 4.2(i) (and Remark 2.8), we conclude that, for some integer  $1 \leq q \ll (\log X)^{O_{A,\varepsilon,d}(1)}$ , we have

$$\|q P_x \pmod{1}\|_{C^\infty(x, x+H]} \ll_d (\log X)^{O_{A,\varepsilon,d}(1)}$$

for all  $x$  in a subset of  $E$  of measure  $\gg_{A,d,\varepsilon} X \log^{-A} X$ . The desired conclusion follows from (2.2). □

### 8.2 Calculating the main terms

In order to prove Theorems 1.5 and 1.6, we must in particular evaluate the correlations of  $\Lambda_w$  and  $d_{k,W} := \prod_{p \leq w} d_{k,p}$  (where  $d_{k,p}$  is defined in Theorem 1.6), since these yield the main terms in the correlations asymptotics for  $\Lambda$  and  $d_k$ , respectively. We do this in a more general setting as follows.

**Proposition 8.1** (Factorizing correlations of multiplicative functions) *Let  $k, \ell \in \mathbb{N}$  be fixed. Let  $3 \leq w \leq \exp((\log X)^{1/2})$  and  $W = \prod_{p \leq w} p$ . Let  $1 \leq h \leq X$  be an integer. Let  $g: \mathbb{N} \rightarrow \mathbb{C}$  be a multiplicative function, and denote by  $g_p$  the multiplicative function defined on prime powers by  $g_p(p^a) = g(p^a)$  and  $g_p(q^a) = 1$  if  $q \neq p$ . Define  $g_W := \prod_{p \leq w} g_p$ . Suppose that  $|g_p(p^a)| \leq d_k(p^a)$  for all  $p \leq w$  and  $a \in \mathbb{N}$ . Then we have*

$$\begin{aligned} & \mathbb{E}_{n \leq X} \prod_{j=0}^{\ell-1} g_W(n + jh) \\ &= \prod_{p \leq w} \left( \lim_{Y \rightarrow \infty} \mathbb{E}_{n \leq Y} \prod_{j=0}^{\ell-1} g_p(n + jh) \right) + O(\exp(-(\log X)^{1/2}/(30\ell))). \end{aligned}$$

**Proof** Let  $\mathcal{B}$  be the set of  $n \leq X$  for which  $(n, W^\infty) > X^{\frac{1}{10\ell}}$ . Since every  $n \in \mathcal{B}$  can be written uniquely as  $e_1 e_2$ , where  $e_1 \mid W^\infty$  with  $e_1 > X^{\frac{1}{10\ell}}$  and  $(e_2, W) = 1$ , we have

$$\begin{aligned} |\mathcal{B}| &= \sum_{\substack{e_1 \mid W^\infty \\ e_1 > X^{\frac{1}{10\ell}}}} \sum_{\substack{(e_2, W)=1 \\ e_2 \leq X/e_1}} 1 \ll X \sum_{\substack{e_1 \mid W^\infty \\ e_1 > X^{\frac{1}{10\ell}}}} \frac{1}{e_1} \\ &\ll X^{1 - \frac{1}{10\ell \log w}} \sum_{\substack{e_1 \mid W^\infty \\ e_1 > X^{\frac{1}{10\ell}}}} \frac{1}{e_1^{1 - \frac{1}{\log w}}} \end{aligned}$$

$$\begin{aligned}
 &\ll X^{1-\frac{1}{10\ell \log w}} \prod_{p \leq w} \left( 1 + \frac{1}{p^{1-\frac{1}{\log w}}} + O\left(\frac{1}{p^{2(1-\frac{1}{\log w})}}\right) \right) \tag{8.1} \\
 &\ll X^{1-\frac{1}{10\ell \log w}} \prod_{p \leq w} \left( 1 + O\left(\frac{1}{p}\right) \right) \\
 &\ll X^{1-\frac{1}{10\ell \log w}} (\log w)^{O(1)}.
 \end{aligned}$$

By Möbius inversion we can write

$$g_W(n) = g((n, W^\infty)) = \sum_{e|(n, W^\infty)} \psi(e),$$

where  $\psi := \mu * g$  is multiplicative. For any prime power  $p^a$  we have  $\psi(p^a) = g_W(p^a) - g_W(p^{a-1})$ . In particular, this implies that

$$|\psi(p^a)| \leq 2d_k(p^a) \cdot 1_{p \leq w}. \tag{8.2}$$

Let us also define the truncated version

$$\tilde{g}_W(n) := \sum_{\substack{e|(n, W^\infty) \\ e \leq X^{\frac{1}{10\ell}}}} \psi(e).$$

For  $n \notin \mathcal{B}$  we have  $g_W(n) = \tilde{g}_W(n)$ . Hence, by the triangle inequality, for some  $i_0 \in \{0, 1, \dots, \ell - 1\}$ , we have

$$\begin{aligned}
 &\sum_{n \leq X} \prod_{j=0}^{\ell-1} g_W(n + jh) \\
 &= \sum_{\substack{n \leq X \\ n+ih \notin \mathcal{B} \forall 0 \leq i \leq \ell-1}} \prod_{j=0}^{\ell-1} \tilde{g}_W(n + jh) + O\left( \max_{0 \leq i \leq \ell-1} \sum_{\substack{n \leq X \\ n+ih \in \mathcal{B}}} \prod_{j=0}^{\ell-1} |g_W(n + jh)| \right) \tag{8.3} \\
 &= \sum_{n \leq X} \prod_{j=0}^{\ell-1} \tilde{g}_W(n + jh) + O\left( \sum_{\substack{n \leq X \\ n+i_0h \in \mathcal{B}}} \prod_{j=0}^{\ell-1} (|g_W| + |\tilde{g}_W|)(n + jh) \right).
 \end{aligned}$$

Since  $|g_W(n)|, |\tilde{g}_W(n)| \leq d_{k+2}(n)$ , the error term here is by Hölder’s inequality (with exponents  $1/2, 1/(2\ell), \dots, 1/(2\ell)$ ), (8.1) and Shiu’s bound (Lemma 2.18)

$$\leq \left( \sum_{\substack{n \leq X \\ n+i_0h \in \mathcal{B}}} 1 \right)^{1/2} \left( \sum_{n \leq X+(\ell-1)h} d_{k+2}^{2\ell} \right)^{1/2}$$

$$\begin{aligned} &\ll X(\log X)^{O(1)} \exp\left(-\frac{\log X}{20\ell \log w}\right) \\ &\ll X \exp\left(-\frac{(\log X)^{1/2}}{30\ell}\right). \end{aligned}$$

We are now left with estimating the main term in (8.3). By the definition of  $\tilde{g}_W$ , this can be expanded out as

$$\sum_{\substack{e_0, \dots, e_{\ell-1} | W^\infty \\ e_0, \dots, e_{\ell-1} \leq X^{\frac{1}{10\ell}}}} \psi(e_0) \cdots \psi(e_{\ell-1}) \sum_{\substack{n \leq X \\ e_i | n+ih \forall 0 \leq i \leq \ell-1}} 1. \tag{8.4}$$

Let

$$\alpha_h(e_0, \dots, e_{\ell-1}) := \lim_{Y \rightarrow \infty} \mathbb{E}_{n \leq Y} \prod_{j=0}^{\ell-1} 1_{e_j | n+jh}. \tag{8.5}$$

Then we have

$$\sum_{\substack{n \leq X \\ e_i | n+ih \forall 0 \leq i \leq \ell-1}} 1 = \alpha_h(e_0, \dots, e_{\ell-1})X + O(1).$$

Now, recalling (8.2), which implies that  $|\psi(n)| \ll n^{o(1)}$ , the expression (8.4) is

$$X \sum_{\substack{e_0, \dots, e_{\ell-1} | W^\infty \\ e_0, \dots, e_{\ell-1} \leq X^{\frac{1}{10\ell}}}} \psi(e_0) \cdots \psi(e_{\ell-1}) \alpha_h(e_0, \dots, e_{\ell-1}) + O(X^{1/10+o(1)}). \tag{8.6}$$

By inclusion–exclusion, the main term in (8.6) is, for some  $I \subset \{0, 1, \dots, \ell - 1\}$ ,

$$\begin{aligned} &X \sum_{e_0, \dots, e_{\ell-1} | W^\infty} \psi(e_0) \cdots \psi(e_{\ell-1}) \alpha_h(e_0, \dots, e_{\ell-1}) \\ &+ O\left( X \sum_{\substack{e_0, \dots, e_{\ell-1} | W^\infty \\ e_i > X^{\frac{1}{10\ell}} \forall i \in I}} \psi(e_0) \cdots \psi(e_{\ell-1}) \alpha_h(e_0, \dots, e_{\ell-1}) \right). \end{aligned} \tag{8.7}$$

We first bound the error term here. Note that, for any  $e_j \in \mathbb{N}$ ,

$$\alpha_h(e_0, \dots, e_{\ell-1}) \leq \frac{1}{[e_0, \dots, e_{\ell-1}]}.$$

Hence, for some  $i_0 \in \{0, \dots, \ell - 1\}$ , the error term in (8.7) is

$$\begin{aligned} &\ll X \sum_{\substack{e_0, \dots, e_{\ell-1} | W^\infty \\ e_{i_0} > X^{\frac{1}{10\ell}}}} \frac{|\psi(e_0)| \cdots |\psi(e_{\ell-1})|}{[e_0, \dots, e_{\ell-1}]} \\ &\ll X^{1 - \frac{1}{10\ell \log w}} \sum_{e_0, \dots, e_{\ell-1} | W^\infty} \frac{|\psi(e_0)| \cdots |\psi(e_{\ell-1})|}{[e_0, \dots, e_{\ell-1}]} e_{i_0}^{\frac{1}{\log w}} = X^{1 - \frac{1}{10\ell \log w}} \prod_{p \leq w} \beta_p, \end{aligned}$$

where

$$\begin{aligned} \beta_p &= \sum_{e_0, \dots, e_{\ell-1} | p^\infty} \frac{|\psi(e_0)| \cdots |\psi(e_{\ell-1})|}{[e_0, \dots, e_{\ell-1}]} e_{i_0}^{\frac{1}{\log w}} \\ &= \sum_{E | p^\infty} \frac{1}{E} \sum_{\substack{e_0, \dots, e_{\ell-1} \\ [e_0, \dots, e_{\ell-1}] = E}} |\psi(e_0)| \cdots |\psi(e_{\ell-1})| e_{i_0}^{\frac{1}{\log w}}. \end{aligned}$$

For  $E = p^a$  with  $a \geq 1$ , by (8.2) the inner sum above is

$$\ll (2d_k(E))^\ell E^{\frac{1}{\log w}} \sum_{\substack{e_0, \dots, e_{\ell-1} \\ [e_0, \dots, e_{\ell-1}] = E}} 1 \ll (2(a+1)^k)^\ell (a+1)^\ell p^{\frac{a}{\log w}}.$$

It follows that

$$\beta_p = 1 + O\left(\sum_{a \geq 1} \frac{(2(a+1))^{(k+1)\ell}}{(p^a)^{1 - \frac{1}{\log w}}}\right) = 1 + O\left(\frac{1}{p}\right)$$

for  $p \leq w$ . Hence the error term in (8.7) is

$$\begin{aligned} &\ll X^{1 - \frac{1}{10\ell \log w}} \prod_{p \leq w} \left(1 + O\left(\frac{1}{p}\right)\right) \\ &\ll X^{1 - \frac{1}{10\ell \log w}} (\log w)^{O(1)} \ll X \exp(-(\log X)^{1/2} / (30\ell)). \end{aligned}$$

The remaining task is to evaluate the main term in (8.7). Since the  $\ell$ -variable function  $\alpha_h(e_0, \dots, e_{\ell-1})$  is multiplicative, we have

$$\begin{aligned} &\sum_{e_0, \dots, e_{\ell-1} | W^\infty} \psi(e_0) \cdots \psi(e_{\ell-1}) \alpha_h(e_0, \dots, e_{\ell-1}) \\ &= \prod_{p \leq w} \sum_{e_0, \dots, e_{\ell-1} | p^\infty} \psi(e_0) \cdots \psi(e_{\ell-1}) \alpha_h(e_0, \dots, e_{\ell-1}). \end{aligned}$$

For each prime  $p \leq w$ , by (8.5) we can write

$$\sum_{e_0, \dots, e_{\ell-1} | p^\infty} \psi(e_0) \cdots \psi(e_{\ell-1}) \alpha_h(e_0, \dots, e_{\ell-1})$$

$$= \sum_{e_0, \dots, e_{\ell-1} | p^\infty} \psi(e_0) \cdots \psi(e_{\ell-1}) \lim_{Y \rightarrow \infty} \mathbb{E}_{n \leq Y} \prod_{j=0}^{\ell-1} 1_{e_j | n+jh}.$$

By the dominated convergence theorem together with the convergence of

$$\sum_{e_0, \dots, e_{\ell-1} | p^\infty} \frac{|\psi(e_0)| \cdots |\psi(e_{\ell-1})|}{[e_0, \dots, e_{\ell-1}]}$$

we can exchange the order of limits and summation. Thus

$$\begin{aligned} & \sum_{e_0, \dots, e_{\ell-1} | p^\infty} \psi(e_0) \cdots \psi(e_{\ell-1}) \alpha_h(e_0, \dots, e_{\ell-1}) \\ &= \lim_{Y \rightarrow \infty} \mathbb{E}_{n \leq Y} \sum_{e_0, \dots, e_{\ell-1} | p^\infty} \prod_{j=0}^{\ell-1} \psi(e_j) 1_{e_j | n+jh} \\ &= \lim_{Y \rightarrow \infty} \mathbb{E}_{n \leq Y} \prod_{j=0}^{\ell-1} g_p(n+jh), \end{aligned}$$

and the claim follows.  $\square$

Applying Proposition 8.1 to the primes or the higher order divisor function, we obtain the following corollary. Here and later we write

$$d_{k,W}(n) := \prod_{p \leq w} d_{k,p}(n)$$

with  $d_{k,p}$  as in Theorem 1.6.

**Corollary 8.2** *Let  $k, \ell \in \mathbb{N}$  be fixed. Let  $\ell < w \leq \exp((\log X)^{1/2})$ ,  $W = \prod_{p \leq w} p$  and  $H \geq 2$ .*

(i) *For  $(1 - O(w^{-1/2}))H$  values of  $1 \leq h \leq H$  we have*

$$\begin{aligned} & \mathbb{E}_{n \leq X} \Lambda_w(n) \Lambda_w(n+h) \cdots \Lambda_w(n+(\ell-1)h) \\ &= (1 + o_{w \rightarrow \infty}(1)) \mathfrak{S}(0, h, \dots, (\ell-1)h), \end{aligned}$$

where  $\mathfrak{S}(h_1, \dots, h_\ell)$  is the singular series defined in (1.11).

(ii) *Let  $H \geq 1$ . For  $(1 - O(w^{-1/2}))H$  values of  $1 \leq h \leq H$  we have*

$$\begin{aligned} & \mathbb{E}_{n \leq X} d_{k,W}(n) d_{k,W}(n+h) \cdots d_{k,W}(n+(\ell-1)h) \\ &= \left( \frac{W}{\varphi(W)} \right)^{k\ell-\ell} (C(h, k, \ell) + o_{w \rightarrow \infty}(1)), \end{aligned}$$

where  $C(h, k, \ell)$  is as in (1.12).

**Proof** For part (i), we apply Proposition 8.1 with  $g = 1_{(\cdot, P(X^{1/2}))=1}$ . This gives

$$\begin{aligned} & \mathbb{E}_{n \leq X} \prod_{j=0}^{\ell-1} \Lambda_w(n + jh) \\ &= \left( \frac{W}{\varphi(W)} \right)^\ell \prod_{p \leq w} \lim_{Y \rightarrow \infty} \mathbb{E}_{n \leq Y} \prod_{j=0}^{\ell-1} 1_{(n+jh, p)=1} + O(\exp(-(\log X)^{1/2}/(30\ell))). \end{aligned}$$

The limit inside the product over  $p$  equals to  $1 - \nu_p/p$ , where  $\nu_p$  is the size of the set  $\{0, h, \dots, (\ell - 1)h\} \pmod p$ . Now the claim follows by noting that

$$\prod_{p > w} \left(1 - \frac{1}{p}\right)^{-\ell} \left(1 - \frac{\nu_p}{p}\right) = \exp\left(O\left(\sum_{p > w} \frac{1}{p^2} + \frac{1_{p|h}}{p}\right)\right) = 1 + O(w^{-1/2})$$

unless

$$\sum_{\substack{p|h \\ p > w}} \frac{1}{p} > w^{-1/2}.$$

But by Markov’s inequality the number of such integers  $1 \leq h \leq H$  is

$$\leq w^{1/2} \sum_{h \leq H} \sum_{\substack{p|h \\ p > w}} \frac{1}{p} \ll w^{1/2} \sum_{p > w} \frac{H}{p^2} \ll Hw^{-1/2},$$

so they can be included in the exceptional set.

Let us now turn to part (ii). We have  $d_{k,p} = 1_p * d_{k-1,p}$ , where  $1_p$  is the multiplicative function such that  $1_p(p^\alpha) = 1$  for every  $\alpha \geq 0$  and  $1_p(q^\alpha) = 0$  for every prime  $q \neq p$  and every  $\alpha \geq 1$ . Thus we have

$$\begin{aligned} \lim_{Y \rightarrow \infty} \mathbb{E}_{n \leq Y} d_{k,p}(n) &= \lim_{Y \rightarrow \infty} \sum_{a \geq 0} d_{k-1,p}(p^a) \frac{1}{Y} \sum_{\substack{n \leq Y \\ p^a | n}} 1 = \sum_{a \geq 0} \frac{1}{p^a} \binom{k-2+a}{a} \\ &= \sum_{a \geq 0} \binom{1-k}{a} \left(\frac{-1}{p}\right)^a = \left(1 - \frac{1}{p}\right)^{1-k}, \end{aligned}$$

so that

$$\prod_{p \leq w} \lim_{Y \rightarrow \infty} \mathbb{E}_{n \leq Y} d_{k,p}(n) = \left(\frac{W}{\varphi(W)}\right)^{k-1}.$$

Now we apply Proposition 8.1 with  $g = d_k$ , obtaining

$$\mathbb{E}_{n \leq X} \prod_{j=0}^{\ell-1} d_{k,w}(n + jh)$$

$$= \left(\frac{W}{\varphi(W)}\right)^{(k-1)\ell} \prod_{p \leq w} C_p(h, k, \ell) + O(\exp(-(\log X)^{1/2}/(30\ell))), \tag{8.8}$$

where

$$C_p(h, k, \ell) := \lim_{Y \rightarrow \infty} \frac{\mathbb{E}_{n \leq Y} \prod_{j=0}^{\ell-1} d_{k,p}(n + jh)}{(\mathbb{E}_{n \leq Y} d_{k,p}(n))^\ell}.$$

Thus it suffices to show that  $\prod_{p > w} C_p(h, k, \ell) = 1 + o_{w \rightarrow \infty}(1)$  for all but at most  $O(w^{-1/2}H)$  values of  $1 \leq h \leq H$ .

Since  $d_{k,p} = 1_p * d_{k-1,p}$ , we can write

$$\begin{aligned} & \mathbb{E}_{n \leq Y} \prod_{j=0}^{\ell-1} d_{k,p}(n + jh) \\ &= \sum_{a_0, \dots, a_{\ell-1} \geq 0} d_{k-1,p}(p^{a_0}) \cdots d_{k-1,p}(p^{a_{\ell-1}}) \cdot \frac{1}{Y} \sum_{\substack{n \leq Y \\ p^a | n + jh \ \forall j \in [0, \ell-1]}} 1. \end{aligned}$$

Thus we have the estimate

$$\begin{aligned} & \lim_{Y \rightarrow \infty} \mathbb{E}_{n \leq Y} \prod_{j=0}^{\ell-1} d_{k,p}(n + jh) \\ &= 1 + O\left(\sum_{a \geq 1} \sum_{\substack{a_0, \dots, a_{\ell-1} \geq 0 \\ \max\{a_0, \dots, a_{\ell-1}\} = a}} d_{k-1,p}(p^{a_0}) \cdots d_{k-1,p}(p^{a_{\ell-1}}) p^{-a}\right) \\ &= 1 + O\left(\sum_{a \geq 1} a^{O(1)} p^{-a}\right) = 1 + O\left(\frac{1}{p}\right), \end{aligned} \tag{8.9}$$

and for any prime  $p > \ell$  such that  $p \nmid h$  we can compute more precisely

$$\begin{aligned} & \lim_{Y \rightarrow \infty} \mathbb{E}_{n \leq Y} \prod_{j=0}^{\ell-1} d_{k,p}(n + jh) \\ &= \sum_{a \geq 0} \sum_{\substack{a_0, \dots, a_{\ell-1} \geq 0 \\ \{a_0, \dots, a_{\ell-1}\} = \{a, 0, \dots, 0\}}} d_{k-1,p}(p^{a_0}) \cdots d_{k-1,p}(p^{a_{\ell-1}}) p^{-a} \\ &= 1 + \frac{(k-1)\ell}{p} + O\left(\frac{1}{p^2}\right). \end{aligned} \tag{8.10}$$

A special case implies that

$$\left(\lim_{Y \rightarrow \infty} \mathbb{E}_{n \leq Y} d_{k,p}(n)\right)^\ell = 1 + \frac{(k-1)\ell}{p} + O\left(\frac{1}{p^2}\right). \tag{8.11}$$

It follows from (8.9) and (8.11) that  $C_p(h, k, \ell) = 1 + O(1/p)$ , and from (8.10) and (8.11) that  $C_p(h, k, \ell) = 1 + O(1/p^2)$  for  $p \nmid h$ . Hence

$$\begin{aligned} \prod_{p > w} C_p(h, k, \ell) &= \prod_{\substack{p > w \\ p \nmid h}} C_p(h, k, \ell) \cdot \prod_{\substack{p > w \\ p \nmid h}} \left(1 + O\left(\frac{1}{p^2}\right)\right) \\ &= \exp\left(O\left(w^{-1} + \sum_{\substack{p > w \\ p \nmid h}} \frac{1}{p}\right)\right). \end{aligned}$$

As in part (i), this is  $1 + O(w^{-1/2})$  for all but  $\ll Hw^{-1/2}$  integers  $1 \leq h \leq H$ , and such integers can be included in the exceptional set. Now the claim follows from (8.8).  $\square$

We shall also need an upper bound of the correct order of magnitude for correlations of multiplicative functions.

**Lemma 8.3** (An upper bound for correlations of multiplicative functions) *Let  $\ell \in \mathbb{N}$  be fixed. For  $1 \leq j \leq \ell$ , let  $1 \leq a_j, h_j \leq \ell X$  be integers with  $(a_j, h_j) = 1$  and  $a_i h_j \neq a_j h_i$  for all  $i \neq j$ . Let  $f_1, \dots, f_\ell: \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$  be multiplicative functions with  $f_j(ab) \leq f_j(a) f_j(b)$  for all  $1 \leq j \leq \ell$  and  $a, b \in \mathbb{N}$ . Suppose that  $f_j(n) \ll n^{o(1)}$  and that  $f_j(p^r) \leq e^{O(r)}$  for all  $1 \leq j \leq \ell$ , all primes  $p$  and all  $r \in \mathbb{N}$ . Then we have*

$$\begin{aligned} &\sum_{X^{1/2} \leq n \leq X} \prod_{j=1}^{\ell} f_j(a_j n + h_j) \\ &\ll_{\Delta_D X} \prod_{\ell < p \leq X^{1/2}} \left(1 - \frac{|\{t \pmod p : \prod_{j=1}^{\ell} (a_j t + h_j) \equiv 0 \pmod p\}|}{p}\right) \\ &\quad \times \prod_{j=1}^{\ell} \sum_{\substack{n \leq X^{1/2} \\ (n, D)=1}} \frac{f_j(n)}{n}, \end{aligned}$$

where  $D = \prod_{1 \leq i < j \leq \ell} (a_i h_j - a_j h_i)^2$  and

$$\Delta_D = \prod_{p \mid D} \left(1 + \sum_{\substack{0 \leq v_1, \dots, v_\ell \leq 1 \\ (v_1, \dots, v_\ell) \neq (0, \dots, 0)}} \frac{|\{n \pmod{p^2} : p \mid a_j n + h_j \forall 1 \leq j \leq \ell\}|}{p^2}\right)$$

$$\times \prod_{j=1}^{\ell} f_j(p^{v_j}) \Bigg).$$

In particular, we have

$$\Delta_D \ll \prod_{p|D} \left(1 + O\left(\frac{1}{p}\right)\right). \tag{8.12}$$

**Proof** We will apply Henriot’s upper bound for correlations of multiplicative functions [13, Corollary 1] with  $x = X^{1/2}$ ,  $y = X$ ,  $\delta = 1/2$ ,  $Q_j(m) = a_j m + h_j$ ,  $Q(m) = \prod_{j=1}^{\ell} Q_j(m)$  and  $F(n_1, \dots, n_{\ell}) = \prod_{j=1}^{\ell} f_j(n)$ . Note that the discriminant of  $Q$  is  $\prod_{1 \leq i < j \leq \ell} (a_i h_j - a_j h_i)^2 = D$ , and that the sum of absolute values of coefficients  $\|Q\|$  of  $Q$  is  $\ll X^{\ell}$ . Note also that by Gauss’ lemma  $Q$  is primitive in the sense that the greatest common divisor of its coefficients is 1. Now the claim follows from [13, Corollary 1] (noting, as in the end of [13, Sect. 3], that  $\tilde{F} = \tilde{G} = F$  in that corollary) and the trivial inequality

$$\sum_{n_1 \cdots n_{\ell} \leq x} a_{n_1,1} \cdots a_{n_{\ell},\ell} \leq \prod_{j=1}^{\ell} \sum_{n \leq x} a_{n,j}$$

for  $a_{n,j} \geq 0$ .<sup>14</sup> □

### 8.3 Proofs of the correlation asymptotics

The proofs of Theorems 1.5 and 1.6 require the Gowers uniformity estimates from Theorem 1.3 together with the generalized von Neumann theorem which we state in the following form.

**Lemma 8.4** (Generalized von Neumann theorem) *Let  $s, d, t, L \geq 1$  be given, and let  $D$  be sufficiently large depending on  $s, d, t, L$ . 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  $|f_i(x)| \leq v(x)$  for all  $i \in [t]$  and  $x \in [N]$ . Let  $\Psi = (\psi_1, \dots, \psi_t)$  be a system of affine-linear forms of complexity  $s$  with integer coefficients such that all the linear coefficients of  $\psi_i$  are bounded by  $L$  in modulus, and such that  $|\psi_i(0)| \leq DN$ . 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

$$\sum_{\mathbf{n} \in K} \prod_{i=1}^t f_i(\psi_i(\mathbf{n})) = o_{\delta \rightarrow 0}(N^d).$$

<sup>14</sup>Note that [13, Theorem 3] contains the assumption that  $Q$  has no fixed prime divisor, which is not necessarily satisfied in our case, but [13, Corollary 1] does not assume this.

**Proof** If  $\Psi$  is in  $s$ -normal form for some  $s$ , this follows from [30, Lemma 10.1]. In general,<sup>15</sup> we can first apply [5, Lemma 4.5] to lift to the  $s$ -normal form case and then argue exactly as in the end of [5, §4].  $\square$

**Proof of Theorems 1.5 and 1.6** Let  $k, \ell \in \mathbb{N}$  and  $\varepsilon > 0$  be fixed. Let  $f \in \{\Lambda, d_k\}$  and  $X^{c_f + \varepsilon} \leq H \leq X^{1 - \varepsilon}$ , where  $c_f = 1/3$  if  $f = \Lambda$  and  $c_f = 0$  if  $f = d_k$ . Let  $X \geq 3$  and let  $w \in \mathbb{N}$  satisfy  $2 \leq w \leq w(X)$ , where  $w(X)$  is some function tending to infinity slowly enough. Also let  $W = \prod_{p < w} p$ . Denote

$$f_w = \begin{cases} \Lambda_w & \text{if } f = \Lambda \\ \left(\frac{\varphi(W)}{W}\right)^{k-1} \frac{\log^{k-1} X}{(k-1)!} \cdot d_{k,W} & \text{if } f = d_k. \end{cases}$$

**Step 1: Splitting into main term and error term.** Writing  $f = f_w + (f - f_w)$ , we can split

$$\begin{aligned} \sum_{n \leq X} \prod_{j=0}^{\ell-1} f(n + jh) &= \sum_{n \leq X} \prod_{j=0}^{\ell-1} f_w(n + jh) + O\left(\max_{\substack{\mathcal{I} \subset \{0, \dots, \ell-1\} \\ \mathcal{I} \neq \emptyset}} |\mathcal{E}_{\mathcal{I}}(h)|\right), \quad \text{where} \\ \mathcal{E}_{\mathcal{I}}(h) &= \sum_{n \leq X} \prod_{j \in \mathcal{I}} (f(n + jh) - f_w(n + jh)) \prod_{j \in \{0, \dots, \ell-1\} \setminus \mathcal{I}} f_w(n + jh). \end{aligned} \tag{8.13}$$

By Corollary 8.2, for  $(1 - o_{w \rightarrow \infty}(1))H$  integers  $1 \leq h \leq H$ , the main term on the right-hand side of (8.13) produces the desired main term for Theorems 1.5 and 1.6. Therefore, we are left with bounding the error term in (8.13).

By Markov’s inequality, for either  $f \in \{\Lambda, d_k\}$  and any  $\eta > 0$ , we have

$$\begin{aligned} &\left| \left\{ h \leq H : |\mathcal{E}_{\mathcal{I}}(h)| \geq \eta \sum_{n \leq X} \prod_{j=0}^{\ell-1} f_w(n + jh) \right\} \right| \\ &\leq \frac{1}{\eta \sum_{n \leq X} \prod_{j=0}^{\ell-1} f_w(n + jh)} \sum_{h \leq H} |\mathcal{E}_{\mathcal{I}}(h)|. \end{aligned}$$

Hence, to complete the proof it suffices to show that

$$\sum_{h \leq H} |\mathcal{E}_{\mathcal{I}}(h)| = o_{w \rightarrow \infty} \left( H \sum_{n \leq X} \prod_{j=0}^{\ell-1} f_w(n + jh) \right). \tag{8.14}$$

**Step 2: Bounding the error term.** By Corollary 8.2, proving (8.14) reduces to showing that, uniformly for all unimodular sequences  $c: \mathbb{N} \rightarrow \mathbb{C}$ , we have

$$\sum_{h \leq H} c(h) \mathcal{E}_{\mathcal{I}}(h) = \begin{cases} o_{w \rightarrow \infty}(HX) & \text{if } f = \Lambda, \\ o_{w \rightarrow \infty}(HX(\log X)^{(k-1)\ell}) & \text{if } f = d_k. \end{cases}$$

<sup>15</sup>We thank Jinseong Kim for stressing the need to cover the non- $s$ -normal form case.

Using for  $m \ll X$  the crude bounds  $f(m), f_w(m) \ll X^{o(1)}$ , we can shift the  $n$  sum in the definition of  $\mathcal{E}_{\mathcal{I}}(h)$  slightly to obtain

$$\begin{aligned} & \sum_{h \leq H} c(h) \mathcal{E}_{\mathcal{I}}(h) \\ &= \frac{1}{H} \sum_{h \leq H} c(h) \sum_{m \leq X} \sum_{r \leq H} \prod_{j \in \mathcal{I}} (f - f_w)(m + jh + r) \\ & \quad \times \prod_{j \in \{0, \dots, \ell-1\} \setminus \mathcal{I}} f_w(m + jh + r) + O(H^2 X^{o(1)}). \end{aligned} \tag{8.15}$$

Let

$$W' := \begin{cases} W & \text{if } f = \Lambda \\ W^w & \text{if } f = d_k. \end{cases}$$

By splitting the variables into residue classes (mod  $W'$ ), the main term on the right-hand side of (8.15) becomes

$$\sum_{0 \leq b_1, b_2, b_3 < W'} S(b_1, b_2, b_3) + O(W'^3 X^{1+o(1)}),$$

where

$$\begin{aligned} & S(b_1, b_2, b_3) \\ &:= \frac{1}{H} \sum_{h \leq H/W'} c(W'h + b_2) \\ & \quad \times \sum_{m \leq X/W'} \sum_{r \leq H/W'} \prod_{j \in \mathcal{I}} (f - f_w)(W'(m + jh + r) + b_1 + jb_2 + b_3) \\ & \quad \cdot \prod_{j \in \{0, \dots, \ell-1\} \setminus \mathcal{I}} f_w(W'(m + jh + r) + b_1 + jb_2 + b_3). \end{aligned}$$

Thus, our goal is to show that

$$\sum_{0 \leq b_1, b_2, b_3 < W'} S(b_1, b_2, b_3) = \begin{cases} o_{w \rightarrow \infty}(HX) & \text{if } f = \Lambda, \\ o_{w \rightarrow \infty}(HX(\log X)^{(k-1)\ell}) & \text{if } f = d_k. \end{cases} \tag{8.16}$$

Now we bound this separately in the cases  $f = \Lambda$  and  $f = d_k$ .

**Case  $f = \Lambda$ .** Note that  $S(b_1, b_2, b_3) \ll (H/W^2)X^{1/2} \log X$  unless  $(b_1 + jb_2 + b_3, W) = 1$  for all  $0 \leq j < \ell$  since if  $\Lambda(Wn + r) > 0$  and  $(W, r) > 1$ , then  $Wn + r$  is

a perfect power. If in turn  $(b_1 + jb_2 + b_3, W) = 1$  for all  $0 \leq j < \ell$ , we can simplify

$$\begin{aligned}
 & S(b_1, b_2, b_3) \\
 &= \frac{1}{H} \sum_{h \leq H/W} c(W'h + b_2) \\
 &\quad \times \sum_{m \leq X/W} \sum_{r \leq H/W} \prod_{j \in \mathcal{I}} \left( \frac{\varphi(W)}{W} \Lambda - 1 \right) (W(m + jh + r) + b_1 + jb_2 + b_3) \\
 &\quad \cdot \prod_{j=0}^{\ell-1} \Lambda_w(b_1 + jb_2 + b_3).
 \end{aligned} \tag{8.17}$$

Let  $D \in \mathbb{N}$  be a large constant. Then by Lemma 7.5(i) and Theorem 1.3(i) there exists a constant  $C_0 \geq 1$  and an exceptional set  $E \subset [1, X/W] \cap \mathbb{N}$  of size  $\ll_A X/\log^A X$  such that for any  $m \in [1, X/W] \setminus E$  the following hold for any integer  $1 \leq b \leq W$  with  $(b, W) = 1$ :

- (i) The function  $(\frac{\varphi(W)}{W} \Lambda(Wn + b) + 1)/C_0$  is majorized by a  $(D, o_{w \rightarrow \infty}(1))$ -pseudorandom function at location  $m$  and scale  $H$ .
- (ii) We have

$$\left\| \frac{\varphi(W)}{W} \Lambda(Wn + b) - 1 \right\|_{U^D(m, m+\ell H]} = o_{w \rightarrow \infty}(1).$$

Separating on the right-hand side of (8.17) the terms with  $m \notin E$  and applying Lemma 8.4, it then follows that

$$\begin{aligned}
 S(b_1, b_2, b_3) &= o_{w \rightarrow \infty} \left( \frac{HX}{W^3} \prod_{j=0}^{\ell-1} \Lambda_w(b_1 + jb_2 + b_3) \right) \\
 &\quad + O \left( \frac{(\log X)^\ell}{H} \left( \frac{H}{W} \right)^2 \cdot |E| \right).
 \end{aligned} \tag{8.18}$$

Since  $|E| \ll_A X/\log^A X$ , the  $O(\cdot)$  error term here is  $\ll_A HX/(W^3 \log^{100} X)$ , which is negligible.

Observe that for any functions  $g_p: \mathbb{Z}^3 \rightarrow \mathbb{C}$  that are  $p$ -periodic in each coordinate, we have the identity

$$\sum_{0 \leq b_1, b_2, b_3 < W} \prod_{p < w} g_p(b_1, b_2, b_3) = \prod_{p < w} \sum_{0 \leq b_1, b_2, b_3 < p} g_p(b_1, b_2, b_3).$$

Applying this with  $g_p(b_1, b_2, b_3) = \prod_{j=0}^{\ell-1} \Lambda_p(b_1 + jb_2 + b_3)$ , where  $\Lambda_p := \frac{p}{p-1} 1_{(\cdot, p)=1}$ , we obtain

$$\begin{aligned} & \frac{1}{W^3} \sum_{0 \leq b_1, b_2, b_3 < W} \prod_{j=0}^{\ell-1} \Lambda_w(b_1 + jb_2 + b_3) \\ &= \prod_{p < w} \frac{1}{p^3} \sum_{0 \leq b_1, b_2, b_3 < p} \prod_{j=0}^{\ell-1} \Lambda_p(b_1 + jb_2 + b_3) \\ &= \prod_{p < w} \left( \frac{p}{p-1} \right)^\ell \left( \frac{p-1}{p} \left( 1 - \frac{\min\{\ell, p\}}{p} \right) + \frac{1}{p} \left( 1 - \frac{1}{p} \right) \right) \\ &\ll \prod_{p < w} \left( 1 + O\left( \frac{1}{p^2} \right) \right) \ll 1. \end{aligned}$$

Now from (8.18) we conclude that

$$\frac{1}{W^3} \sum_{0 \leq b_1, b_2, b_3 < W} S(b_1, b_2, b_3) = o_{w \rightarrow \infty}(HX),$$

as desired.

**Case  $f = d_k$ .** Note that  $d_{k,W}(W'n + b) = d_{k,W}(b)$  whenever  $p^w \nmid b$  for all  $p \leq w$ . Let us say that an integer tuple  $(b_1, b_2, b_3) \in [0, W')^3$  is *good* if

$$p^w \nmid b_1 + jb_2 + b_3 \text{ for all } p \leq w \text{ and } j = 0, \dots, \ell - 1.$$

Recall that we need to prove (8.16). We estimate the sum there separately for good tuples and tuples that are not good.

**Contribution of good tuples.** For any good  $(b_1, b_2, b_3)$ , denote for brevity

$$G_j = G_j(b_1, b_2, b_3) = (W', b_1 + jb_2 + b_3).$$

Note that with this notation, for any good  $(b_1, b_2, b_3)$ , we have

$$\begin{aligned} & d_k(W'(m + jh + r) + b_1 + jb_2 + b_3) \\ &= d_k(G_j) d_k \left( \frac{W'(m + jh + r) + b_1 + jb_2 + b_3}{G_j} \right) \\ &= d_{k,W}(b_1 + jb_2 + b_3) d_k \left( \frac{W'(m + jh + r) + b_1 + jb_2 + b_3}{G_j} \right). \end{aligned}$$

Hence

$$\begin{aligned} & S(b_1, b_2, b_3) \\ &= \left( \frac{\log^{k-1} X}{(k-1)!} \left( \frac{\varphi(W)}{W} \right)^{k-1} \right)^\ell \frac{1}{H} \sum_{h \leq H/W'} c(W'h + b_2) \prod_{j=0}^{\ell-1} d_{k,W}(b_1 + jb_2 + b_3) \end{aligned}$$

$$\sum_{m \leq X/W'} \sum_{r \leq H/W'} \prod_{j \in \mathcal{I}} \left( d_k \left( \frac{W'}{G_j} (m + jh + r) + \frac{b_1 + jb_2 + b_3}{G_j} \right) \times \left( \frac{W}{\varphi(W)} \right)^{k-1} \frac{(k-1)!}{\log^{k-1} X} - 1 \right).$$

Let  $D \in \mathbb{N}$  be a large constant. Then by Lemma 7.5(ii) and Theorem 1.3(iii) there exists a constant  $C_0 \geq 1$  and an exceptional set  $E \subset [1, X/W'] \cap \mathbb{N}$  of size  $\ll_A X/\log^A X$  such that for any  $m \in [1, X/W'] \setminus E$  the following hold for any integer  $1 \leq b \leq W'$  with  $(b, W') = 1$ :

(i) The function

$$\frac{\log^{1-k} X \left( \frac{W}{\varphi(W)} \right)^{k-1} d_k(W'n + b) + 1}{C_0}$$

is majorized by a  $(D, o_{w \rightarrow \infty}(1))$ -pseudorandom function at location  $m$  and scale  $H$ .

(ii) We have

$$\left\| \left( \frac{W}{\varphi(W)} \right)^{k-1} \frac{(k-1)!}{(\log X)^{k-1}} d_k(W'n + b) - 1 \right\|_{U^D(m, m+\ell H/W')} = o_{w \rightarrow \infty}(1).$$

Write  $B_j = b_1 + jb_2 + b_3$  for brevity. From Lemma 8.4, it then follows that

$$\begin{aligned} & S(b_1, b_2, b_3) \\ &= o_{w \rightarrow \infty} \left( \frac{HX}{W'^3} (\log X)^{(k-1)\ell} \left( \frac{\varphi(W)}{W} \right)^{(k-1)\ell} \prod_{j=0}^{\ell-1} d_{k,W}(b_1 + jb_2 + b_3) \right) \\ &+ O \left( \frac{(\log X)^{(k-1)\ell}}{H} \sum_{h \leq H/W'} \sum_{\substack{m \leq X/W' \\ m \in E}} \sum_{r \leq H/W'} \prod_{j=0}^{\ell-1} d_k(W'(m + jh + r) + B_j) \right). \end{aligned} \tag{8.19}$$

Using the inequality  $x_0 \cdots x_{\ell-1} \leq x_0^\ell + \cdots + x_{\ell-1}^\ell$  for  $x_j \geq 0$  and the estimate

$$\max_{1 \leq x \leq 2X} \max_{1 \leq b \leq (\ell+1)W'} \sum_{x \leq n \leq x+H} d_k(W'n + b)^\ell \ll H(\log X)^{O(1)}$$

coming from Shiu’s bound (Lemma 2.18), and recalling that  $|E| \ll_A X/\log^A X$ , the  $O(\cdot)$  term in (8.19) gives a contribution of  $\ll HX/(W'^3(\log X)^{100})$ , say, which is negligible.

Since the function  $d_{k,W}$  is constant on those progressions  $b \pmod{W'}$  for which  $p^w \nmid b$  for all  $p \leq w$ , we obtain

$$\begin{aligned} & \left(\frac{\varphi(W)}{W}\right)^{(k-1)\ell} \frac{1}{W'^3} \sum_{\substack{0 \leq b_1, b_2, b_3 < W' \\ b_1 + jb_2 + b_3 \text{ good}}} \prod_{j=0}^{\ell-1} d_{k,W}(b_1 + jb_2 + b_3) \\ &= \left(\frac{\varphi(W)}{W}\right)^{(k-1)\ell} \frac{1}{W'^3 \lfloor X/W' \rfloor^2} \sum_{\substack{b_1, b_2 \leq W' \lfloor X/W' \rfloor \\ 0 \leq b_3 < W' \\ b_1 + jb_2 + b_3 \text{ good}}} \prod_{j=0}^{\ell-1} d_{k,W}(b_1 + jb_2 + b_3) \\ &\ll \left(\frac{\varphi(W)}{W}\right)^{(k-1)\ell} \frac{1}{X^2} \sum_{b_1, b_2 \leq X+W'} \prod_{j=0}^{\ell-1} d_{k,W}(b_1 + jb_2). \end{aligned}$$

By Lemma 8.3, this is

$$\begin{aligned} & \ll \left(\frac{\varphi(W)}{W}\right)^{(k-1)\ell} \frac{1}{X} \sum_{b_2 \leq 2X} \prod_{p|b_2} \left(1 + o\left(\frac{1}{p}\right)\right) \\ & \quad \times \prod_{\ell < p \leq (2X)^{1/2}} \left(1 - \frac{\ell}{p}\right) \left(\sum_{n \leq (2X)^{1/2}} \frac{d_{k,W}(n)}{n}\right)^\ell \\ & \ll \frac{1}{X} \sum_{b_2 \leq 2X} \prod_{p|b_2} \left(1 + o\left(\frac{1}{p}\right)\right) \ll 1. \end{aligned}$$

Returning to (8.19), we obtain

$$\sum_{\substack{0 \leq b_1, b_2, b_3 < W' \\ (b_1, b_2, b_3) \text{ good}}} S(b_1, b_2, b_3) = o_{w \rightarrow \infty}(HX(\log X)^{k\ell - \ell}).$$

**Contribution of non-good tuples.** It remains to bound the contribution of the non-good tuples  $(b_1, b_2, b_3) \in [0, W']^3$  to (8.16). We can split such tuples according to the greatest common divisors

$$G_j = (W', b_1 + jb_2 + b_3).$$

Let

$$\Sigma(G_0, \dots, G_{\ell-1}) := \frac{1}{W'^3} \sum_{\substack{0 \leq b_1, b_2, b_3 < W' \\ (W', b_1 + jb_2 + b_3) = G_j \forall j}} S(b_1, b_2, b_3). \tag{8.20}$$

Note that since  $(b_1, b_2, b_3)$  is not good,  $p^w \mid G_j$  for some  $0 \leq j \leq \ell - 1$  and  $2 \leq p \leq w$ .

Then it suffices to show that

$$\sum_{\substack{G_0, \dots, G_{\ell-1} | W^\infty \\ \exists j: p^w | G_j \text{ for some } p \leq w}} |\Sigma(G_0, \dots, G_{\ell-1})| = o_{w \rightarrow \infty}(HX(\log X)^{(k-1)\ell}).$$

Let  $(b_1, b_2, b_3)$  be a tuple that is not good. Write  $B_j = b_1 + jb_2 + b_3$  for brevity. Using the inequalities  $d_k(ab) \leq d_k(a)d_k(b)$ ,  $d_{k,W}(ab) \leq d_{k,W}(a)d_{k,W}(b)$  for  $a, b \in \mathbb{N}$ , we see that the contribution of  $S(b_1, b_2, b_3)$  is

$$\begin{aligned} &\ll \frac{1}{H} \sum_{h \leq H/W'} \sum_{m \leq X/W'} \sum_{r \leq H/W'} \prod_{j=0}^{\ell-1} \left( d_k + (\log^{k-1} X) \left( \frac{\varphi(W)}{W} \right)^{k-1} d_{k,W} \right) \\ &\quad \times (W'(m + jh + r) + B_j) \\ &\ll \frac{1}{W'} \sum_{h \leq H/W'} \sum_{m \leq 2X/W'} \prod_{j=0}^{\ell-1} \left( d_k + (\log^{k-1} X) \left( \frac{\varphi(W)}{W} \right)^{k-1} d_{k,W} \right) \\ &\quad \times (W'(m + jh) + B_j) \\ &\ll \frac{\prod_{j=0}^{\ell-1} d_k(G_j)}{W'} \sum_{h \leq H/W'} \sum_{m \leq 2X/W'} \prod_{j=0}^{\ell-1} \left( d_k + (\log^{k-1} X) \left( \frac{\varphi(W)}{W} \right)^{k-1} d_{k,W} \right) \\ &\quad \times \left( \frac{W'}{G_j}(m + jh) + \frac{B_j}{G_j} \right). \end{aligned}$$

We have the crude bound  $\prod_{j=0}^{\ell-1} d_k(G_j) \ll (G_0 \cdots G_{\ell-1})^{1/(100\ell)}$ , say. Hence, for some  $\mathcal{J} \subset \{0, \dots, \ell - 1\}$ , the above is

$$\begin{aligned} &\ll \frac{(G_0 \cdots G_{\ell-1})^{1/(100\ell)}}{W'} \sum_{h \leq H/W'} \sum_{m \leq 2X/W'} \prod_{j \in \mathcal{J}} (\log^{k-1} X) d_{k,W} \\ &\quad \times \left( \frac{W'}{G_j}(m + jh) + \frac{B_j}{G_j} \right) \tag{8.21} \\ &\cdot \prod_{j \in \{0, \dots, \ell-1\} \setminus \mathcal{J}} d_k \left( \frac{W'}{G_j}(m + jh) + \frac{B_j}{G_j} \right). \end{aligned}$$

We now apply Lemma 8.3 to upper bound this. We take  $a_j = W'/G_j$ ,  $h_j = (W'/G_j)jh + B_j/G_j$  there and note that  $(a_j, h_j) = 1$ . Denote  $D = \prod_{i < j} (a_i h_j - a_j h_i)^2$ . Then

$$p \mid D \implies p \mid W' \prod_{\substack{|i| \leq \ell \\ i \neq 0}} (W'ih + ib_2).$$

Then Lemma 8.3, (8.12) and Shiu’s bound (Lemma 2.18) imply that (8.21) is

$$\begin{aligned} &\ll \frac{(G_0 \cdots G_{\ell-1})^{1/(100\ell)}}{W'} \sum_{h \leq H/W'} \prod_{p|W'} \left(1 + o\left(\frac{1}{p}\right)\right) \prod_{\substack{|i| \leq \ell \\ i \neq 0}} \prod_{p|W'ih+ib_2} \left(1 + o\left(\frac{1}{p}\right)\right) \\ &\quad \cdot \frac{X}{W'} \prod_{\ell < p \leq (2X/W')^{1/2}} \left(1 - \frac{\ell \cdot 1_{p>w}}{p}\right) \\ &\quad \times \prod_{j=0}^{\ell-1} \sum_{n \leq (2X/W')^{1/2}} \frac{\max\{d_k(n), (\log^{k-1} X)d_{k,W}(n)\}}{n} \\ &\ll \frac{(G_0 \cdots G_{\ell-1})^{1/(100\ell)}}{W'} \sum_{h \leq H/W'} (\log w)^{O(1)} \frac{X}{W'} (\log X)^{(k-1)\ell} \\ &\quad \times \prod_{\substack{|i| \leq \ell \\ i \neq 0}} \prod_{p|W'ih+ib_2} \left(1 + o\left(\frac{1}{p}\right)\right). \end{aligned}$$

Using the inequality  $x_1 \cdots x_\ell \leq \ell \max_i x_i^\ell$  for  $x_j \geq 0$ , we see from Shiu’s bound that

$$\begin{aligned} \sum_{h \leq H/W'} \prod_{\substack{|i| \leq \ell \\ i \neq 0}} \prod_{p|W'ih+ib_2} \left(1 + o\left(\frac{1}{p}\right)\right) &\ll \max_{\substack{|i| \leq \ell \\ i \neq 0}} \sum_{h \leq H/W'} \prod_{p|W'ih+ib_2} \left(1 + o\left(\frac{1}{p}\right)\right)^\ell \\ &\ll \max_{\substack{|i| \leq \ell \\ i \neq 0}} \sum_{\substack{h' \leq H\ell \\ h' \equiv ib_2 \pmod{W'}}} \prod_{p|h'} \left(1 + o\left(\frac{1}{p}\right)\right)^\ell \\ &\ll (\log w)^{O(1)} \frac{H}{W'}. \end{aligned}$$

Since  $(\log w)^B \ll_B (\max G_j)^{1/100}$  for any  $B > 0$  and the number of tuples  $(b_1, b_2, b_3) \in [1, W']^3$  with  $(W', b_1 + jb_2 + b_3) = G_j$  for all  $j$  is  $\ll (W')^3 / \max_j G_j$ , recalling (8.20) we conclude that

$$|\Sigma(G_0, \dots, G_{\ell-1})| \ll \frac{1}{W'^3} (\log^{(k-1)\ell} X) H X (\max_j G_j)^{2/100-1} \ll \frac{(\log^{(k-1)\ell} X) H X}{(\max_j G_j)^{1/2}}.$$

Hence

$$\begin{aligned} &(\log X)^{-(k-1)\ell} (H X)^{-1} \sum_{\substack{G_0, \dots, G_{\ell-1} | W^\infty \\ \exists j: p^w | G_j \text{ for some } p \leq w}} \Sigma(G_0, \dots, G_{\ell-1}) \\ &\ll \sum_{\substack{G_0, \dots, G_{\ell-1} | W^\infty \\ \exists j: p^w | G_j \text{ for some } p \leq w}} (G_0 \cdots G_{\ell-1})^{-1/(2\ell)} \end{aligned}$$

$$\begin{aligned} &\ll 2^{-w/(6\ell)} \sum_{G_0, \dots, G_{\ell-1} | W^\infty} (G_0 \cdots G_{\ell-1})^{-1/(3\ell)} \\ &= 2^{-w/(6\ell)} \prod_{p \leq w} \left( 1 + O\left(\frac{1}{p^{1/(3\ell)}}\right) \right)^\ell \\ &\ll 2^{-w/(6\ell)} \exp(O(w^{1-1/(3\ell)})) \ll 2^{-w/(10\ell)}, \end{aligned}$$

say. This completes the proof. □

### 9 Fixed Fourier estimate for $d_2$

In this section we establish Theorem 1.8.

**Proof of Theorem 1.8** Let  $Q = X^{\varepsilon/2} \leq H^{1/2}$ . By Dirichlet’s approximation theorem, there exist some integers  $a, q$  with  $(a, q) = 1$  and  $1 \leq q \leq Q$  such that  $|\alpha - a/q| \leq 1/(qQ)$ . We take  $H' = qQ^{1/2}$ . Then, for all  $n \in [x, x + H']$  we have

$$\begin{aligned} e(\alpha n) &= e(an/q) \cdot e((\alpha - a/q)x) + O(H'/(qQ)) \\ &= e(an/q) \cdot e((\alpha - a/q)x) + O(X^{-\varepsilon/4}). \end{aligned}$$

Hence, by splitting into intervals of length  $H' \leq H^{3/4}$ , it suffices to show that we have

$$\left| \sum_{x < n \leq x + H'} (d(n) - d^\sharp(n))e(an/q) \right|^* \leq H' X^{-\varepsilon/1000} / 2.$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_\varepsilon(X^{1-\varepsilon/400})$ .

Let  $H_1 = X^{1-1/100}$ . By splitting into progressions modulo  $q$  and applying [30, Proposition 3.2(ii)], we see that for any  $x \in [X, 2X]$  we have

$$\begin{aligned} \left| \sum_{x < n \leq x + H_1} (d(n) - d^\sharp(n))e(an/q) \right|^* &\leq q \left| \sum_{x < n \leq x + H_1} (d(n) - d^\sharp(n)) \right|^* \\ &\ll Q \frac{H_1^2}{X} \ll H_1 X^{-\varepsilon}. \end{aligned}$$

Hence, it suffices to show that

$$\begin{aligned} &\max_{P \text{ arithmetic progression}} \left| \frac{1}{H'} \sum_{\substack{x < n \leq x + H' \\ n \in P}} d(n)e\left(\frac{an}{q}\right) - \frac{1}{H_1} \sum_{\substack{x < n \leq x + H_1 \\ n \in P}} d(n)e\left(\frac{an}{q}\right) \right| \\ &\ll X^{-\varepsilon/1000} \end{aligned}$$

and

$$\max_{P \text{ arithmetic progression}} \left| \frac{1}{H'} \sum_{\substack{x < n \leq x+H' \\ n \in P}} d^\sharp(n) e\left(\frac{an}{q}\right) - \frac{1}{H_1} \sum_{\substack{x < n \leq x+H_1 \\ n \in P}} d^\sharp(n) e\left(\frac{an}{q}\right) \right| \ll X^{-\varepsilon/1000}$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_\varepsilon(X^{1-\varepsilon/400})$ . In the maxima above, we may restrict to arithmetic progressions of modulus  $\leq X^{\varepsilon/999}$  thanks to the triangle inequality and the divisor bound. Hence, by the union bound, it suffices to show that for any  $1 \leq b \leq r \leq X^{\varepsilon/999}$  we have

$$\left| \frac{1}{H'} \sum_{\substack{x < n \leq x+H' \\ n \equiv b \pmod{r}}} d(n) e\left(\frac{an}{q}\right) - \frac{1}{H_1} \sum_{\substack{x < n \leq x+H_1 \\ n \equiv b \pmod{r}}} d(n) e\left(\frac{an}{q}\right) \right| \leq \frac{1}{10} X^{-\varepsilon/1000}$$

and

$$\left| \frac{1}{H'} \sum_{\substack{x < n \leq x+H' \\ n \equiv b \pmod{r}}} d^\sharp(n) e\left(\frac{an}{q}\right) - \frac{1}{H_1} \sum_{\substack{x < n \leq x+H_1 \\ n \equiv b \pmod{r}}} d^\sharp(n) e\left(\frac{an}{q}\right) \right| \leq \frac{1}{10} X^{-\varepsilon/1000}$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_\varepsilon(X^{1-\varepsilon/300})$ . By Fourier expanding the summation condition  $n \equiv b \pmod{r}$ , we reduce to showing that for any  $1 \leq q' \leq qX^{\varepsilon/999}$  and  $(a', q') = 1$  we have

$$\left| \frac{1}{H'} \sum_{x < n \leq x+H'} d(n) e\left(\frac{a'n}{q'}\right) - \frac{1}{H_1} \sum_{x < n \leq x+H_1} d(n) e\left(\frac{a'n}{q'}\right) \right| \leq \frac{1}{10} X^{-\varepsilon/1000} \tag{9.1}$$

and

$$\left| \frac{1}{H'} \sum_{x < n \leq x+H'} d^\sharp(n) e\left(\frac{a'n}{q'}\right) - \frac{1}{H_1} \sum_{x < n \leq x+H_1} d^\sharp(n) e\left(\frac{a'n}{q'}\right) \right| \leq \frac{1}{10} X^{-\varepsilon/1000} \tag{9.2}$$

for all  $x \in [X, 2X]$  outside of a set of measure  $O_\varepsilon(X^{1-\varepsilon/300})$ .

By Lemma 2.19(ii), the function  $d_k^\sharp$  is a sum of  $O(1)$  terms of the form  $a * \psi$ , where  $|a(m)| \ll d_{k-1}(m)$  and  $a$  is supported on  $[1, X^{\varepsilon/5}]$  and each function  $\psi$  is of the form  $\psi(n) = (\log^\ell n)/(\log^\ell X)$  with  $0 \leq \ell \leq k - 1$  an integer. The estimate (9.2) follows from this combined with the triangle inequality and partial summation (splitting into residue classes  $\pmod{q'}$ ).

We are left with showing (9.1). By Markov’s inequality, it suffices to show that

$$\int_X^{2X} \left| \frac{1}{H'} \sum_{x < n \leq x+H'} d(n)e\left(\frac{a'n}{q'}\right) - \frac{1}{H_1} \sum_{x < n \leq x+H_1} d(n)e\left(\frac{a'n}{q'}\right) \right| dx \ll X^{1-\varepsilon/100}.$$

Factorizing  $n = \ell m$  with  $\ell \mid (q')^\infty$  and  $(m, q') = 1$ , and applying the triangle inequality, we reduce to showing that

$$\sum_{\ell \mid (q')^\infty} d(\ell) \int_X^{2X} \left| \frac{1}{H'} \sum_{x/\ell < m \leq x/\ell + H'/\ell} d(m)e\left(\frac{a'\ell m}{q'}\right) 1_{(m, q')=1} - \frac{1}{H_1} \sum_{x/\ell < m \leq x/\ell + H_1/\ell} d(m)e\left(\frac{a'\ell m}{q'}\right) 1_{(m, q')=1} \right| dx \ll X^{1-\varepsilon/100}.$$

For any  $L \geq 1$ , we have the bound

$$\begin{aligned} \sum_{\substack{\ell \mid (q')^\infty \\ \ell \geq L}} \frac{1}{\ell} &\leq \frac{1}{\sqrt{L}} \sum_{\substack{\ell \mid (q')^\infty \\ \ell \geq L}} \frac{1}{\ell^{1/2}} \\ &\leq \frac{1}{\sqrt{L}} \prod_{p \mid q'} \left(1 + \frac{1}{p^{1/2}} + o\left(\frac{1}{p}\right)\right) \\ &\ll \frac{1}{\sqrt{L}} \exp(O((\log q')^{1/2})). \end{aligned}$$

Combining this with the triangle inequality, we see that it suffices to show that

$$\int_X^{2X} \left| \frac{1}{H'} \sum_{x/\ell < m \leq x/\ell + H'/\ell} d(m)e\left(\frac{a'\ell m}{q'}\right) 1_{(m, q')=1} - \frac{1}{H_1} \sum_{x/\ell < m \leq x/\ell + H_1/\ell} d(m)e\left(\frac{a'\ell m}{q'}\right) 1_{(m, q')=1} \right| dx \ll X^{1-\varepsilon/30}$$

for any  $1 \leq \ell \leq X^{\varepsilon/40}$ .

For any integer  $b$ , we have the Fourier expansion

$$e\left(\frac{bm}{q'}\right) 1_{(m, q')=1} = \frac{1}{\varphi(q')} \sum_{\chi \pmod{q'}} \tau(\bar{\chi}) \chi(bm),$$

where  $\tau(\chi)$  is the Gauss sum. Using the bound  $|\tau(\chi)| \leq q^{1/2}$  for non-principal  $\chi \pmod{q}$  together with the triangle inequality and Cauchy–Schwarz, we reduce to

showing that

$$\sum_{\chi \pmod{q'}} \int_X^{2X} \left| \frac{1}{H'} \sum_{x/\ell < m \leq (x+H')/\ell} d(m)\chi(m) - \frac{1}{H_1} \sum_{x/\ell < m \leq (x+H_1)/\ell} d(m)\chi(m) \right|^2 dx \ll X^{1-\varepsilon/14}.$$

By Lemma 3.3(i), we reduce to showing that

$$\sum_{\chi \pmod{q'}} \max_{T \geq \frac{X}{H'}} \frac{X/H'}{T} \int_{X^{1/400} \leq |t| \leq T} |D(1+it, \chi)|^2 dt \ll X^{-\varepsilon/14}, \tag{9.3}$$

where

$$D(s, \chi) := \sum_{X/8 < n \leq 8X} d(n)\chi(n)n^{-s}.$$

By the mean value theorem for Dirichlet polynomials and crude estimation, we may restrict the maximum over  $T$  to  $T \leq X^{1-\varepsilon/1000}$ .

Using Perron’s formula, we have, for  $|t| \leq X^{1-\varepsilon/1000}$

$$D(1+it, \chi) = \frac{1}{2\pi i} \int_{1-i|t|/2}^{1+i|t|/2} L(1+it+w, \chi)^2 \frac{(8X)^w - (X/8)^w}{w} dw + O\left(\frac{\log X}{|t|}\right).$$

Moving the line of integration to  $\text{Re}(w) = -1/2$  and applying the triangle inequality, we obtain

$$|D(1+it, \chi)| \ll X^{-1/2} \int_{-|t|/2}^{|t|/2} |L(1/2+i(t+u), \chi)|^2 \frac{du}{1+|u|} + O\left(\frac{\log X}{|t|}\right).$$

Substituting this to (9.3) and applying the Cauchy–Schwarz inequality, we see that it suffices to show that

$$\sum_{\chi \pmod{q'}} \max_{\frac{X}{H'} \leq T \leq X} \frac{X/H'}{T} \int_{X^{1/500} \leq |t| \leq T} |L(1/2+it, \chi)|^4 dt \ll X^{1-\varepsilon/10}.$$

By splitting characters according to their conductors, and using the Euler product formula, we reduce to showing that for all  $r \mid q'$  we have

$$\sum_{\chi' \pmod{r}}^* \max_{\frac{X}{H'} \leq T \leq X} \frac{X/H'}{T} \int_{X^{1/500} \leq |t| \leq T} |L(1/2+it, \chi')|^4 \prod_{p \mid \frac{q'}{r}} \left(1+p^{-1/2}\right)^4 dt \ll X^{1-\varepsilon/9}.$$

where  $\sum^*$  denotes a sum over primitive characters. Without loss of generality, we may restrict the maximum over  $T$  to powers of two, so, upper bounding the product over  $p$  by  $x^\varepsilon/100$ , it suffices to show that

$$\max_{\frac{x}{H'} \leq T \leq x} \frac{X/H'}{T} \sum_{\chi' \pmod{r}}^* \int_{X^{1/200} \leq |t| \leq T} |L(1/2 + it, \chi')|^4 dt \ll X^{1-\varepsilon/8}. \quad (9.4)$$

But by the fourth moment bound for Dirichlet  $L$ -functions [32, Theorem 10.1], the left-hand side of (9.4) is

$$\ll \varphi(r) \frac{X}{H'} \log^4 X.$$

The bound (9.4) immediately follows from this since  $r \leq q' \leq qX^{\varepsilon/999}$  and  $H' = qQ^{1/2} = qX^{\varepsilon/4}$ .  $\square$

**Acknowledgements** We are grateful to the anonymous referee for a careful reading of the paper and for numerous helpful comments and corrections. We thank Jiseong Kim for a correction.

**Funding information** KM was supported by Academy of Finland grant no. 285894. MR acknowledges the support of NSF grant DMS-2401106. XS was supported by NSF grant DMS-2200565. 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 Academy of Finland grant no. 362303, a von Neumann Fellowship (NSF grant DMS-1926686), and funding from European Union's Horizon Europe research and innovation programme under Marie Skłodowska-Curie grant agreement no. 101058904 and ERC grant agreement no. 101162746.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Baker, R.C., Harman, G., Pintz, J.: The difference between consecutive primes. II. Proc. Lond. Math. Soc. (3) **83**(3), 532–562 (2001)
2. Conrey, J.B., Gonek, S.M.: High moments of the Riemann zeta-function. Duke Math. J. **107**(3), 577–604 (2001)
3. Frantzikinakis, N., Host, B.: Higher order Fourier analysis of multiplicative functions and applications. J. Am. Math. Soc. **30**(1), 67–157 (2017)
4. Friedlander, J.B., Iwaniec, H.: The divisor problem for arithmetic progressions. Acta Arith. **45**(3), 273–277 (1985)
5. Green, B., Tao, T.: Linear equations in primes. Ann. Math. (2) **171**(3), 1753–1850 (2010)
6. Green, B., Tao, T.: The Möbius function is strongly orthogonal to nilsequences. Ann. Math. (2) **175**(2), 541–566 (2012)
7. Green, B., Tao, T.: The quantitative behaviour of polynomial orbits on nilmanifolds. Ann. Math. (2) **175**(2), 465–540 (2012)

8. Green, B., Tao, T., Ziegler, T.: An inverse theorem for the Gowers  $U^{s+1}[N]$ -norm. *Ann. Math. (2)* **176**(2), 1231–1372 (2012)
9. Guth, L., Maynard, J.: New large value estimates for Dirichlet polynomials. arXiv e-prints, [arXiv:2405.20552](https://arxiv.org/abs/2405.20552) (2024)
10. Hardy, G.H., Littlewood, J.E.: Some problems of ‘Partitio numerorum’; III: on the expression of a number as a sum of primes. *Acta Math.* **44**(1), 1–70 (1923)
11. Harman, G.: Prime-Detecting Sieves. London Mathematical Society Monographs Series, vol. 33. Princeton University Press, Princeton (2007)
12. He, X., Wang, Z.: Möbius disjointness for nilsequences along short intervals. *Trans. Am. Math. Soc.* **374**(6), 3881–3917 (2021)
13. Henriot, K.: Nair-Tenenbaum bounds uniform with respect to the discriminant. *Math. Proc. Camb. Philos. Soc.* **152**(3), 405–424 (2012)
14. Hilbert, D.: Ueber die Irreducibilität ganzer rationaler Functionen mit ganzzahligen Coefficienten. *J. Reine Angew. Math.* **110**, 104–129 (1892)
15. Ivić, A.: The general additive divisor problem and moments of the zeta-function. In: *New Trends in Probability and Statistics*, Palanga, 1996, vol. 4, pp. 69–89. VSP, Utrecht (1997)
16. Iwaniec, H., Kowalski, E.: *Analytic Number Theory*. American Mathematical Society Colloquium Publications., vol. 53. Am. Math. Soc., Providence (2004)
17. Jutila, M.: On the divisor problem for short intervals. *Ann. Univ. Turku. Ser. A I* **186**, 23–30 (1984). Studies in honour of Arto Kustaa Salomaa on the occasion of his fiftieth birthday
18. Leng, J., Sah, A., Sawhney, M.: Quasipolynomial bounds on the inverse theorem for the Gowers  $U^{s+1}[N]$ -norm. arXiv e-prints [arXiv:2402.17994](https://arxiv.org/abs/2402.17994) (2024)
19. Mangerel, A.P.: Divisor-bounded multiplicative functions in short intervals. *Res. Math. Sci.* **10**(12) (2023)
20. Matomäki, K., Radziwiłł, M.: A note on the Liouville function in short intervals. arXiv e-prints, [arXiv:1502.02374](https://arxiv.org/abs/1502.02374) (2015)
21. Matomäki, K., Radziwiłł, M.: Multiplicative functions in short intervals. *Ann. Math. (2)* **183**(3), 1015–1056 (2016)
22. Matomäki, K., Radziwiłł, M.: Multiplicative functions in short intervals II. arXiv e-prints, [arXiv:2007.04290](https://arxiv.org/abs/2007.04290) (2020)
23. Matomäki, K., Shao, X.: Discorrelation between primes in short intervals and polynomial phases. *Int. Math. Res. Not.* **16**, 12330–12355 (2021)
24. Matomäki, K., Teräväinen, J.: On the Möbius function in all short intervals. *J. Eur. Math. Soc.* **25**(4), 1207–1225 (2023)
25. Matomäki, K., Radziwiłł, M., Tao, T.: An averaged form of Chowla’s conjecture. *Algebra Number Theory* **9**(9), 2167–2196 (2015)
26. Matomäki, K., Radziwiłł, M., Tao, T.: Correlations of the von Mangoldt and higher divisor functions I. Long shift ranges. *Proc. Lond. Math. Soc. (3)* **118**(2), 284–350 (2019)
27. Matomäki, K., Radziwiłł, M., Tao, T.: Correlations of the von Mangoldt and higher divisor functions II: divisor correlations in short ranges. *Math. Ann.* **374**(1–2), 793–840 (2019)
28. Matomäki, K., Radziwiłł, M., Tao, T.: Fourier uniformity of bounded multiplicative functions in short intervals on average. *Invent. Math.* **220**(1), 1–58 (2020)
29. Matomäki, K., Radziwiłł, M., Tao, T., Teräväinen, J., Ziegler, T.: Higher uniformity of bounded multiplicative functions in short intervals on average. *Ann. Math. (2)* **197**(2), 739–857 (2023)
30. Matomäki, K., Shao, X., Tao, T., Teräväinen, J.: Higher uniformity of arithmetic functions in short intervals I. All intervals. *Forum Math. Pi* **11**, e29 (2023)
31. Matthesen, L.: Linear correlations of multiplicative functions. *Proc. Lond. Math. Soc. (3)* **121**(2), 372–425 (2020)
32. Montgomery, H.L.: *Topics in Multiplicative Number Theory*. Lecture Notes in Mathematics, vol. 227. Springer, Berlin (1971)
33. Montgomery, H.L.: *Ten Lectures on the Interface Between Analytic Number Theory and Harmonic Analysis*. CBMS Regional Conference Series in Mathematics, vol. 84. Am. Math. Soc., Providence (1994). Published for the Conference Board of the Mathematical Sciences, Washington, DC
34. Robert, O., Sargos, P.: Three-dimensional exponential sums with monomials. *J. Reine Angew. Math.* **591**, 1–20 (2006)
35. Shiu, P.: A Brun-Titchmarsh theorem for multiplicative functions. *J. Reine Angew. Math.* **313**, 161–170 (1980)
36. Sun, Y.-C.: On divisor bounded multiplicative functions in short intervals. arXiv e-prints, [arXiv:2401.08432](https://arxiv.org/abs/2401.08432) (2024)

37. Tao, T.: Higher Order Fourier Analysis. Graduate Studies in Mathematics, vol. 142. Am. Math. Soc., Providence (2012)
38. Tao, T.: Hilbert's Fifth Problem and Related Topics. Graduate Studies in Mathematics, vol. 153. Am. Math. Soc., Providence (2014)
39. Tao, T.: Equivalence of the logarithmically averaged Chowla and Sarnak conjectures. In: Number Theory—Diophantine Problems, Uniform Distribution and Applications, pp. 391–421. Springer, Cham (2017)
40. Tao, T.: Heuristic computation of correlations of higher order divisor functions. Available at <https://terrytao.wordpress.com/2016/08/31/>
41. Vinogradov, A.I.:  $SL_n$  techniques and the density conjecture. Zap. Nauč. Semin. LOMI **168**, 5–10 (1988), 187
42. Walsh, M.N.: Local uniformity through larger scales. Geom. Funct. Anal. **31**(4), 981–991 (2021)
43. Walsh, M.N.: Phase relations and pyramids. arXiv e-prints, [arXiv:2304.09792](https://arxiv.org/abs/2304.09792) (2023)
44. Walsh, M.N.: Stability under scaling in the local phases of multiplicative functions. arXiv e-prints, [arXiv:2310.07873](https://arxiv.org/abs/2310.07873) (2023)
45. Wang, M.: Local Fourier uniformity of higher divisor functions on average. arXiv e-prints, [arXiv:2402.18342](https://arxiv.org/abs/2402.18342) (2024)
46. Zhan, T.: On the representation of large odd integer as a sum of three almost equal primes. Acta Math. Sin. **7**(3), 259–272 (1991)

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Authors and Affiliations

**Kaisa Matomäki<sup>1</sup> · Maksym Radziwiłł<sup>2</sup> · Xuancheng Shao<sup>3</sup> · Terence Tao<sup>4</sup> · Joni Teräväinen<sup>5</sup>**

✉ J. Teräväinen  
[joni.p.teravainen@gmail.com](mailto:joni.p.teravainen@gmail.com)

K. Matomäki  
[ksmato@utu.fi](mailto:ksmato@utu.fi)

M. Radziwiłł  
[maksym.radziwill@northwestern.edu](mailto:maksym.radziwill@northwestern.edu)

X. Shao  
[xuancheng.shao@uky.edu](mailto:xuancheng.shao@uky.edu)

T. Tao  
[tao@math.ucla.edu](mailto:tao@math.ucla.edu)

<sup>1</sup> Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland

<sup>2</sup> Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208, USA

<sup>3</sup> Department of Mathematics, University of Kentucky, 715 Patterson Office Tower, Lexington, KY 40506, USA

<sup>4</sup> Department of Mathematics, UCLA, 405 Hilgard Ave, Los Angeles, CA 90095, USA

<sup>5</sup> Department of Pure Mathematics and Mathematical Statistics, University of Cambridge, Cambridge CB3 0WB, UK