J. Eur. Math. Soc. (Online first) DOI 10.4171/JEMS/1404

Terence Tao · Joni Teräväinen

Quantitative bounds for Gowers uniformity of the Möbius and von Mangoldt functions

Received July 19, 2021; revised September 14, 2023

Abstract. We establish quantitative bounds on the $U^k[N]$ Gowers norms of the Möbius function μ and the von Mangoldt function Λ for all k, with error terms of the shape $O((\log\log N)^{-c})$. As a consequence, we obtain quantitative bounds for the number of solutions to any linear system of equations of finite complexity in the primes, with the same shape of error terms. We also obtain the first quantitative bounds on the size of sets containing no k-term arithmetic progressions with shifted prime difference.

Keywords. Gowers uniformity norms, von Mangoldt function, Möbius function

1. Introduction

Throughout this paper we fix an integer $k \geq 1$, and let $N > 1$ be a real parameter that is assumed to be sufficiently large depending on k. We will also make frequent use of the somewhat smaller quantity

$$Q := \exp(\log^{1/10} N), \tag{1.1}$$

for instance by sieving out multiples of all primes less than Q. We use c to denote various small positive constants depending on k that are allowed to vary from line to line, or even within the same line. All the constants in our asymptotic notation are permitted to depend on k (see Section 3 for a more detailed description of the asymptotic notation conventions used in this paper). The implied constants will be effective, except when otherwise stated.

In this paper we will be interested in quantitatively controlling the Gowers norm uniformity of the Möbius function μ and the von Mangoldt function Λ on the interval $[N] := \{n \in \mathbb{N} : 1 \leq n \leq N\}$, as well as various related statistics. Our methods can extend to some other arithmetic functions, such as sufficiently "non-pretentious" bounded multiplicative functions, but we focus on the classical functions μ, Λ here for ease of

Terence Tao: Department of Mathematics, University of California, Los Angeles, CA 90095-1555, USA; tao@math.ucla.edu

Joni Teräväinen: Department of Mathematics and Statistics, University of Turku, 20014 Turku, Finland; joni.p.teravainen@gmail.com

Mathematics Subject Classification (2020): Primary 11B30; Secondary 11N37, 37A44

exposition. Such quantitative control on the Gowers norms will be used to quantify the asymptotics for linear equations in primes obtained in [22].

We begin by recalling the definition of the Gowers uniformity norms, first introduced by Gowers [13]; we largely follow the notation of [22, Appendix B] here, except that we will find it convenient to work with both normalized and unnormalized Gowers norms.

Definition 1.1 (Gowers norms). Let $k \geq 1$ be a natural number.

(i) If $\omega \in \{0,1\}^k$ is a k-tuple, we write $\omega_1, \dots, \omega_k \in \{0,1\}$ for the components of ω, and $|\omega| := \omega_1 + \cdots + \omega_k$. Similarly, if $\vec{h} \in G^k$ is a k-tuple in some additive group G, we write $h_1, \dots, h_k \in G$ for the components of h, and write $\omega \cdot \vec{h}$ for the "dot product"

$$\omega \cdot \vec{h} := \omega_1 h_1 + \cdots + \omega_k h_k.$$

We often identify G^{k+1} with $G \times G^k$, thus for instance the assertion $(n, \vec{h}) \in G^{k+1}$ means that $n \in G$ and $\vec{h} \in G^k$.

(ii) If $f : G \to \mathbb{C}$ is a finitely supported function on an additive group G, we define the (*unnormalized*) *Gowers uniformity norm* $\|f\|_{\tilde{U}^k(G)}$ to be the quantity

$$\|f\|_{\tilde{U}^k(G)} := \Big(\sum_{(n,\vec{h}) \in G^{k+1}} \prod_{\omega \in \{0,1\}^k} \mathcal{C}^{|\omega|} f(n + \omega \cdot \vec{h}) \Big)^{1/2^k},$$

where $\mathcal{C} : z \mapsto \overline{z}$ denotes complex conjugation. If G is finite, we then define the normalized norm

$$\|f\|_{U^k(G)} := \|f\|_{\tilde{U}^k(G)} / \|1\|_{\tilde{U}^k(G)}.$$

(iii) For any function $f : \mathbb{Z} \to \mathbb{C}$ and natural number N, we define the *local* (*normalized*) *Gowers uniformity norm*

$$\|f\|_{U^k[N]} := \|f 1_{[N]}\|_{\tilde{U}^k(\mathbb{Z})} / \|1_{[N]}\|_{\tilde{U}^k(\mathbb{Z})},$$

where $1_{[N]}$ is the indicator function of $[N]$.

Thus for instance

$$\begin{aligned} \|f\|_{U^1[N]} &= \big(\mathbb{E}_{n,h:\,(n,n+h) \in [N]^2} f(n) \overline{f(n+h)}\big)^{1/2} \\ &= |\mathbb{E}_{n \in [N]} f(n)|, \end{aligned}$$

where throughout this paper we use the averaging notation

$$\mathbb{E}_{a \in A} f(a) := \frac{1}{\#A} \sum_{a \in A} f(a)$$

for any non-empty set A of some finite cardinality $\#A$, and by the orthogonality of additive characters we can compute

$$\begin{aligned} \|f\|_{U^2[N]} &= \big(\mathbb{E}_{n,h:\,(n,n+h,n+k,n+h+k) \in [N]^4} f(n) \overline{f(n+h) f(n+k)} f(n+h+k)\big)^{1/4} \\ &\asymp N^{-3/4} \Big(\int_0^1 \Big| \sum_{n \in [N]} f(n) e(n\theta) \Big|^4 d\theta \Big)^{1/4}, \end{aligned}$$

where we adopt the usual asymptotic notation (see Section 3), and $e(\theta) := e^{2\pi i\theta}$. While we will permit the functions f to be complex-valued for compatibility with previous literature (particularly those that invoke the circle method), in this paper we will deal almost exclusively with real-valued functions. As is well known, the Gowers uniformity norms are indeed norms for $k \geq 2$, and seminorms for $k = 1$; see for instance [22, Appendix B]. In particular, they obey the triangle inequality

$$\|f + g\|_{U^k[N]} \leq \|f\|_{U^k[N]} + \|g\|_{U^k[N]} \tag{1.2}$$

(and similarly for the other variants of the Gowers norms in Definition 1.1), which we will rely on frequently in this paper.

The *Möbius pseudorandomness principle* (see e.g. [29, p. 338]) informally makes the prediction

$$\mu(n) \approx 0$$

in the metric given by the Gowers norms $U^k[N]$. Similarly, the usual modification of the Cramér random model [6], as refined by Granville [15] in order to take into account the distribution at primes below some threshold w, makes the prediction

$$\Lambda(n) \approx \Lambda_{\text{Cramér},w}(n)$$

for various small $2 \leq w \ll N$, where $\Lambda_{\text{Cramér},w} : \mathbb{Z} \to \mathbb{R}$ is the function

$$\Lambda_{\text{Cramér},w}(n) := \frac{P(w)}{\phi(P(w))} 1_{(n,P(w))=1} = \prod_{p<w} \frac{p}{p-1} 1_{p\nmid n},$$

where $P(w)$ is the primorial[1] of w,

$$P(w) := \prod_{p<w} p,$$

with ϕ the Euler totient function and $(n, P(w))$ the greatest common divisor of n and $P(w)$. Thus for instance $\Lambda_{\text{Cramér},2} = 1$ (which corresponds to the original model of Cramér). The precise choice of the parameter w is not too important, as can be shown by the following standard sieve-theoretic calculation:

Proposition 1.2 (Gowers norm stability of the Cramér model). *If $2 \leq w, z \leq Q$, then*

$$\|\Lambda_{\text{Cramér},w} - \Lambda_{\text{Cramér},z}\|_{U^k[N]} \ll \log^{-c} N + w^{-c} + z^{-c}. \tag{1.3}$$

We establish this proposition in Section 5. In our applications it will be convenient to focus on the Cramér models $\Lambda_{\text{Cramér},w}, \Lambda_{\text{Cramér},z}$ with $w = \log^\kappa N$, $z = Q$, for $\kappa > 0$ a sufficiently small constant which may depend on k (usually we can take $\kappa = 1/100$). However, using Proposition 1.2 it is not difficult to also work with other suitable choices

[1] In some texts the constraint $p \leq w$ is used in place of $p < w$; the precise convention is not too important for our applications, but the choice $p < w$ is consistent with the conventions in [11].

of parameters if desired, at least up to logarithmic decay (and probably up to *pseudopolynomial decay*[2] as well, see Remark 5.4).

We summarize the previous Gowers uniformity results on Möbius and von Mangoldt as follows.

Theorem 1.3 (Gowers uniformity of Möbius and von Mangoldt).

(i) (Pseudopolynomial U^1 uniformity) *We have*

$$\|\mu\|_{U^1[N]}, \|\Lambda - 1\|_{U^1[N]} \ll \exp(-c(\log N)^{3/5}(\log\log N)^{-1/5}).$$

(ii) (Logarithmic and strongly logarithmic U^2 uniformity) *We have*

$$\|\mu\|_{U^2[N]} \ll_A^{\text{ineff}} \log^{-A} N$$

and

$$\|\Lambda - \Lambda_{\text{Cramér},w}\|_{U^2[N]} \ll^{\text{ineff}} \log^{-c} N + w^{-c} \tag{1.4}$$

for all $A > 0$ and all $2 \le w \le Q$.

(iii) (Qualitative higher uniformity) *For any fixed $k > 2$, we have*

$$\|\mu\|_{U^k[N]} = o^{\text{ineff}}(1)$$

and

$$\|\Lambda - \Lambda_{\text{Cramér},w}\|_{U^k[N]} \ll w^{-c} + o^{\text{ineff}}(1) \tag{1.5}$$

as $N \to \infty$ uniformly for any $2 \le w \le Q$.

In the asymptotic notation superscripted with ineff*, the implied constants are permitted to be ineffective.*

A short deduction of this theorem from results stated in the literature is given in Appendix B for the sake of completeness.

The first main objective of this paper is to quantify (and make effective) the qualitative rate of decay $o^{\text{ineff}}(1)$ in Theorem 1.3 (iii). We are able to obtain doubly logarithmic bounds which are weaker than the $k = 2$ logarithmic bound in Theorem 1.3 (ii) only by a single additional logarithm:

Theorem 1.4 (Doubly logarithmic uniformity of Möbius and von Mangoldt). *For $k \ge 2$, we have*

$$\|\mu\|_{U^k[N]} \ll (\log\log N)^{-c}$$

and

$$\|\Lambda - \Lambda_{\text{Cramér},w}\|_{U^k[N]} \ll (\log\log N)^{-c} + w^{-c}$$

whenever $2 \le w \le Q$.

[2] By a *pseudopolynomially decaying* function we mean one that decays faster than $\exp(-\log^c N)$ for some $c > 0$.

This is new for $k \geq 3$; henceforth we will assume $k \geq 2$ in our arguments to avoid some minor degeneracies. We remark that this theorem (and hence all of our subsequent results) are dependent on the results in [33] (see also [1]), which are currently available in preprint form as of this time of writing.

For later use, we also state a version of Theorem 1.4 for Λ where the W-trick has been implemented.

Corollary 1.5 (W-tricked quantitative Gowers uniformity). *Let $w = (\log\log N)^{1/2}$ and $W = \prod_{p \leq w} p$. Then for $k \geq 2$ we have*

$$\left\| \frac{\phi(W)}{W} \Lambda(W \cdot + b) - 1 \right\|_{U^k[\frac{N-b}{W}]} \ll (\log\log N)^{-c}$$

whenever $1 \leq b \leq W$ is coprime to W.

In Corollary 1.5, unlike in Theorem 1.4, the size of w turns out to be important. Indeed, if we had $w/\log\log N \to \infty$, then for all we know there could be a Siegel zero to some modulus $q \leq Q$ such that all its prime factors divided W, and this would bias the main term 1 in Corollary 1.5; cf. Theorem 2.6.

1.1. Applications to linear equations in primes and to progressions with shifted prime difference

The main application of the qualitative uniformity result (1.5) in [22] was to obtain qualitative asymptotics on linear equations in the primes; now using Theorem 1.4 we can make that result quantitative.

Theorem 1.6 (Quantitative linear equations in primes). *Let N, d, t, L be positive integers, and let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms $\psi_i : \mathbb{Z}^d \to \mathbb{Z}$ of the form*

$$\psi_i(n) = n \cdot \dot{\psi}_i + \psi_i(0)$$

where $\dot{\psi}_i \in \mathbb{Z}^d$ and $\psi_i(0) \in \mathbb{Z}$ are such that $|\dot{\psi}_i| \leq L$ and $|\psi_i(0)| \leq LN$. Suppose that no two of the $\dot{\psi}_i$ are linearly dependent. Let $\Omega \subset [-N, N]^d$ be a convex body. Then

$$\sum_{\vec{n} \in \Omega \cap \mathbb{Z}^d} \prod_{i=1}^{t} \Lambda(\psi_i(\vec{n})) = \beta_\infty \prod_p \beta_p + O_{t,d,L}(N^d (\log\log N)^{-c}) \tag{1.6}$$

as $N \to \infty$, where $c = c_{t,d,L} > 0$ depends only on t, d, L, Λ is extended by zero to the integers, β_∞ is the Archimedean factor

$$\beta_\infty = \mathrm{vol}(\Omega \cap \Psi^{-1}(\mathbb{R}^t_{>0})),$$

and for each prime p, β_p is the local factor

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

(viewing each ψ_i also as an affine map from $(\mathbb{Z}/p\mathbb{Z})^d$ to $\mathbb{Z}/p\mathbb{Z}$).

Note that, in the language of [22], the assumption that $\dot{\psi}_i$ are pairwise linearly independent is equivalent to these forms having "finite Cauchy–Schwarz complexity".

In [22], the result of Theorem 1.6 was established with the qualitative error term $o^{\mathrm{ineff}}_{t,d,L}(N^d)$ in (1.6) (initially under the hypotheses of the Möbius and nilsequences conjecture and the inverse Gowers-norm conjecture, but these were later proved in [23, 27]).

In Section 9, we outline the (rather straightforward) details of the deduction of Theorem 1.6 from Theorem 1.4.

Example 1.7. In [22, Example 8] it is shown that the number of (increasing) arithmetic progressions of primes of a given length $k \geq 2$ in $[N]$ is equal to

$$\left(\frac{1}{2(k-1)}\prod_p \beta_p + o^{\mathrm{ineff}}(1)\right)\frac{N^2}{\log^k N}$$

where β_p is equal to $\frac{1}{p}(\frac{p}{p-1})^{k-1}$ when $p \leq k$, and $(1-\frac{k-1}{p})(\frac{p}{p-1})^{k-1}$ otherwise. Inserting Theorem 1.6 into the arguments from [22], the qualitative error term $o^{\mathrm{ineff}}(1)$ can now be improved to the doubly logarithmic error $O((\log\log N)^{-c})$. This is new for $k \geq 4$.

Another application of Theorem 1.6 is to sets containing no progressions with shifted prime difference.[3] It was shown by Sárközy [39] that (for N large) any subset of $[N]$ of size $\gg N$ contains a pattern of the form $x, x+p-1$ with p a prime. After several improvements [32, 37, 46], the current best known quantitative version of this, proved recently by Green [16], is that any subset of $[N]$ of size $\geq N^{1-c}$ contains a pattern of this form. Sárközy's theorem was later generalized to longer progressions by Frantzikinakis–Host–Kra [10], and Wooley–Ziegler [47], who showed that, for any $k \geq 3$ and N large enough in terms of k, any subset of $[N]$ of size $\gg N$ contains a pattern of the form $x, x+p-1, x+2(p-1), \dots, x+(k-1)(p-1)$ with p a prime, that is, a k-term arithmetic progression with shifted prime difference. These proofs however did not provide quantitative bounds for the density of a set avoiding k-term progressions with shifted prime difference. Using our main theorem, we can now obtain the first quantitative bound for this problem.

Theorem 1.8 (A quantitative bound for sets missing progressions with shifted prime difference). *Let $k \geq 3$, and let N be large enough in terms of k. Then any subset of $[N]$ of size $\geq N(\log\log\log\log N)^{-c}$ contains a k-term arithmetic progression whose common difference is a shifted prime of the form $p-1$. Moreover, if $k=4$, one can replace $N(\log\log\log\log N)^{-c}$ with $N(\log\log\log N)^{-c}$ above, and if $k=3$, one can replace it with $N\exp(-(\log\log\log N)^c)$.*

The proof of this is given in Section 10.

Remark 1.9. It is likely that one can similarly now make other qualitative consequences of (1.5) quantitative. Certainly the version of the generalized Hardy–Littlewood conjecture in [22, Conjecture 1.2] (in the finite complexity case) can now be made quantitative,

[3] We are indebted to Sean Prendiville for bringing this application to our attention.

with doubly logarithmic savings, in a manner perfectly analogous to Theorem 1.6, as can the version of the main theorem in [22, Theorem 1.8]; we omit the details. The more recent asymptotics on linear inequalities in primes in [45] are also likely to now have a doubly logarithmic quantitative version, but we do not pursue this matter here.

Lastly, one can also use Theorem 1.6 to quantify a result of the authors [42] on the logarithmically averaged Chowla conjecture for odd order correlations (whose proof relied on the Gowers uniformity of Λ). A back of the envelope calculation suggests that one could quantify the error term there, for fixed odd $k \geq 3$, to triply logarithmic; thus,

$$\frac{1}{\log x} \sum_{n \leq x} \frac{\mu(n+h_1)\mu(n+h_2)\cdots\mu(n+h_k)}{n} \ll (\log\log\log x)^{-c} \tag{1.7}$$

for any fixed integers $0 \leq h_1 < \cdots < h_k$ (and the same with the Liouville function in place of μ). Very briefly, by the entropy decrement argument [42, Theorem 3.1] one can locate a scale $\exp((\log\log x)^{1/2}) \leq P \leq \log x$ such that the left-hand side of (1.7) can be replaced up to a triply logarithmic error term with

$$\frac{(-1)^k}{\log\log P} \sum_{p \leq P} \frac{1}{p} \frac{1}{\log x} \sum_{n \leq x} \frac{\mu(n+ph_1)\mu(n+ph_2)\cdots\mu(n+ph_k)}{n}.$$

One would then split the p sum into dyadic scales and proceed as in [42] by replacing the average over primes p with an average over w-rough integers, using Theorem 1.6 and a quantitative version of the generalized von Neumann theorem as a substitute for Theorem 1.3, producing an admissible $O((\log\log P)^{-c})$ error term. The triply logarithmic error terms at this step are much worse than any other error terms arising in the rest of the proof, therefore leading to (1.7). We leave the details to the interested reader.

2. Discussion and set-up of the proof

Until recently, there were two main obstacles to achieving the sort of quantitative (and effective) bound stated in Theorem 1.4. Firstly, the first proofs of the inverse conjecture for the Gowers norms in the large k regime $k \geq 5$ were ineffective (using tools such as non-standard analysis) and did not provide any quantitative dependence of constants. Secondly, in order to overcome certain logarithmic losses in the estimates, it was necessary to invoke Siegel's theorem to control the correlation of the Möbius function with nilsequences, and the decay rate in the $o(1)$ bounds in Theorem 1.3 (iii) then depended on the rate at which the constants in Siegel's theorem $|L(1,\chi)| \gg_\varepsilon^{\text{ineff}} q^{-\varepsilon}$ depended on ε, which is completely ineffective with known methods.

The first issue was resolved recently with the quantitative inverse theorem of Manners [33], which provided a good quantitative dependence on all parameters in the inverse theory of Gowers norms. To resolve the second issue, we perform the technique of isolating out the contribution of a potential Siegel zero to obtain more refined approximations

$$\mu \approx \mu_{\text{Siegel}}(n), \quad \Lambda \approx \Lambda_{\text{Siegel}}(n)$$

to the arithmetic functions μ, Λ. To make this precise we introduce some notation:

Definition 2.1 (Siegel model). Recall that the quantity Q was defined in (1.1).

(i) We define a *Q-Siegel zero* to be a real number $1 - \frac{c_0}{\log Q} < \beta < 1$ for which there exists a primitive real Dirichlet character χ_{Siegel} (which we call the *Q-Siegel character*) of conductor $q_{\text{Siegel}} < Q$ such that $L(\beta, \chi_{\text{Siegel}}) = 0$, where $L(s, \chi)$ denotes the Dirichlet L-function associated to χ. Here c_0 is a sufficiently small absolute constant (and henceforth all implied constants are permitted to depend on c_0). Note from the Landau–Page theorem (see e.g. [35, Corollary 11.10]) that if a Q-Siegel zero exists, then it is unique (and similarly for the Q-Siegel character), and the zero β is simple (so that $L'(\beta, \chi_{\text{Siegel}}) \neq 0$).

(ii) We define the *Q-Siegel model* Λ_{Siegel} for the von Mangoldt function Λ to be

$$\Lambda_{\text{Siegel}}(n) := \Lambda_{\text{Cramér},Q}(n) = \frac{P(Q)}{\phi(P(Q))} 1_{(n,P(Q))=1}$$

if no Q-Siegel zero exists, and

$$\Lambda_{\text{Siegel}}(n) := \Lambda_{\text{Cramér},Q}(n)(1 - n^{\beta-1}\chi_{\text{Siegel}}(n))$$

otherwise.

(iii) We define the *Q-Siegel model* μ_{Siegel} for the Möbius function μ to be

$$\mu_{\text{Siegel}}(n) := 0$$

if no Q-Siegel zero exists, and

$$\mu_{\text{Siegel}}(n) := \mu_{\text{local}} * \mu'(n) \tag{2.1}$$

otherwise, where μ_{local} is the local Möbius function

$$\mu_{\text{local}}(n) := \mu(n) 1_{n|P(Q)},$$

μ' is the function

$$\mu'(n) := \alpha n^{\beta-1} \chi_{\text{Siegel}}(n) 1_{(n,P(Q))=1} = \alpha \frac{\phi(P(Q))}{P(Q)} (\Lambda_{\text{Cramér},Q}(n) - \Lambda_{\text{Siegel}}(n)), \tag{2.2}$$

α is the quantity

$$\alpha := \frac{1}{L'(\beta, \chi_{\text{Siegel}})} \prod_{p<Q} \left(1 - \frac{1}{p}\right)^{-1} \left(1 - \frac{\chi_{\text{Siegel}}(p)}{p^\beta}\right)^{-1}, \tag{2.3}$$

and $\mu_{\text{local}} * \mu'$ is the Dirichlet convolution of μ_{local} and μ':

$$\mu_{\text{local}} * \mu'(n) := \sum_{d|n} \mu_{\text{local}}(d) \mu'(n/d).$$

(Note from the supports of μ_{local}, μ' that at most one term in this sum is non-zero for any given n.)

The significance of these models is that Λ and $\Lambda_{\mathrm{Siegel}}$ have very nearly the same statistics on arithmetic progressions (with error terms that improve over the main term by pseudopolynomial factors $O(\exp(-\log^c N))$, which are superior to the strongly logarithmic gains $O_A^{\mathrm{ineff}}(\log^{-A} N)$ provided by the Siegel–Walfisz theorem), and similarly for μ and μ_{Siegel}. Indeed, in Section 7 we will show the following estimates:

Proposition 2.2 (Pseudopolynomial equidistribution in arithmetic progressions). *For any arithmetic progression $P \subset [N]$, we have*

$$\sum_{n\in P}(\mu(n) - \mu_{\mathrm{Siegel}}(n)) \ll N\exp(-c\log^{1/10} N) \tag{2.4}$$

and

$$\sum_{n\in P}(\Lambda(n) - \Lambda_{\mathrm{Siegel}}(n)) \ll N\exp(-c\log^{1/10} N). \tag{2.5}$$

Remark 2.3. The construction of μ_{Siegel} appears to be complicated, but it is a multiplicative construction and can be justified as follows. If χ is any character induced from χ_{Siegel} of some period $q|[q_{\mathrm{Siegel}}, P(Q)]$, a short calculation reveals the Euler products

$$\sum_{n=1}^{\infty}\frac{\mu_{\mathrm{Siegel}}\chi(n)}{n^s} = \frac{\zeta(s+1-\beta)}{L'(\beta,\chi_{\mathrm{Siegel}})}\prod_{\substack{p<Q\\ p|q}}\frac{1-\frac{\chi_{\mathrm{Siegel}}(p)}{p^s}}{1-\frac{\chi_{\mathrm{Siegel}}(p)}{p^\beta}}\prod_{p<Q}\frac{1-\frac{1}{p^{s+1-\beta}}}{1-\frac{1}{p}}$$
$$\times\prod_{p|q}\left(1-\frac{\chi_{\mathrm{Siegel}}(p)}{p^\beta}\right)^{-1} \tag{2.6}$$

and

$$\sum_{n=1}^{\infty}\frac{\mu\chi(n)}{n^s} = \frac{1}{L(s,\chi_{\mathrm{Siegel}})}\prod_{p|q}\left(1-\frac{\chi_{\mathrm{Siegel}}(p)}{p^s}\right)^{-1} \tag{2.7}$$

whenever $\mathrm{Re}(s) > 1$. One can then check that the meromorphic continuations of the two Dirichlet series (2.6), (2.7) both have a simple pole at $s = \beta$ with the same residue (and when χ is not induced from χ_{Siegel} there is no such pole), which helps justify why we expect μ_{Siegel} to be a good approximation to μ. We experimented with simpler models to μ than μ_{Siegel}, but in order to get the pseudopolynomial error terms $\exp(-\log^c N)$ in (2.4) it seems essential that the model μ_{Siegel} behaves almost identically to μ with respect to primes p as large as $\exp(\log^c N)$, which necessitates a complicated construction such as (2.1). We remark that a similar (though slightly less refined) approximant $\lambda_{\mathrm{Siegel}}$ to the Liouville function λ was introduced by Germán and Katai [12], and recently used in [2] to establish Chowla's conjecture in the presence of a Siegel zero.

For future reference we also observe the following crude pointwise bounds on Λ, μ and their approximate models:

Lemma 2.4 (Pointwise bounds). *For $n \in [N]$ and $2 \le w \le Q$, one has*

$$\Lambda(n), \Lambda_{\mathrm{Cramér},w}(n), \Lambda_{\mathrm{Siegel}}(n) \ll \log N$$

and

$$\mu(n), \mu_{\mathrm{Siegel}}(n) \ll 1.$$

Proof. All of these bounds are either trivial or immediate consequences of Mertens' theorem, except for the bound on μ_{Siegel}, which would follow if the quantity α in (2.3) were bounded. This turns out to follow from standard bounds on the L-function $L(s, \chi_{\text{Siegel}})$ near a Q-Siegel zero β; see Lemma 5.5. ∎

In view of Proposition 1.2 and the triangle inequality (1.2), Theorem 1.4 then follows from the following two statements.

Theorem 2.5 (Siegel corrections are logarithmically Gowers uniform). *We have*

$$\|\mu_{\text{Siegel}}\|_{U^k[N]} \ll q_{\text{Siegel}}^{-c} \ll \log^{-c} N \tag{2.8}$$

and

$$\|\Lambda_{\text{Siegel}} - \Lambda_{\text{Cramér},Q}\|_{U^k[N]} \ll q_{\text{Siegel}}^{-c} \ll \log^{-c} N \tag{2.9}$$

with the convention that the expression q_{Siegel}^{-c} vanishes when no Q-Siegel zero exists.

Theorem 2.6 (Doubly logarithmic uniformity of Möbius and von Mangoldt, II). *We have*

$$\|\mu - \mu_{\text{Siegel}}\|_{U^k[N]} \ll (\log\log N)^{-c} \tag{2.10}$$

and

$$\|\Lambda - \Lambda_{\text{Siegel}}\|_{U^k[N]} \ll (\log\log N)^{-c}. \tag{2.11}$$

Theorem 2.5 is an application of sieve-theoretic methods, smooth number estimates and the Weil bound, and is established in Section 5.2. The main difficulty is to establish Theorem 2.6. In principle, one can directly apply the quantitative inverse theory of Manners [33], and reduce matters to controlling the correlation of $\mu - \mu_{\text{Siegel}}$, $\Lambda - \Lambda_{\text{Siegel}}$ with nilsequences arising from nilmanifolds (although in the case of $\Lambda - \Lambda_{\text{Siegel}}$ we have the obstacle that the function is unbounded – the resolution of this is discussed below). Indeed, in Section 7 we will establish the following bounds that significantly extend the bounds in Proposition 2.2:

Theorem 2.7 (Pseudopolynomial orthogonality of Möbius and von Mangoldt with nilsequences). *Let $\epsilon > 0$ and $k \geq 1$. Let $c_1(\epsilon) > 0$ be small enough in terms of ϵ. Then we have the bounds*

$$\sum_{n\in P}(\mu - \mu_{\text{Siegel}})(n)\overline{F}(g(n)\Gamma) \ll_\epsilon N \exp(-\log^{1/10-\epsilon} N) \tag{2.12}$$

and

$$\sum_{n\in P}(\Lambda - \Lambda_{\text{Siegel}})(n)\overline{F}(g(n)\Gamma) \ll_\epsilon N \exp(-\log^{1/10-\epsilon} N) \tag{2.13}$$

whenever $P \subset [N]$ is an arithmetic progression, G/Γ is a filtered nilmanifold of degree $k-1$, dimension at most $(\log\log N)^{c_1(\epsilon)}$, and complexity at most $\exp(\log^{c_1(\epsilon)} N)$, $F : G/\Gamma \to \mathbb{C}$ is a 1-bounded Lipschitz function[4] of Lipschitz constant at most

[4] A function $F\colon X \to \mathbb{C}$ is 1-*bounded* if $|F(x)| \leq 1$ for all $x \in X$. More generally, given any $\nu\colon X \to \mathbb{R}^+$, we say that $F\colon X \to \mathbb{C}$ is ν-*bounded* if $|F(x)| \leq \nu(x)$ for all $x \in X$.

$\exp(\log^{1/10-\epsilon} N)$, *and* $g\colon \mathbb{Z} \to G$ *is a polynomial map.* (*The relevant definitions of filtered nilmanifolds, etc., are reviewed in Definition* 6.1.)

Remark 2.8. If one redefined the Siegel models $\mu_{\mathrm{Siegel}}, \Lambda_{\mathrm{Siegel}}$ by assigning the parameter Q the larger value $\exp(\log^{1/2} N)$, one could inspect that the exponent of the logarithm in (2.12) and (2.13) (and in particular in Proposition 2.2) could be increased to $1/2-\epsilon$, hence essentially matching the shape of the error term in the classical prime number theorem. For this modification, one would have to tweak the exponents in Section 5 a little; in particular, in Proposition 5.2 the exponents $3/5$ and $4/5$ would have to be replaced with $1/2$. As the precise value of the exponent has very little influence on our bounds, we leave the details of this strengthening to the interested reader.

For the sake of comparison, in [23] the strongly logarithmic bound

$$\mathbb{E}_{n\in[N]}\mu(n)\overline{F}(g(n)\Gamma) \ll_{A,M}^{\mathrm{ineff}} \log^{-A} N$$

was established for any $A > 0$ assuming that the dimension and complexity of G/Γ and the Lipschitz constant of F were all bounded by M; using this bound, in [22] the qualitative bound

$$\mathbb{E}_{n\in[N]}\left(\frac{\phi(W)}{W}\Lambda(Wn+b)-1\right)\overline{F}(g(n)\Gamma) = o^{\mathrm{ineff}}(1)$$

was shown for the same type of nilsequences $F(g(n)\Gamma)$, where $W = P(w)$ for some $w = w(N)$ growing sufficiently slowly to infinity with N and any $b \in [W]$ coprime to W. With a little additional effort, the latter bound then also implies the qualitative bound

$$\sum_{n\in P}(\Lambda - \Lambda_{\mathrm{Cramér},w})(n)\overline{F}(g(n)\Gamma) = o^{\mathrm{ineff}}(N)$$

for these nilsequences and arbitrary arithmetic progressions $P \subset [N]$. The arguments relied upon (and in fact imply) the Siegel–Walfisz theorem and thus could not give error terms better than strongly logarithmic, which would be unsuitable for our applications (particularly those involving the von Mangoldt function). It is therefore necessary to account for the correction terms $\mu_{\mathrm{Siegel}}, \Lambda_{\mathrm{Siegel}} - \Lambda_{\mathrm{Cramér},Q}$ to avoid any appeal to the Siegel–Walfisz theorem and to improve the bounds to be of pseudopolynomial type, despite the fact (from Theorem 2.5) that these correction terms are already logarithmically small in the Gowers norm sense.

Our proof of Theorem 2.7 will broadly follow the same strategy as that in [23], relying on Proposition 2.2 in the "major arc" case and on decomposition into "Type I" and "Type II" sums, followed by Cauchy–Schwarz and an appeal to the equidistribution theory of nilmanifolds, in the "minor arc" case. A key new feature, compared to previous work, is that the dimension of the nilsequences is no longer bounded, but grows at a roughly doubly logarithmic rate in N. Because of this, we are forced to perform a careful accounting on the dependence on dimension in the aforementioned equidistribution theory, and in particular ensure that the bounds only depend at most doubly exponentially on the dimension. This is in fact one of the main reasons why our bounds in Theorem 1.4 are limited to be doubly logarithmic in nature; see Remarks 2.9 and 6.4 below.

The estimate (2.10) can be directly obtained from (2.12) using the inverse theorem of Manners [33], which we review in Section 6; note that this theorem basically applies a double logarithm to the quantitative bounds, which is why the pseudopolynomial type terms in Theorem 2.7 are reduced to doubly logarithmic type terms in Theorem 2.6. For the von Mangoldt estimate (2.11), we encounter the familiar problem that $\Lambda - \Lambda_{\mathrm{Siegel}}$ is not bounded (see Lemma 2.4), so that Manners' quantitative inverse theorem does not immediately apply. In [22], this difficulty was resolved at the qualitative level by first using the "W-trick" of passing to an arithmetic progression $\{Wn + b : n \in \mathbb{N}\}$ for some $W = P(w)$ and some w growing slowly with N, and then dominating (an appropriately normalized version of) the von Mangoldt function on that progression by a divisor sum ν of Goldston–Yıldırım type that obeyed some "pseudorandomness" conditions. This enabled one to then apply a transference principle that roughly speaking allowed one to behave "as if" the normalized von Mangoldt function was bounded on this progression, at least for the purposes of applying an inverse theorem for the Gowers norms.

Here the biggest source of quantitative inefficiency is the transference principle, as the first few proofs of this principle [14, 19, 22, 36] involved the Weierstrass approximation theorem, quantitative versions of which can generate exponential type losses. However, in [5] (see also [4]), Conlon, Fox, and Zhao introduced the method of *densification*, which they used to obtain a transference principle in the context of Szemerédi-type theorems that involved only polynomial dependencies on the bounds (and they also relaxed the pseudorandomness hypotheses on the enveloping sieve ν by dropping the so-called "correlation condition"). As it turns out, the densification method can be adapted to inverse theorems as well with efficient quantitative bounds, at least when the correlation in the inverse theorem enjoys polynomial bounds; we formalize this observation (which seems to be of independent interest) as Theorem 8.1. Fortunately for us, the arguments of Manners in [33, Section 5] already provide such a polynomial bound. Using our quantitative transference result for the inverse theorem, it becomes a relatively routine matter to derive (2.11) from (2.13), after making various necessary quantitative refinements (for instance, the parameter w will now be taken to be of the shape $\log^{\varepsilon} N$ for some small $\varepsilon > 0$, rather than growing in some unspecified slow fashion with N). This will all be performed in Section 8.

Remark 2.9. Perhaps surprisingly, the bounds in Theorem 1.4 are not significantly improved if one assumes the generalized Riemann hypothesis; some pseudopolynomial bounds can now be sharpened to polynomial bounds (such as Theorem 2.7), but for the logarithmic and doubly logarithmic bounds only minor improvements in the unspecified constants c are available under GRH (though of course in this case any terms involving Q-Siegel zeroes can simply be deleted). On the other hand, it is tempting to conjecture that the doubly logarithmic bounds in our main results can be improved to logarithmic, given that several of the key estimates already have this quality of error term or better. This is particularly appealing in the $k = 3$ case where we have quite a good inverse U^3 theorem [18]. The main difficulty is that to achieve this goal, it appears that one needs an equidistribution theory for 2-step nilmanifolds (or quadratic bracket polynomials) that

involves exponents that are merely *polynomial* in the dimension of the nilmanifold (or complexity of the bracket polynomial) rather than exponential. In analogy with the well known quadratic Diophantine approximation theory of Schmidt [40], it seems reasonable to expect such a theory to be feasible,[5] but we will not pursue this matter here. On the other hand, we note that by combining Theorem 2.7 with the circle method one can obtain the pseudopolynomial bounds

$$\|\mu - \mu_{\mathrm{Siegel}}\|_{U^2[N]}, \|\Lambda - \Lambda_{\mathrm{Siegel}}\|_{U^2[N]} \ll \exp(-c \log^c N),$$

and one could optimistically conjecture that such pseudopolynomial (or even polynomial) bounds are also true for higher Gowers norms as well (such bounds would follow from a sufficiently uniform version of the Hardy–Littlewood prime tuples conjecture).

3. Notation

As stated in the introduction, throughout this paper we fix an integer $k \geq 1$, and assume N is a positive real number that is sufficiently large depending on k (and Q is given in terms of N by (1.1)). We abbreviate $\{n \in \mathbb{N} : 1 \leq n \leq N\}$ as $[N]$ (even when N is not an integer).

We use the asymptotic notation $X \ll Y$, $Y \gg X$, or $X = O(Y)$ to denote an estimate of the form $|X| \leq CY$ for some constant $C > 0$. If C depends on additional parameters, we indicate this by subscripts, for instance $X = O_d(Y)$ denotes the estimate $|X| \leq C_d Y$ for some $C_d > 0$ depending on d. However, as all of our constants will depend on the fixed parameter k, we omit this parameter from this subscripting notation. Unless otherwise specified, the constants will depend in an effective fashion on the parameters; on the rare occasions (mostly involving citing previous literature) in which ineffective constants are used, we will use the superscript ineff to indicate this. We write $X \asymp Y$ as an abbreviation for $X \ll Y \ll X$, subject to the same subscripting and superscripting conventions as before. If X, Y depend on an additional parameter N, we write $X = o(Y)$ as $N \to \infty$ to denote the claim that $|X| \leq c(N)Y$ for some quantity $c(N)$ that goes to zero as $N \to \infty$, again subject to the same subscripting and superscripting conventions as before. As stated in the introduction, we use c to denote various small positive constants depending on k that can vary from line to line.

We often refer to the following hierarchy of decay estimates, in increasing order of strength:

- *qualitative (and ineffective) decay*, in which $X = o^{\mathrm{ineff}}(Y)$ as $N \to \infty$;
- *doubly logarithmic decay*, in which $X \ll (\log\log N)^{-c} Y$;

[5] Another option is to exploit the improved dimension bounds for the inverse U^3 theory now available [38], using the equivalences from [21]. Since the initial release of this article as a preprint, this option has in fact been carried out by Leng [31], who significantly improved the $(\log\log N)^{-c}$ type bounds in Theorem 2.6 to $\exp(-\log^c N)$ type bounds in the $k = 3$ case.

- *logarithmic decay*, in which $X \ll (\log N)^{-c} Y$;
- *strongly (but ineffectively) logarithmic decay*, in which $X \ll_A^{\text{ineff}} (\log N)^{-A} Y$ for any $A > 0$ (this is a typical shape for bounds obtained using the Siegel–Walfisz theorem);
- *pseudopolynomial decay*, in which $X \ll \exp(-c \log^c N) Y$;
- *polynomial decay*, in which $X \ll N^{-c} Y$.

As the terminology suggests, pseudopolynomial decay will be a satisfactory substitute for polynomial decay in many of our arguments.

We use 1_E to denote the indicator function of a set E, thus $1_E(n)$ equals 1 when $n \in E$ and 0 otherwise. We also use 1_S to denote the indicator of a statement S, thus 1_S equals 1 when S is true and 0 otherwise.

If A is a finite set, we use $\#A$ to denote its cardinality.

All sums and products over the variable p are understood to be over primes, and similarly all sums and products over variables such as n or d are understood to be over natural numbers, unless otherwise indicated.

4. Some lemmas on Gowers norms

We state here a few lemmas concerning the Gowers norms that will be used later on.

In addition to the triangle inequality (1.2), we shall also often use the closely related *Gowers–Cauchy–Schwarz inequality*

$$\left|\mathbb{E}_{(x,\vec{h})\in G^{k+1}} \prod_{\omega\in\{0,1\}^k} f_\omega(x+\omega\cdot\vec{h})\right| \le \prod_{\omega\in\{0,1\}^k} \|f_\omega\|_{U^k(G)} \tag{4.1}$$

for any finite additive group G and any functions $f_\omega\colon G \to \mathbb{C}$ for $\omega \in \{0,1\}^k$; see for instance [22, Lemma B.2]. For arbitrary additive groups, we also have the non-normalized variant

$$\left|\sum_{(x,\vec{h})\in G^{k+1}} \prod_{\omega\in\{0,1\}^k} f_\omega(x+\omega\cdot\vec{h})\right| \le \prod_{\omega\in\{0,1\}^k} \|f_\omega\|_{\tilde{U}^k(G)}. \tag{4.2}$$

Observe that the Gowers norms behave well with respect to tensor products: if $f_1 : G_1 \to \mathbb{C}$, $f_2\colon G_2 \to \mathbb{C}$ are finitely supported functions on additive groups G_1, G_2, then a short computation reveals that

$$\|f_1 \otimes f_2\|_{\tilde{U}^k(G_1\times G_2)} = \|f_1\|_{\tilde{U}^k(G_1)} \|f_2\|_{\tilde{U}^k(G_2)} \tag{4.3}$$

for any $k \ge 1$.

We now develop a variant of this identity. We localize the Gowers norm to cosets $a + H$ of a subgroup H of an additive group G as follows: if $k \ge 1$ and $f\colon G \to \mathbb{C}$ is finitely supported, we define $\|f\|_{\tilde{U}^k(a+H)} := \|f(a+\cdot)\|_{\tilde{U}^k(H)}$, and similarly $\|f\|_{U^k(a+H)} := \|f(a+\cdot)\|_{U^k(H)}$ if H is finite. Note that this definition does not depend on the choice of coset representative. We have the following convenient Fubini type inequality (which is reasonably well known "folklore", although the only explicit prior reference to such an inequality that we are aware of is [3, Lemma 4.3]):

Lemma 4.1 (Fubini type inequality). *Let $k \geq 1$, let G be an additive group, let H be a subgroup of G, and let $f : G \to \mathbb{C}$ be a finitely supported function. For each coset $a + H$ in the quotient group G/H, let $F(a + H)$ denote the quantity*

$$F(a + H) := \|f\|_{\tilde{U}^k(a+H)};$$

note that $F : G/H \to \mathbb{C}$ is also a finitely supported function. Then

$$\|f\|_{\tilde{U}^k(G)} \leq \|F\|_{\tilde{U}^k(G/H)}. \tag{4.4}$$

Informally, this lemma asserts that to bound the $U^k(G)$ norm of a function f, one can first evaluate the $\tilde{U}^k$ norm along the various cosets of H, and then compute the $\tilde{U}^k$ norm of the numbers obtained in that fashion. If G, H are finite we can obtain similar claims for the normalized U^k norms in the obvious fashion. Note that the Fubini–Tonelli theorem establishes a similar claim for the ℓ^1 (or more generally ℓ^p) norms (and in this case one has equality in (4.4) instead of inequality. One can also verify that (4.4) is consistent with (4.3).

Proof of Lemma 4.1. From Definition 1.1 we have

$$\|f\|^{2^k}_{\tilde{U}^k(G)} = \sum_{(n,\vec{h}) \in G^{k+1}} \prod_{\omega \in \{0,1\}^k} \mathcal{C}^{|\omega|} f(n + \omega \cdot \vec{h}).$$

Consider the contribution to the right-hand side where n lies in a coset $a + H$ and h_i lies in a coset $b_i + H$ for $i = 1, \ldots, k$. By the Gowers–Cauchy–Schwarz inequality (4.2), this contribution can be bounded in magnitude by

$$\prod_{\omega \in \{0,1\}^k} F(a + \omega \cdot \vec{b} + H)$$

where $\vec{b} := (b_1, \ldots, b_k)$. Summing over all choices of $a, \vec{b}$ and applying Definition 1.1 again, we conclude that

$$\|f\|^{2^k}_{\tilde{U}^k(G)} \leq \|F\|^{2^k}_{\tilde{U}^k(G/H)},$$

giving (4.4). ∎

As a corollary of this inequality, we can estimate the Gowers norm of a function on $[N]$ in terms of its values on various arithmetic progressions:

Corollary 4.2 (W-trick). *Let $1 \leq W \leq N/10$, and let $f : [N] \to \mathbb{C}$ be a function supported on the set $\{n \in [N] : (n, W) = 1\}$ that obeys the bounds*

$$\left\| \frac{\phi(W)}{W} f(W \cdot + b) \right\|_{U^k[\frac{N-b}{W}]} \leq A$$

for all $b \in [W]$ coprime to W and some $A > 0$. Then

$$\|f\|_{U^k[N]} \ll A.$$

Proof. We extend f by zero to the integers $\mathbb{Z}$ and work with the unnormalized Gowers norms. Since

$$\|1_{[N]}\|_{\tilde{U}^k(\mathbb{Z})} \asymp N^{\frac{k+1}{2^k}}$$

and

$$\|1_{[\frac{N-b}{W}]}\|_{\tilde{U}^k(\mathbb{Z})} \asymp (N/W)^{\frac{k+1}{2^k}}$$

we have

$$\|f\|_{\tilde{U}^k(W\mathbb{Z}+b)} \ll \frac{W}{\phi(W)} A (N/W)^{\frac{k+1}{2^k}}$$

for all $b \in [W]$ coprime to W, and it will suffice to show that

$$\|f\|_{\tilde{U}^k(\mathbb{Z})} \ll A N^{\frac{k+1}{2^k}}.$$

Applying Lemma 4.1 with $G = \mathbb{Z}$ and $H = W\mathbb{Z}$, and normalizing the Gowers norms, it suffices to show that

$$\left\| \frac{W}{\phi(W)} 1_{(\cdot,W)=1} \right\|_{U^k(\mathbb{Z}/W\mathbb{Z})} \ll 1.$$

Expressing W as the product of primes $p^{v_p(W)}$ and using the Chinese remainder theorem and (4.3) repeatedly, the left-hand side can be written as

$$\prod_p \left\| \frac{p}{p-1} 1_{(\cdot,p)=1} \right\|_{U^k(\mathbb{Z}/p^{v_p(W)}\mathbb{Z})}.$$

However, direct computation using the inclusion-exclusion principle shows that

$$\|1_{(\cdot,p)=1}\|^{2^k}_{U^k(\mathbb{Z}/p^{v_p(W)}\mathbb{Z})} = 1 - \frac{2^k}{p} + O_k\left(\frac{1}{p^2}\right),$$

and hence

$$\left\| \frac{p}{p-1} 1_{(\cdot,p)=1} \right\|_{U^k(\mathbb{Z}/p^{v_p(W)}\mathbb{Z})} = 1 + O_k\left(\frac{1}{p^2}\right).$$

The claim follows. ∎

Next, we give a variant of the triangle inequality that estimates a Gowers norm based on the greatest common divisor with a fixed modulus.

Lemma 4.3 (Variant of triangle inequality). *Let $N \geq 100$, let $1 \leq q \leq N$, and let $k \geq 1$ be an integer. Let $f : [N] \to [-1,1]$ be a function. Then*

$$\|f\|^{2^k}_{U^k[N]} \ll \sum_{d|q} \frac{1}{d} \|f(d\cdot) 1_{(\cdot,q/d)=1}\|_{U^k[N/d]}.$$

The key point here is the presence of the factor $1/d$, which ensures that the summation over d can be estimated manageably.

Proof of Lemma 4.3. We extend f by zero outside of $[N]$. From Definition 1.1, it suffices to show the unnormalized estimate

$$\sum_{(n,\vec{h})\in\mathbb{Z}^{k+1}}\prod_{\omega\in\{0,1\}^k} f(n+\omega\cdot\vec{h}) \ll N^{k+1}\sum_{d|q}\frac{1}{d(N/d)^{\frac{k+1}{2^k}}}\|f1_{(\cdot,q)=d}\|_{\tilde{U}^k(d\mathbb{Z})}.$$

The left-hand side can be written as

$$\sum_{n\in\mathbb{Z}} f(n)F(n)$$

where the *dual function* $F(n)$ is defined as

$$F(n) := \sum_{\vec{h}\in\mathbb{Z}^k}\prod_{\omega\in\{0,1\}^k\setminus\{0\}^k} f(n+\omega\cdot\vec{h}).$$

We split this sum in terms of the value of (n,q) as

$$\sum_{n\in\mathbb{Z}} f(n)F(n) = \sum_{d|q}\sum_{n\in d\mathbb{Z}} f(n)1_{(n,q)=d}F(n).$$

By the triangle inequality, it thus suffices to show that

$$\sum_{n\in d\mathbb{Z}} f(n)1_{(n,q)=d}F(n) \ll \frac{N^{k+1}}{d(N/d)^{\frac{k+1}{2^k}}}\|f1_{(\cdot,q)=d}\|_{\tilde{U}^k(d\mathbb{Z})}$$

for each $d\mid q$. Decomposing $h_1,\dots,h_k$ in the definition of $F(n)$ into cosets mod d, the left-hand side may be written as

$$\sum_{\vec{b}\in[d]^k}\sum_{(n,\vec{h})\in(d\mathbb{Z})^{k+1}}\prod_{\omega\in\{0,1\}^k} f(n+\omega\cdot(\vec{h}+\vec{b}))1_{(n,q)=d}.$$

By the Gowers–Cauchy–Schwarz inequality (4.1), and noting that f is bounded by $1_{[N]}$, we have

$$\sum_{(n,\vec{h})\in(d\mathbb{Z})^{k+1}}\prod_{\omega\in\{0,1\}^k} f(n+\omega\cdot(\vec{h}+\vec{b}))1_{(n,q)=d} \ll \|f1_{(\cdot,q)=d}\|_{\tilde{U}^k(d\mathbb{Z})}\big((N/d)^{\frac{k+1}{2^k}}\big)^{2^k-1}.$$

Summing over all the d^k choices of $\vec{b}$, we thus obtain

$$\sum_{n\in d\mathbb{Z}} f(n)1_{(n,q)=d}F(n) \ll \|f1_{(\cdot,q)=d}\|_{\tilde{U}^k(d\mathbb{Z})}d^k\big((N/d)^{\frac{k+1}{2^k}}\big)^{2^k-1}$$

and the claim follows after a little algebra. ∎

5. Some sieve theory

5.1. The Cramér model

In this section we use some standard sieve-theoretic tools to establish several estimates involving the Cramér models $\Lambda_{\text{Cramér},w}$, some of which will also be useful in controlling the Siegel models $\Lambda_{\text{Siegel}}, \mu_{\text{Siegel}}$ in later sections.

We first recall a form of the fundamental lemma of sieve theory (arising from an analysis of the beta sieve).

Lemma 5.1 (Fundamental lemma of sieve theory). *Let $(a_n)_{n\in\mathbb{Z}}$ be a collection of non-negative reals, let $\kappa > 0$, $z \geq 2$, and $D \geq z^{9\kappa+1}$. Let $g\colon \mathbb{N} \to [0,1)$ be a multiplicative function obeying the estimates*

$$\prod_{w\leq p<z} (1-g(p))^{-1} \leq K\Big(\frac{\log z}{\log w}\Big)^{\kappa} \tag{5.1}$$

for all $2 \leq w \leq z$ and some $K > 0$. Suppose that for every $d \leq D$ dividing $P(z)$ one has the formula

$$\sum_{d|n} a_n = Xg(d) + r_d \tag{5.2}$$

for some $X > 0$ and some remainder r_d. Then

$$\sum_{\substack{n\\(n,P(z))=1}} a_n = X\Big(\prod_{p<z}(1-g(p))\Big)(1+O(e^{9\kappa-s}K^{10})) + O\Big(\sum_{\substack{d\leq D\\ d|P(z)}} |r_d|\Big)$$

where $s := \frac{\log D}{\log z}$.

Proof. See [11, Theorem 6.9]. ∎

In our applications, the ratio $s = \frac{\log D}{\log z}$ will grow at a logarithmic rate, leading to pseudopolynomial accuracy when applying the fundamental lemma.

Using the fundamental lemma we can obtain satisfactory estimates (with pseudopolynomial accuracy) for counting linear equations in the Cramér model (compare with Theorem 1.6).

Proposition 5.2 (Linear equations in the Cramér model). *Let $t, m \geq 1$ be integers, and let $N \geq 100$. Let Ω be a convex subset of the cube $[-N,N]^d$, and let $\psi_1, \dots, \psi_t\colon \mathbb{Z}^m \to \mathbb{Z}$ be linear forms*

$$\psi_i(\vec{n}) = \vec{n}\cdot\dot{\psi}_i + \psi_i(0)$$

for some $\dot{\psi}_i \in \mathbb{Z}^m$ and $\psi_i(0) \in \mathbb{Z}$. Assume that the linear coefficients $\dot{\psi}_1, \dots, \dot{\psi}_t \in \mathbb{Z}^m$ are all pairwise linearly independent and have magnitude at most $\exp(\log^{3/5} N)$ (say). Then for any $2 \leq z \leq Q$, one has

$$\sum_{\vec{n}\in\Omega\cap\mathbb{Z}^m} \prod_{i=1}^{t} \Lambda_{\text{Cramér},z}(\psi_i(\vec{n})) = \text{vol}(\Omega)\prod_{p<z}\beta_p + O_{t,m}(N^m \exp(-c\log^{4/5} N))$$

for some $c > 0$ depending only on t, m, where for each p, β_p is the local factor

$$\beta_p := \mathbb{E}_{\vec{n} \in (\mathbb{Z}/p\mathbb{Z})^m} \prod_{i=1}^{t} \frac{p}{p-1} 1_{\psi_i(\vec{n}) \neq 0},$$

where ψ_i is also viewed as a map from $(\mathbb{Z}/p\mathbb{Z})^m$ to $\mathbb{Z}/p\mathbb{Z}$ in the obvious fashion.

Proof. Without loss of generality we may assume that N is sufficiently large depending on t, m; we now allow all implied constants to depend on t, m.

For any d dividing $P(z)$, let $g(d) \in [0,1]$ denote the quantity

$$g(d) := \mathbb{E}_{\vec{n} \in (\mathbb{Z}/d\mathbb{Z})^m} 1_{\prod_{i=1}^{t} \psi_i(\vec{n}) = 0},$$

with the convention that $g(d) = 0$ if d does not divide $P(z)$. In particular, we have

$$g(p) = 1 - \left(\frac{p-1}{p}\right)^t \beta_p \tag{5.3}$$

for all $p < z$. From the Chinese remainder theorem we see that g is multiplicative. Suppose first that $g(p) = 1$ for some $p < z$, then $\beta_p = 0$ and $\prod_{i=1}^{k} \Lambda_{\text{Cramér},z}(\psi_i(n))$ is identically zero. Thus the proposition is trivial in this case, so we may assume that $g(p) < 1$ for all p. From the construction we then have the crude bound

$$g(p) \leq 1 - \frac{1}{p^m}. \tag{5.4}$$

Also, from the construction we see that for any two distinct linear forms ψ_i, ψ_j, there is a positive integer $A_{ij} = \exp(O(\log^{3/5} N))$ such that $\dot{\psi}_i, \dot{\psi}_j$ are linearly independent in $(\mathbb{Z}/p\mathbb{Z})^k$ whenever p does not divide A_{ij} (indeed, one can take A_{ij} to be one of the non-zero coefficients of the wedge product of $\dot{\psi}_i$ and $\dot{\psi}_j$). If we let $A = \exp(O(\log^{3/5} N))$ be the product of all the A_{ij}, we conclude in particular that

$$\mathbb{E}_{n \in (\mathbb{Z}/p\mathbb{Z})^m} 1_{\psi_i(\vec{n}) = \psi_j(\vec{n}) = 0} \leq \frac{1}{p^2}$$

whenever p does not divide A, hence by using the inclusion-exclusion formula (or the Bonferroni inequalities) we have

$$g(p) = \frac{t}{p} + O\left(\frac{1}{p^2}\right) \tag{5.5}$$

whenever p does not divide A. In particular, we have

$$(1 - g(p))^{-1} = \left(\frac{p}{p-1}\right)^t \left(1 + O\left(\frac{1}{p^2}\right)\right)$$

unless p divides A (using (5.4) to handle the case when p is bounded). For p dividing A, (5.4) instead gives $(1 - g(p))^{-1} \leq p^m$. We conclude that for any $2 \leq w \leq z$, we have

$$\prod_{w \leq p < z} (1 - g(p))^{-1} \ll \left(\prod_{p|A} p\right)^m \prod_{w \leq p < z} \left(\frac{p}{p-1}\right)^t \leq A^m \prod_{w \leq p < z} \left(\frac{p}{p-1}\right)^t$$

and hence by Mertens' theorem the axiom (5.1) is obeyed with $\kappa = t$ and some $K = O(\exp(O(\log^{3/5} N)))$.

We introduce the sequence

$$a_n := \sum_{\vec{n}\in\Omega\cap\mathbb{Z}^m} 1_{\prod_{i=1}^t \psi_i(\vec{n})=n}.$$

Observe that the a_n are non-negative with

$$\sum_{\vec{n}\in\Omega\cap\mathbb{Z}^m} \prod_{i=1}^t \Lambda_{\text{Cramér},z}(\psi_i(n)) = \left(\prod_{p<z} \frac{p}{p-1}\right)^t \sum_{\substack{n \\ (n,P(z))=1}} a_n.$$

Set

$$D := \exp(\log^{9/10} N).$$

For any $d \le D$ dividing $P(z)$, we have

$$\sum_{n\equiv 0\ (d)} a_n = \sum_{\vec{n}\in\Omega\cap\mathbb{Z}^m} 1_{d\mid\prod_{i=1}^t \psi_i(\vec{n})}.$$

The condition $d \mid \prod_{i=1}^t \psi_i(\vec{n})$ restricts d to $g(d)d^m$ cosets of $(d\mathbb{Z})^m$. Applying a volume packing argument using [22, Corollary A.2] gives

$$\sum_{\vec{n}\in\Omega\cap\mathbb{Z}^m} 1_{d\mid\prod_{i=1}^t \psi_i(\vec{n})} = g(d)\operatorname{vol}(\Omega) + O(d^{O(1)}N^{m-1})$$

and hence axiom (5.2) is obeyed with $X := \operatorname{vol}(\Omega)$ and some $r_d = O(D^{O(1)}N^{m-1})$. Applying Lemma 5.1, we conclude that

$$\begin{aligned}\sum_{\vec{n}\in\Omega\cap\mathbb{Z}^m} \prod_{i=1}^t \Lambda_{\text{Cramér},z}(\psi_i(\vec{n})) &= \left(\prod_{p<z} \frac{p}{p-1}\right)^t \operatorname{vol}(\Omega) \prod_{p<z}(1-g(p)) \\ &\quad \times \left(1 + O\left(e^{-s}\exp(O(\log^{3/5} N))\right)\right) \\ &\quad + O\left(\left(\prod_{p<z} \frac{p}{p-1}\right)^t D^{O(1)}N^{m-1}\right)\end{aligned}$$

with $s = \frac{\log D}{\log Q} \gg \log^{4/5} N$. We can then simplify the right-hand side using (5.3) and Mertens' theorem to

$$\begin{aligned}\sum_{\vec{n}\in\Omega\cap\mathbb{Z}^m} \prod_{i=1}^t \Lambda_{\text{Cramér},z}(\psi_i(\vec{n})) &= \operatorname{vol}(\Omega)\Big(\prod_{p<z} \beta_p\Big)\left(1 + O\left(\exp(-c\log^{4/5} N)\right)\right) \\ &\quad + O(N^{m-1/2})\end{aligned}$$

(say) for some constant $c > 0$ depending on t, m. From (5.1), (5.3) and Mertens' theorem we have the crude bound

$$\prod_{p<z} \beta_p \ll \exp(O(\log^{3/5} N))$$

and the claim follows. ∎

As a first application of this estimate, we have good estimates (basically of logarithmic type) for the Cramér model in the Gowers norm.

Corollary 5.3 (Gowers uniformity of the Cramér model on arithmetic progressions). *Let $2 \leq w \leq z \leq Q$ be such that $w \leq \log^{1/100} N$. Set $W := P(w)$. Then for any $1 \leq b \leq W$ coprime to W, one has*

$$\left\| \frac{\phi(W)}{W} \Lambda_{\text{Cramér},z}(W \cdot + b) - 1 \right\|_{U^k[\frac{N-b}{W}]} \ll w^{-c}.$$

Proof. Write $N' := \frac{N-b}{W}$. We can rewrite the desired estimate (after adjusting c appropriately) as

$$\sum_{(n,\vec{h}) \in \Omega \cap \mathbb{Z}^{k+1}} \prod_{\omega \in \{0,1\}^k} \left(\frac{\phi(W)}{W} \Lambda_{\text{Cramér},z}(W(n + \omega \cdot \vec{h}) + b) - 1 \right) \ll (N')^{k+1} w^{-c},$$

where Ω is the convex body of tuples $(n, \vec{h}) \in \mathbb{R}^{k+1}$ such that

$$0 < n + \omega \cdot \vec{h} \leq N'$$

for all $\omega \in \{0,1\}^k$. By inclusion-exclusion, it suffices to establish the bounds

$$\sum_{(n,\vec{h}) \in \Omega \cap \mathbb{Z}^{k+1}} \prod_{\omega \in S} \frac{\phi(W)}{W} \Lambda_{\text{Cramér},z}(W(n + \omega \cdot \vec{h}) + b) = \text{vol}(\Omega) + O((N')^{k+1} w^{-c})$$

for all subsets $S \subset \{0,1\}^k$. Applying Proposition 5.2 (and Mertens' theorem), the left-hand side is equal to

$$\left(\frac{\phi(W)}{W} \right)^{\#S} \text{vol}(\Omega) \prod_{p<z} \beta_p + O((N')^{k+1} w^{-c})$$

(in fact, there is plenty of room to spare in the error term), where

$$\beta_p := \mathbb{E}_{(n,\vec{h}) \in (\mathbb{Z}/p\mathbb{Z})^{k+1}} \prod_{\omega \in S} \frac{p}{p-1} 1_{W(n+\omega\cdot\vec{h})+b \neq 0}.$$

If $p < w$, then W vanishes modulo p and b is coprime to p, and hence $\beta_p = (\frac{p}{p-1})^{\#S}$. Thus

$$\left(\frac{\phi(W)}{W} \right)^{\#S} \prod_{p<z} \beta_p = \prod_{w \leq p < z} \beta_p.$$

By the inclusion-exclusion argument used to establish (5.5) one has

$$\beta_p = \left(\frac{p}{p-1} \right)^{\#S} \left(1 - \frac{\#S}{p} + O\left(\frac{1}{p^2} \right) \right) = 1 + O\left(\frac{1}{p^2} \right)$$

for any $w \leq p < z$, hence

$$\prod_{w \leq p < z} \beta_p = 1 + O(w^{-1}).$$

Since $\mathrm{vol}(\Omega) \ll (N')^{k+1}$, the claim follows. ∎

Proof of Proposition 1.2. Combining Corollary 5.3 with Corollary 4.2, we see that

$$\| \Lambda_{\mathrm{Cramér},z} - \Lambda_{\mathrm{Cramér},w} \|_{U^k[N]} \ll w^{-c}$$

whenever $2 \leq w \leq z \leq \exp(\log^{1/10} N)$ are such that $w \leq \log^{1/100} N$. Proposition 1.2 now follows from the triangle inequality (1.2) (note the case $N = O(1)$ is trivial, so we may assume N is large enough that $\log^{1/100} N > 2$). ∎

Remark 5.4. With more effort it may be possible to delete the $\log^{-c} N$ term in (1.3), but we will not need to do so here as there are several other error terms in our analysis that are of the same order of magnitude as $\log^{-c} N$, or worse.

5.2. *Controlling the Siegel correction*

Now suppose that there is a Q-Siegel zero β, with associated quadratic character χ_{Siegel} and conductor q_{Siegel}. In this subsection we combine the previous sieve-theoretic estimates with Weil sum estimates to obtain good control on the Siegel models $\Lambda_{\mathrm{Siegel}}, \mu_{\mathrm{Siegel}}$.

We begin with some basic estimates on the Q-Siegel zero β and the Q-Siegel conductor q_{Siegel}. As χ_{Siegel} is a primitive real character, q_{Siegel} is must be either square-free or four times a square-free number or eight times a square-free number. From the construction one has the upper bound

$$q_{\mathrm{Siegel}} \leq Q = \exp(\log^{1/10} N).$$

From [8, Chapter 14, (12)] one has the estimate

$$1 - \beta \gg q_{\mathrm{Siegel}}^{-1/2} \log^{-2} q_{\mathrm{Siegel}},$$

which when combined with the upper bound $1 - \beta \ll \frac{1}{\log Q} \ll \log^{-1/10} N$ gives the lower bound

$$q_{\mathrm{Siegel}} \gg \frac{\log^{1/5} N}{(\log\log N)^2}. \tag{5.6}$$

One could improve this lower bound using Siegel's theorem to strongly logarithmic, but we will not do so here in order to keep the estimates effective. In particular, any bound of the shape $O(q_{\mathrm{Siegel}}^{-c})$ will lead to logarithmic decay.

From [35, Theorem 2.9] we observe the doubly logarithmic bound

$$\prod_{p | q_{\mathrm{Siegel}}} \left(1 - \frac{1}{p}\right)^{-1} = \frac{q_{\mathrm{Siegel}}}{\phi(q_{\mathrm{Siegel}})} \ll \log\log q_{\mathrm{Siegel}} \ll \log\log N. \tag{5.7}$$

Next, we show that the quantity α in (2.3) is bounded, which was the missing step needed to establish Lemma 2.4:

Lemma 5.5. *We have $\alpha \ll 1$. In particular, Lemma 2.4 holds.*

Proof. Consider the meromorphic function

$$F(s) = \frac{1}{L(s, \chi_{\text{Siegel}})} \prod_{p<Q} \left(1 - \frac{\chi_{\text{Siegel}}(p)}{p^s}\right)^{-1}.$$

This function has a simple pole at β with residue

$$\text{Res}(F, \beta) = \frac{1}{L'(\beta, \chi_{\text{Siegel}})} \prod_{p<Q} \left(1 - \frac{\chi_{\text{Siegel}}(p)}{p^\beta}\right)^{-1} = \alpha \prod_{p<Q} \left(1 - \frac{1}{p}\right)$$

and no other poles in the disk $\{s : |s - \beta| \leq \frac{2c_0}{\log Q}\}$ if c_0 is small enough, by [35, Theorem 11.3]. By Mertens' theorem, it thus suffices to establish the bound

$$\text{Res}(F, \beta) \ll \frac{1}{\log Q}.$$

By the residue theorem, it suffices to show that

$$F(s) \ll 1 \tag{5.8}$$

on the circle $|s - \beta| = \frac{2c_0}{\log Q}$. On the rightmost point $s_0 = \beta + \frac{2c_0}{\log Q} \geq 1 + \frac{c_0}{\log Q}$ of this circle, we can use the Euler product representation

$$F(s_0) = \prod_{p \geq Q} \left(1 - \frac{\chi_{\text{Siegel}}(p)}{p^{s_0}}\right)$$

followed by the triangle inequality to estimate

$$|F(s_0)| \leq \prod_{p \geq Q} \left(1 + \frac{1}{p^{1+\frac{c_0}{\log Q}}}\right) \ll 1 \tag{5.9}$$

thanks to Mertens' theorem. For more general points s on this circle, we deduce from [35, Theorem 11.4] that

$$\frac{L'}{L}(s, \chi_{\text{Siegel}}) \ll \log Q.$$

Since

$$\begin{aligned} \frac{F'}{F}(s) &= -\frac{L'}{L}(s, \chi_{\text{Siegel}}) - \sum_{p<Q} \frac{\chi_{\text{Siegel}}(p) \log p}{p^s - \chi_{\text{Siegel}}(p)} \\ &= -\frac{L'}{L}(s, \chi_{\text{Siegel}}) + O\left(\sum_{p<Q} \frac{\log p}{p^{1-\frac{3c_0}{\log Q}}}\right) \end{aligned}$$

(noting that $\mathrm{Re} s \geq 1 - \frac{3c_0}{\log Q}$), we conclude from Mertens' theorem that

$$\frac{F'}{F}(s) \ll \log Q$$

on the entire circle; integrating this and using (5.9), we obtain (5.8) as required. ∎

From [35, Theorem 11.4] we have

$$\frac{L'}{L}(s, \chi_{\mathrm{Siegel}}) = \frac{1}{s-\beta} + O(\log q_{\mathrm{Siegel}})$$

and

$$L(s, \chi_{\mathrm{Siegel}}) \gg |s-\beta|$$

for $s \neq \beta$ sufficiently close to β; multiplying the two estimates and taking limits as $s \to \beta$, we also obtain the bound

$$\frac{1}{L'(\beta, \chi_{\mathrm{Siegel}})} \ll 1. \tag{5.10}$$

We can view χ_{Siegel} as a function on $\mathbb{Z}/q_{\mathrm{Siegel}}\mathbb{Z}$. Crucially, it exhibits some cancellation in the Gowers norms (of polynomial type in q_{Siegel}, and hence of logarithmic type in N):

Lemma 5.6 (Gowers norm cancellation). *For any $\varepsilon > 0$, we have*

$$\|\chi_{\mathrm{Siegel}}\|_{U^k(\mathbb{Z}/q_{\mathrm{Siegel}}\mathbb{Z})} \ll_\varepsilon q_{\mathrm{Siegel}}^{-1/2^{k+1}+\varepsilon}.$$

Proof. By the Chinese remainder theorem, we can express $\mathbb{Z}/q_{\mathrm{Siegel}}\mathbb{Z}$ as the product of prime cyclic groups $\mathbb{Z}/p\mathbb{Z}$ of odd order, as well as $\mathbb{Z}/2^j\mathbb{Z}$ for some $0 \leq j \leq 3$. The quadratic character χ_{Siegel} can then be expressed as the tensor product of quadratic characters on these groups. Using (4.3) and the divisor bound, it thus suffices to show that

$$\|\chi\|_{U^k(\mathbb{Z}/p\mathbb{Z})} \ll p^{-1/2^{k+1}}$$

for all odd primes p, with χ the quadratic character on $\mathbb{Z}/p\mathbb{Z}$. By Definition 1.1, this is equivalent to

$$\mathbb{E}_{(n,\vec{h})\in(\mathbb{Z}/p\mathbb{Z})^{k+1}} \prod_{\omega\in\{0,1\}^k} \chi(n+\omega\cdot\vec{h}) \ll p^{-1/2}.$$

The contribution of any given tuple $\vec{h} \in (\mathbb{Z}/p\mathbb{Z})^k$ to the left-hand side is trivially bounded by $O(p^{-k})$. When the dot products $\omega\cdot\vec{h}$ are all distinct, the Weil bounds (see e.g. [29, Corollary 11.24]) give instead the bound $O(p^{-k-1/2})$. Since there are p^k tuples h and collisions between the $\omega\cdot\vec{h}$ only occur for $O(p^{k-1})$ of these tuples, the claim follows. ∎

We can now use this cancellation to prove Theorem 2.5.

Proof of Theorem 2.5. We may assume N is sufficiently large depending on k, and allow all implied constants to depend on k. Obviously we may assume that a Q-Siegel zero exists, as the claim is trivial otherwise.

We first establish (2.9). It suffices to show the polynomial (in q_{Siegel}) bound

$$\|\Lambda_{\text{Cramér},Q}\chi_{\text{Siegel}}(\cdot)^{\beta-1}\|_{U^k[N]} \ll q_{\text{Siegel}}^{-c}, \tag{5.11}$$

where $(\cdot)^{\beta-1}$ denotes the function $n \mapsto n^{\beta-1}$. By the fundamental theorem of calculus, we have

$$1_{[N]}(t)t^{\beta-1} = \int_1^N 1_{[M]}(t)(1-\beta)M^{\beta-2}\,dM + N^{\beta-1}1_{[N]}(t) \tag{5.12}$$

and

$$1 = \int_1^N (1-\beta)M^{\beta-2}\,dM + N^{\beta-1}. \tag{5.13}$$

Substituting (5.12) and (5.13) on the left and right-hand sides of (5.11), respectively, and applying Minkowski's integral inequality to the Banach space norm $\|\cdot\|_{U^k[N]}$, it suffices to show that[6]

$$\|\Lambda_{\text{Cramér},Q}\chi_{\text{Siegel}}1_{[M]}\|_{U^k[N]} \ll q_{\text{Siegel}}^{-c}$$

uniformly for all $1 \le M \le N$. By Definition 1.1, we can rewrite this estimate as

$$\sum_{(n,\vec{h})\in\Omega\cap\mathbb{Z}^{k+1}} \prod_{\omega\in\{0,1\}^k} \Lambda_{\text{Cramér},Q}\chi_{\text{Siegel}}(n+\omega\cdot\vec{h}) \ll N^{k+1}q_{\text{Siegel}}^{-c} \tag{5.14}$$

for some $c > 0$ and all $1 \le M \le N$, where $\Omega = \Omega_M$ is the convex body

$$\Omega := \big\{(x,\vec{y}) \in \mathbb{R}^{k+1} : 0 < x + \omega\cdot\vec{y} \le M \text{ for all } \omega \in \{0,1\}^k\big\}.$$

Splitting $n, h_1, \dots, h_k$ into cosets of q_{Siegel}, we can write the left-hand side of (5.14) as

$$\sum_{(a,\vec{b})\in[q_{\text{Siegel}}]^{k+1}} \prod_{\omega\in\{0,1\}^k} \chi_{\text{Siegel}}(a+\omega\cdot\vec{b})G(a,\vec{b}), \tag{5.15}$$

where

$$G(a,\vec{b}) := \sum_{(n,\vec{h})\in\frac{1}{q_{\text{Siegel}}}(\Omega-(a,\vec{b}))\cap\mathbb{Z}^{k+1}} \prod_{\omega\in\{0,1\}^k} \Lambda_{\text{Cramér},Q}(q_{\text{Siegel}}n + q_{\text{Siegel}}\omega\cdot\vec{h} + a + \omega\cdot\vec{b}).$$

Applying Proposition 5.2 (with N replaced by N/q_{Siegel}), we can estimate

$$G(a,\vec{b}) = q_{\text{Siegel}}^{-k-1}\,\text{vol}(\Omega)\prod_{p<z}\beta_p + O\big((N/q_{\text{Siegel}})^{k+1}\exp(-c\log^{4/5}N)\big),$$

where

$$\beta_p := \mathbb{E}_{(n,\vec{h})\in(\mathbb{Z}/p\mathbb{Z})^{k+1}} \prod_{\omega\in\{0,1\}^k} \frac{p}{p-1}1_{q_{\text{Siegel}}n+q_{\text{Siegel}}\omega\cdot\vec{h}+a+\omega\cdot\vec{b}\neq 0}.$$

[6] Alternatively, instead of applying Minkowski's integral inequality one could open the definition of the $U^k[N]$ norm, exchange the order of integration and averaging, and apply the Gowers–Cauchy–Schwarz inequality.

Because of the χ_{Siegel} factor in (5.15), we can restrict attention to the case where $a+\omega\cdot\vec{b}$ is coprime to q_{Siegel}. This implies that $\beta_p = (\frac{p}{p-1})^{2^k}$ when $p \mid q_{\mathrm{Siegel}}$. When $p \nmid q_{\mathrm{Siegel}}$, we can dilate $n, \vec{h}$ by $1/q_{\mathrm{Siegel}}$ (performing the division over the field $\mathbb{Z}/p\mathbb{Z}$) and then shift both variables to simplify

$$\beta_p = \mathbb{E}_{(n,\vec{h})\in(\mathbb{Z}/p\mathbb{Z})^{k+1}} \prod_{\omega\in\{0,1\}^k} \frac{p}{p-1} 1_{n+\omega\cdot\vec{h}\neq 0}.$$

In particular, the β_p are not dependent on $a, \vec{b}, q_{\mathrm{Siegel}}$. Summing in $a, \vec{b}$, we can thus write the left-hand side of (5.14) as

$$\mathrm{vol}(\Omega) \prod_{p<z} \beta_p \mathbb{E}_{(a,\vec{b})\in[q_{\mathrm{Siegel}}]^{k+1}} \prod_{\omega\in\{0,1\}^k} \chi_{\mathrm{Siegel}}(a+\omega\cdot\vec{b}) + O(N^{k+1}\exp(-c\log^{4/5} N)).$$

The error term is certainly negligible. From Lemma 5.6 we have

$$\mathbb{E}_{(a,\vec{b})\in[q_{\mathrm{Siegel}}]^{k+1}} \prod_{\omega\in\{0,1\}^k} \chi_{\mathrm{Siegel}}(a+\omega\cdot\vec{b}) \ll q_{\mathrm{Siegel}}^{-1/4}$$

(say), and we can of course bound $\mathrm{vol}(\Omega) \ll N^{k+1}$. Finally, direct calculation shows that $\beta_p = 1 + O(1/p^2)$ when $p \nmid q_{\mathrm{Siegel}}$, thus

$$\prod_{p<z} \beta_p \ll \prod_{p\mid q_{\mathrm{Siegel}}} \left(\frac{p}{p-1}\right)^{2^k} \ll \left(\frac{q_{\mathrm{Siegel}}}{\phi(q_{\mathrm{Siegel}})}\right)^{2^k} \ll (\log\log q_{\mathrm{Siegel}})^{2^k}$$

thanks to (5.7). Putting these estimates together, we obtain the claim (2.9).

Now we establish (2.8), which is a similar calculation but a little more involved because of the μ_{local} factor. By Lemma 4.3 and (5.6) it suffices to show that

$$\sum_{d\mid q_{\mathrm{Siegel}}} \frac{1}{d} \|\mu_{\mathrm{Siegel}}(d\cdot) 1_{(\cdot, q_{\mathrm{Siegel}}/d)=1}\|_{U^k[N/d]} \ll q_{\mathrm{Siegel}}^{-c}$$

for some $c>0$ depending on k. From (5.7) we have

$$\sum_{d\mid q_{\mathrm{Siegel}}} \frac{1}{d} \ll \log\log q_{\mathrm{Siegel}},$$

so it suffices to show that

$$\|\mu_{\mathrm{Siegel}}(d\cdot) 1_{(\cdot, q_{\mathrm{Siegel}}/d)=1}\|_{U^k[N/d]} \ll q_{\mathrm{Siegel}}^{-c}$$

for each $d \mid q_{\mathrm{Siegel}}$.

Fix d. We rewrite this estimate as

$$\|\mu_{\mathrm{Siegel}}(d\cdot) 1_{(\cdot, q_{\mathrm{Siegel}}/d)=1} 1_{[N/d]}\|_{\tilde{U}^k(\mathbb{Z})}^{2^k} \ll q_{\mathrm{Siegel}}^{-c} \|1_{[N/d]}\|_{\tilde{U}^k(\mathbb{Z})}^{2^k}. \tag{5.16}$$

Using Definition 2.1, we can write

$$\mu_{\mathrm{Siegel}}(dn)1_{(\cdot,q_{\mathrm{Siegel}}/d)=1}(n)1_{[N/d]}(n)$$
$$= \alpha \sum_{d'\in\mathcal{D}} \mu(d)\mu(d')1_{d'|n}(n/d')^{\beta-1}\chi_{\mathrm{Siegel}}(n/d')1_{(n/d',P(Q))=1}1_{[N/d]}(n) \tag{5.17}$$

where $\mathcal{D}$ is the set of all $d' \mid P(Q)$ with $(d', q_{\mathrm{Siegel}}) = 1$. By Lemma 5.5, it thus suffices to show that

$$\Big\| \sum_{d'\in\mathcal{D}} \mu(d')1_{d'|\cdot}(\cdot/d')^{\beta-1}\chi_{\mathrm{Siegel}}(\cdot/d')1_{(\cdot/d',P(Q))=1}1_{[N/d]}\Big\|^{2^k}_{\tilde{U}^k(\mathbb{Z})}$$
$$\ll q_{\mathrm{Siegel}}^{-c}\|1_{[N/d]}\|^{2^k}_{\tilde{U}^k(\mathbb{Z})}.$$

Using (5.12), (5.13) and Minkowski's integral inequality, it suffices to show

$$\Big\| \sum_{d'\in\mathcal{D}} \mu(d')1_{d'|\cdot}\chi_{\mathrm{Siegel},M}(\cdot/d')1_{[N/d]}\Big\|^{2^k}_{\tilde{U}^k(\mathbb{Z})} \ll q_{\mathrm{Siegel}}^{-c}\|1_{[N/d]}\|^{2^k}_{\tilde{U}^k(\mathbb{Z})} \tag{5.18}$$

for any $M \geq 1$, where

$$\chi_{\mathrm{Siegel},M}(n) := \chi_{\mathrm{Siegel}}(n)1_{[M]}(n)1_{(n,P(Q))=1}.$$

We decompose $\mathcal{D} = \mathcal{D}_{\leq} \cup \mathcal{D}_{>}$, where $\mathcal{D}_{\leq}$ are those $d' \in \mathcal{D}$ with $d' \leq \exp(\log^{1/2} N)$ (say) and $\mathcal{D}_{>}$ are those $d' \in \mathcal{D}$ with $d' > \exp(\log^{1/2} N)$. We first dispose of the contribution of the large d', i.e. $d' \in \mathcal{D}_{>}$. Their contribution to the expression inside the norm on the left-hand side of (5.18) is supported on a set of numbers n of size

$$\sum_{d'\in\mathcal{D}_{>}} \frac{N}{dd'}.$$

From basic estimates on smooth numbers [28, Theorem 1.1], the number of elements of $\mathcal{D}_{>}$ in any dyadic range $[M, 2M]$ with $M \in [Q, N]$ is $O(Mu^{-u/2})$ (say) where $u := \frac{\log M}{\log Q}$. From this and a routine dyadic decomposition we see that

$$\sum_{d'\in\mathcal{D}_{>}} \frac{N}{dd'} \ll N\exp(-\log^{-1/10} N)$$

(say). We thus see that the contribution to the left-hand side of (5.18) can be bounded by $O(N^{k+1}\exp(-\log^{-1/10} N))$, which is acceptable. Thus, by the triangle inequality (1.2), it suffices to control the contribution of $\mathcal{D}_{\leq}$, i.e. to show that

$$\Big\| \sum_{d'\in\mathcal{D}_{\leq}} \mu(d')1_{d'|\cdot}\chi_{\mathrm{Siegel},M}1_{[N/d]}\Big\|^{2^k}_{\tilde{U}^k(\mathbb{Z})} \ll q_{\mathrm{Siegel}}^{-c}\|1_{[N/d]}\|^{2^k}_{\tilde{U}^k(\mathbb{Z})}.$$

We can expand out the left-hand side as

$$\sum_{d'\in\mathcal{D}_{\leq}^{\{0,1\}^k}} A_{d'},$$

where for $d' = (d'_\omega)_{\omega\in\{0,1\}^k}$ we have

$$A_{d'} := \sum_{(n,\vec{h})\in\Omega} \prod_{\omega\in\{0,1\}^k} \mu(d'_\omega) 1_{d'_\omega | n+\omega\cdot\vec{h}} \chi_{\mathrm{Siegel},M}\left(\frac{n+\omega\cdot\vec{h}}{d'_\omega}\right),$$

where Ω is the set of all tuples $(n,\vec{h}) \in \mathbb{Z}^{k+1}$ such that $n+\omega\cdot\vec{h} \in [N/d]$ for all ω in $\{0,1\}^k$. Meanwhile, using the pointwise bound

$$0 \le \sum_{d'\in\mathcal{D}_\le} 1_{d'|.} 1_{[N/d]} 1_{(./d',P(Q))=1} \le 1_{[N/d]}$$

(reflecting the fact that every number n has a unique decomposition $n = d'(n/d')$ where $d' \mid P(Q)$ and $(n/d', P(Q)) = 1$) one has

$$\sum_{d'\in\mathcal{D}_\le^{\{0,1\}^k}} B_{d'} \le \|1_{[N/d]}\|^{2^k}_{\tilde{U}^k(\mathbb{Z})},$$

where

$$B_{d'} := \sum_{(n,\vec{h})\in\Omega} \prod_{\omega\in\{0,1\}^k} 1_{d'_\omega | n+\omega\cdot\vec{h}} 1_{(\frac{n+\omega\cdot\vec{h}}{d'_\omega},P(Q))=1}.$$

Hence it will suffice to show that

$$A_{d'} \ll q_{\mathrm{Siegel}}^{-c} B_{d'}$$

for all $d' \in \mathcal{D}_\le^{\{0,1\}^k}$.

The constraints $1_{d'_\omega | n+\omega\cdot\vec{h}}$ restrict $(n,\vec{h})$ to some finite union of cosets $(a,\vec{b}) + D\mathbb{Z}^{k+1}$ of $D\mathbb{Z}^{k+1}$ where $D := \prod_{\omega\in\{0,1\}^k} d'_\omega$, with the property that d'_ω divides $a+\omega\cdot\vec{b}$ for all $\omega \in \{0,1\}^k$. Note from the construction that D is coprime to q_{Siegel} and of size $O(\exp(O(\log^{1/2} N)))$. So, denoting for brevity $\Omega^{(a,\vec{b})} := \Omega \cap ((a,\vec{b}) + D\mathbb{Z}^{k+1})$, it will suffice to show that

$$\sum_{(n,\vec{h})\in\Omega^{(a,\vec{b})}} \prod_{\omega\in\{0,1\}^k} \chi_{\mathrm{Siegel},M}\left(\frac{n+\omega\cdot\vec{h}}{d'_\omega}\right) \ll q_{\mathrm{Siegel}}^{-c} \sum_{(n,\vec{h})\in\Omega^{(a,\vec{b})}} \prod_{\omega\in\{0,1\}^k} 1_{(\frac{n+\omega\cdot\vec{h}}{d'_\omega},P(Q))=1} \tag{5.19}$$

for all such cosets $(a,\vec{b}) + D\mathbb{Z}^{k+1}$. Using Proposition 5.2 and some elementary rescaling, we have

$$\sum_{(n,\vec{h})\in\Omega^{(a,\vec{b})}} \prod_{\omega\in\{0,1\}^k} 1_{(\frac{n+\omega\cdot\vec{h}}{d'_\omega},P(Q))=1} = D^{-k-1}\,\mathrm{vol}(\Omega) \prod_{p<Q} \tilde{\beta}_p + O\big((N/D)^{k+1}\exp(-c\log^{4/5} N)\big),$$

where

$$\tilde{\beta}_p := \mathbb{E}_{(n,\vec{h})\in(\mathbb{Z}/p\mathbb{Z})^{k+1}} \prod_{\omega\in\{0,1\}^k} 1_{\frac{a+\omega\cdot\vec{b}}{d'_\omega}+(n+\omega\cdot\vec{h})\frac{D}{d'_\omega}\neq 0}.$$

If any of the $\tilde{\beta}_p$ vanish then both sides of (5.19) vanish and we are done. For p not dividing D we have the crude bound

$$\tilde{\beta}_p = 1 - O(1/p), \tag{5.20}$$

and for all p we have the lower bound

$$\tilde{\beta}_p \geq \frac{1}{p^{k+1}} \tag{5.21}$$

since the $\tilde{\beta}_p$ are non-vanishing integer multiples of $1/p^{k+1}$. This gives the crude lower bound

$$\prod_{p<Q} \tilde{\beta}_p \gg D^{-O(1)} \log^{-O(1)} N \tag{5.22}$$

and hence the right-hand side of (5.19) is comparable to $q_{\text{Siegel}}^{-c}(N/D)^{k+1}\prod_{p<Q}\tilde{\beta}_p$. Next, we partition the left-hand side of (5.19) as

$$\sum_{(r,\vec{s})\in[q_{\text{Siegel}}]^{k+1}} \prod_{\omega\in\{0,1\}^k} \chi_{\text{Siegel}}(d'_\omega)\chi_{\text{Siegel}}(r+\omega\cdot\vec{s})F_{r,\vec{s}}, \tag{5.23}$$

where

$$F_{r,\vec{s}} := \sum_{(n,\vec{h})\in\Omega\cap((a,\vec{b})+D\mathbb{Z}^{k+1})\cap((r,\vec{s})+q_{\text{Siegel}}\mathbb{Z}^{k+1})} \prod_{\omega\in\{0,1\}^k} 1_{[M]}\Big(\frac{n+\omega\cdot\vec{h}}{d'_\omega}\Big) 1_{(n+\omega\cdot\vec{h},P(Q))=1}.$$

We can restrict attention to those $(r,\vec{s})$ for which $r+\omega\cdot\vec{s}$ is coprime to q_{Siegel} for all $\omega\in\{0,1\}^k$, since otherwise the product in (5.23) vanishes. Under this assumption, we can apply Proposition 5.2, the Chinese remainder theorem, and some further rescaling (using the fact that D, q_{Siegel} are coprime), to conclude that

$$F_{r,\vec{s}} = (Dq_{\text{Siegel}})^{-k-1}\operatorname{vol}(\Omega')\prod_{\substack{p<Q\\ p\nmid q_{\text{Siegel}}}}\tilde{\beta}_p + O\big((N/Dq_{\text{Siegel}})^{k+1}\exp(-c\log^{4/5}N)\big),$$

where

$$\Omega' := \Big\{(n,\vec{h})\in\Omega : \frac{n+\omega\cdot\vec{h}}{d'_\omega}\in[M]\ \forall\omega\in\{0,1\}^k\Big\}.$$

Note the main term here is independent of $r,\vec{s}$. In particular, we can rewrite (5.23) as

$$\Big(\prod_{\omega\in\{0,1\}^k}\chi_{\text{Siegel}}(d'_\omega)\Big)\|\chi_{\text{Siegel}}\|_{\tilde{U}^k(\mathbb{Z}/q_{\text{Siegel}})}^{2^k}(Dq_{\text{Siegel}})^{-k-1}\operatorname{vol}(\Omega')\prod_{\substack{p<Q\\ p\nmid q_{\text{Siegel}}}}\tilde{\beta}_p + O\big((N/D)^{k+1}\exp(-c\log^{4/5}N)\big).$$

By Lemma 5.6, this quantity is

$$\ll q_{\text{Siegel}}^{-c}(N/D)^{k+1} \prod_{\substack{p<Q \\ p\nmid q_{\text{Siegel}}}} \tilde{\beta}_p + (N/D)^{k+1}\exp(-c\log^{4/5} N). \tag{5.24}$$

The second term in (5.24) is acceptable thanks to (5.22). From (5.20), (5.21), (5.7) we have

$$\prod_{p\mid q_{\text{Siegel}}} \tilde{\beta}_p \gg (\log\log q_{\text{Siegel}})^{-O(1)}$$

and so the first term in (5.24) is also acceptable. ∎

6. The Manners inverse theorem

We are now ready to state a version of the inverse theorem of Manners [33], though formulated in a slightly different language (in particular, using the complexity notions from [24] rather than [33]).

Definition 6.1 (Nilmanifolds). Let $s \geq 1$ be an integer, and let $M > 0$. A (filtered) *nilmanifold* G/Γ of degree s and complexity at most M consists of the following data:

(i) a nilpotent connected and simply connected Lie group G of some dimension m, which can be identified with its Lie algebra $\log G$ via the exponential map $\exp\colon \log G \to G$ or its inverse $\log\colon G \to \log G$;

(ii) a filtration $G_\bullet = (G_i)_{i\geq 0}$ of closed connected subgroups G_i of G with

$$G = G_0 = G_1 \geq \cdots \geq G_s \geq G_{s+1} = \{\mathrm{id}_G\}$$

(and G_i trivial for all $i \geq s+1$) such that[7] $[G_i, G_j] \subset G_{i+j}$ for all $i, j \geq 0$ (or equivalently $[\log G_i, \log G_j] \subset \log G_{i+j}$ in the Lie algebra $\log G$);

(iii) a discrete cocompact subgroup Γ of G;

(iv) a linear basis $X_1, \ldots, X_{\dim(G)}$ of $\log G$, known as a *Mal'cev basis* (*of the second kind*).

We require this data to obey the following axioms:

(a) For $1 \leq i, j \leq \dim(G)$, one has

$$[X_i, X_j] = \sum_{i,j<k\leq \dim(G)} c_{ijk} X_k \tag{6.1}$$

for some rational numbers c_{ijk} with numerator and denominator bounded in magnitude by M.

[7] We use $[,]$ to denote both the commutator in the Lie group G and the Lie bracket in the Lie algebra $\log G$, with the two being related to each other by the Baker–Campbell–Hausdorff formula.

(b) For each $1 \leq i \leq s$, the Lie algebra $\log G_i$ is spanned by the X_j with $\dim(G) - \dim(G_i) < j \leq \dim(G)$.

(c) The subgroup Γ consists of all elements of the form $\exp(t_1 X_1) \cdots \exp(t_{\dim G} X_{\dim G})$ with $t_1, \ldots, t_{\dim G} \in \mathbb{Z}$.

This data defines a metric on G/Γ as described in [24, Definition 2.2], as well as the notion of a polynomial map $g: \mathbb{Z} \to G$, defined in [24, Definition 1.8].

A function $f : X \to \mathbb{C}$ is said to be 1-*bounded* if $|f(n)| \leq 1$ for all $n \in X$.

Theorem 6.2 (Manners inverse theorem). *Let $0 < \delta < 1$. Let $f : [N] \to \mathbb{C}$ be a 1-bounded function such that*

$$\|f\|_{U^k[N]} \geq \delta.$$

Then there exist a (filtered) nilmanifold G/Γ of degree $k-1$, dimension $O(\delta^{-O(1)})$, and complexity at most $\exp\exp(O(1/\delta^{O(1)}))$, a 1-bounded Lipschitz function $F: G/\Gamma \to \mathbb{C}$ of Lipschitz constant at most $\exp\exp(O(1/\delta^{O(1)}))$, and a polynomial map $g: \mathbb{Z} \to G$, such that

$$|\mathbb{E}_{n \in [N]} f(n) \overline{F}(g(n)\Gamma)| \gg \exp(-\exp(O(1/\delta^{O(1)}))).$$

Proof. By Bertrand's postulate we can find a prime N' such that $10N \leq N' \leq 20N'$. If we embed $[N]$ into the cyclic group $\mathbb{Z}/N'\mathbb{Z}$ and extend f by zero we may view f as a 1-bounded function on $\mathbb{Z}/N'\mathbb{Z}$, and a brief calculation reveals that

$$\|f\|_{U^k(\mathbb{Z}/N'\mathbb{Z})} \gg \delta.$$

We now apply [33, Theorem 1.1.2] with $s := k-1$ to produce the required data G/Γ, g, F, X_i, save for two differences. Firstly, the polynomial g is described as a map from $\mathbb{Z}/N'\mathbb{Z}$ to G/Γ rather than from $\mathbb{Z}$ to G, but one can lift the map from the former to the latter using [33, Proposition C.17]. Secondly, instead of axiom (a) of Definition 6.1, the basis elements X_i are instead required to obey a decomposition

$$[\exp(X_i), \exp(X_j)] = \prod_{i,j < l \leq \dim(G)} \exp(a_{ijl} X_l) \tag{6.2}$$

for some integers a_{ijl} bounded in magnitude by some bound $M_0 \ll \exp\exp(O(1/\delta^{O(1)}))$, where the product is taken from left to right. However, as briefly noted in [33, Section C.2], one can pass from this control (6.2) to the control (6.1) (with M a suitable polynomial of M_0), as follows. For any $1 \leq a \leq k-1$, we let $P(a)$ denote the claim that one has (6.1) with M of the form $\exp\exp(O(1/\delta^{O(1)}))$ whenever one of X_i, X_j lies in $\log G_a$. The claim $P(a)$ is certainly true for $a = k-1$ since $\log G_{k-1}$ is central, and we will be done if $P(1)$ is true, so it suffices by downward induction (with at most $k-2$ steps) to show that $P(a+1)$ implies $P(a)$ for any $1 \leq a \leq k-2$, where the implied constants in the $O_k()$ notation are allowed to vary with each step of the induction. Call a rational number *good* if its numerator and denominator are bounded in magnitude by $\exp\exp(O(1/\delta^{O(1)}))$. If one of X_i, X_j lies in $\log G_a$, then from (6.2), the induction

hypothesis, and the Baker–Campbell–Hausdorff formula we see that

$$\log[\exp(X_i), \exp(X_j)] = \sum_{l>i,j} c'_{ijl} X_l \tag{6.3}$$

for some good rationals c'_{ijl} (and furthermore one can restrict to those X_k lying in $\log G_{a+1}$). On the other hand, a further application of Baker–Campbell–Hausdorff reveals that $\log[\exp(X_i), \exp(X_j)]$ is equal to $[X_i, X_j]$ plus $O_k(1)$ additional terms, which consist of a good rational number times an iterated Lie bracket formed by starting with $[X_i, X_j]$ and taking the Lie bracket with either X_i or X_j one or more times (but no more than $O(1)$ times in all). Inverting this formula, we can then write $[X_i, X_j]$ as $\log[\exp(X_i), \exp(X_j)]$ plus $O(1)$ additional terms, which consist of a good rational number times an iterated Lie bracket formed by starting with $\log[\exp(X_i), \exp(X_j)]$ and taking the Lie bracket with either X_i or X_j one or more times (but no more than $O(1)$ times in all). Using (6.3) and the induction hypothesis $P(a+1)$ repeatedly, we conclude $P(a)$, thus closing the induction. ∎

Remark 6.3. As noted in [33], improved bounds are available for $k \leq 4$ [18, 26], but we will not be able to take advantage of these bounds due to inefficiencies elsewhere in the arguments (in particular, our nilsequence equidistribution theory involves exponents that are exponential in the dimension rather than polynomial).

From Lemma 2.4 we see that the function $\mu - \mu_{\text{Siegel}}$ can be made 1-bounded by multiplying by a small absolute constant. Applying Theorem 6.2 in the contrapositive (setting δ equal to a small power of $(\log\log N)^{-1}$, we conclude that the bound (2.10) is an immediate consequence of (2.12). The same argument does not work directly for $\Lambda - \Lambda_{\text{Siegel}}$ due to the additional factor of $\log N$ in the pointwise bounds; but we will be able to get around this in Section 8 by employing the densification technology of Conlon, Fox, and Zhao [5]. Assuming this for the moment, the only remaining step needed to establish Theorem 1.4 is to prove Theorem 2.7, to which we now turn.

Remark 6.4. When $k = 3$, one can appeal instead of Theorem 6.2 to the quantitative inverse theorem in [18], and when $k = 4$ one can use the fact that Manners proved in [33] a stronger form of Theorem 6.2 for $k = 4$ than for $k \geq 5$. If one does so, one eventually finds that one would be able to improve the doubly logarithmic bounds in Theorem 1.4 for $k \leq 4$ to singly logarithmic, provided that one could increase the bound on the dimension of G/Γ in Theorem 2.7 from $(\log\log N)^{c_1}$ to $\log^{c_1} N$. Unfortunately, our equidistribution theory on nilmanifolds is currently not satisfactory at this high a dimension, although in principle it is conceivable that some variant of the methods of Schmidt [40] could resolve this issue. We will not pursue this question further here.

7. Orthogonality to nilsequences

In this section we prove Theorem 2.7. We begin by establishing Proposition 2.2, which will be used to establish the "major arc" case of Theorem 2.7.

Proof of Proposition 2.2. We adopt the convention that any factor involving the Q-Siegel character χ_{Siegel} is deleted if no such character exists. Any arithmetic progression $P \subset [N]$ can be expressed in the form $\{N'' < n \leq N' : n \equiv a\ (q)\}$ for some $1 \leq a \leq q$ and $0 < N'' \leq N' \leq N$. By the triangle inequality, it thus suffices to establish the bounds

$$\sum_{\substack{n \leq N' \\ n \equiv a\ (q)}} \Lambda(n) = \sum_{\substack{n \leq N' \\ n \equiv a\ (q)}} \Lambda_{\text{Siegel}}(n) + O(N \exp(-c \log^{1/10} N)) \tag{7.1}$$

and

$$\sum_{\substack{n \leq N' \\ n \equiv a\ (q)}} \mu(n) = \sum_{\substack{n \leq N' \\ n \equiv a\ (q)}} \mu_{\text{Siegel}}(n) + O(N \exp(-c \log^{1/10} N)) \tag{7.2}$$

for any $1 \leq a \leq q$ and $0 < N' \leq N$.

If $q > \exp(c_2 \log^{1/10} N)$ for any constant $c_2 > 0$ then the triangle inequality (and Lemma 2.4) give the desired bounds after adjusting the value of c, so we may assume that $q \leq \exp(c_2 \log^{1/10} N)$ for some small absolute constant c_2. In particular, $q \leq Q$. Similarly we may assume $N' \geq N \exp(-c_2 \log^{1/10} N)$.

We begin with (7.1). From [29, Theorem 5.27] one has

$$\sum_{\substack{n \leq N' \\ n \equiv a\ (q)}} \Lambda(n) = \frac{N'}{\phi(q)} \left(1 - \chi_{\text{Siegel}}(a) 1_{q_{\text{Siegel}}|q} \frac{(N')^{\beta-1}}{\beta}\right) 1_{(a,q)=1} + O(N \exp(-c \log^{1/10} N)).$$

Therefore, it will certainly suffice from the triangle inequality to show for $1 \leq a \leq q \leq \exp(\log^{3/5} N)$ that[8]

$$\sum_{\substack{n \leq N' \\ n \equiv a\ (q)}} \Lambda_{\text{Cramér},Q}(n) = \frac{N'}{\phi(q)} 1_{(a,q)=1} + O(N \exp(-c \log^{4/5} N)) \tag{7.3}$$

and

$$\sum_{\substack{n \leq N' \\ n \equiv a\ (q)}} (\Lambda_{\text{Cramér},Q}(n) - \Lambda_{\text{Siegel}}(n)) = \frac{(N')^{\beta}}{\beta \phi(q)} \chi_{\text{Siegel}}(a) 1_{q_{\text{Siegel}}|q} 1_{(a,q)=1} + O(N \exp(-c \log^{4/5} N)). \tag{7.4}$$

We first show (7.3). By a change of variables we have

$$\sum_{\substack{n \leq N' \\ n \equiv a\ (q)}} \Lambda_{\text{Cramér},Q}(n) = \sum_{\frac{-a}{q} \leq n \leq \frac{N'-a}{q}} \Lambda_{\text{Cramér},Q}(qn + a)$$

[8] It would of course suffice to show this for $q \leq \exp(\log^{1/10} N)$ and with savings $\exp(-c \log^{1/10} N)$, but the larger powers of $\log N$ will be useful later on.

and then on applying Proposition 5.2 we have

$$\sum_{\substack{n\le N'\\ n\equiv a\ (q)}} \Lambda_{\mathrm{Cramér},Q}(n) = \frac{N'}{q}\prod_{p<Q}\beta_p + O(N\exp(-c\log^{4/5}N)),$$

where

$$\beta_p := \mathbb{E}_{n\in\mathbb{Z}/p\mathbb{Z}}\frac{p}{p-1}1_{qn+a\neq 0}.$$

If $(a,q)>1$ then (a,q) will be divisible by some prime $p\le q<Q$, in which case $\beta_p=0$ and the claim follows. If instead $(a,q)=1$, then $\beta_p=1$ for all $p<Q$ not dividing q, and $\beta_p=\frac{p}{p-1}$ for all $p<Q$ dividing q, and the claim (7.3) follows.

Now we show (7.4). We may of course assume there is a Q-Siegel zero, in which case (by Definition 2.1 (ii)) our task is to show that

$$\sum_{\substack{n\le N'\\ n\equiv a\ (q)}} \Lambda_{\mathrm{Cramér},Q}(n)n^{\beta-1}\chi_{\mathrm{Siegel}}(n) = \frac{(N')^\beta}{\beta\phi(q)}\chi_{\mathrm{Siegel}}(a)1_{q_{\mathrm{Siegel}}|q}1_{(a,q)=1} + O(N\exp(-c\log^{4/5}N)).$$

From the fundamental theorem of calculus we have

$$n^{\beta-1}1_{[N']}(n) = \int_1^{N'}(1-\beta)M^{\beta-2}1_{[M]}(n)\,dM + (N')^{\beta-1}1_{[N']}(n)$$

and

$$\frac{(N')^\beta}{\beta} - \frac{1}{\beta} + 1 = \int_1^{N'}(1-\beta)M^{\beta-2}M\,dM + (N')^{\beta-1}N',$$

so from the triangle inequality it suffices to show that

$$\sum_{\substack{n\le M\\ n\equiv a\ (q)}} \Lambda_{\mathrm{Cramér},Q}(n)\chi_{\mathrm{Siegel}}(n) = \frac{M}{\phi(q)}\chi_{\mathrm{Siegel}}(a)1_{q_{\mathrm{Siegel}}|q}1_{(a,q)=1} + O(N\exp(-c\log^{4/5}N))$$

for all $1\le M\le N$. We split the left-hand side as

$$\sum_{\substack{1\le b\le q'\\ b\equiv a\ (q)}}\chi_{\mathrm{Siegel}}(b)\sum_{\substack{n\le M\\ n\equiv b\ (q')}}\Lambda_{\mathrm{Cramér},Q}(n),$$

where $q':=[q,q_{\mathrm{Siegel}}]$ is the least common multiple of q and q_{Siegel}. By (7.3) we have

$$\sum_{\substack{n\le M\\ n\equiv b\ (q')}} \Lambda_{\mathrm{Cramér},Q}(n) = \frac{M}{\phi(q')}1_{(b,q')=1} + O(N\exp(-c\log^{4/5}N))$$

and thus

$$\sum_{\substack{n\le M\\ n\equiv a\ (q)}} \Lambda_{\mathrm{Cramér}}(n)\chi_{\mathrm{Siegel}}(n) = \frac{M}{\phi(q')}\sum_{\substack{1\le b\le q'\\ b\equiv a\ (q)}}\chi_{\mathrm{Siegel}}(b)1_{(b,q')=1} + O(N\exp(-c\log^{4/5}N)).$$

The right-hand side vanishes if $(a,q) > 1$, and also vanishes if $q' > q$ due to the orthogonality properties of Dirichlet characters. If instead $(a,q) = 1$ and $q' = q$ then the right-hand side is equal to $\frac{M}{\phi(q)}\chi_{\mathrm{Siegel}}(a)$, and the claim (7.4) follows.

Now we turn to (7.2). We first do an easy reduction to the case of primitive residue classes. Let $d := (a,q)$. Observe that for any natural number n one has

$$\mu(dn) = \mu(d)\mu(n)1_{(n,d)=1}$$

and also from Definition 2.1 (ii) we similarly have

$$\mu_{\mathrm{Siegel}}(dn) = \mu(d)\mu_{\mathrm{Siegel}}(n)1_{(n,d)=1}$$

and thus

$$\begin{aligned}\sum_{\substack{n\le N'\\ n\equiv a\ (q)}} (\mu(n)-\mu_{\mathrm{Siegel}(n)}) &= \mu(d) \sum_{\substack{n\le N'/d\\ n\equiv a/d\ (q/d)\\ (n,d)=1}} (\mu(n)-\mu_{\mathrm{Siegel}}(n))\\ &= \mu(d)\sum_{\substack{1\le b\le d\\ (b,d)=1}} \sum_{\substack{n\le N'/d\\ n\equiv a/d\ (q/d)\\ n\equiv b\ (d)}} (\mu(n)-\mu_{\mathrm{Siegel}}(n)). \qquad (7.5)\end{aligned}$$

Since $d \le q \le \exp(c_2 \log^{1/10} N)$, it thus suffices to establish the pseudopolynomial decay estimate

$$\sum_{\substack{n\le N'/d\\ n\equiv a/d\ (q/d)\\ n\equiv b\ (d)}} (\mu(n)-\mu_{\mathrm{Siegel}}(n)) \ll N\exp(-c\log^{1/10} N)$$

for all $1 \le b \le d$ coprime to d (where the constant c here is uniform in c_2). Writing $q' := [q/d, d]$, we see from the Chinese remainder theorem that the constraints $n \equiv a/d\ (q/d)$ and $n \equiv b\ (d)$ are either inconsistent, or constrain n to precisely one primitive residue class $a'\ (q')$ with $(a',q') = 1$. Thus it suffices to show the pseudopolynomial decay bound

$$\sum_{\substack{n\le N'\\ n\equiv a'\ (q')}} (\mu(n)-\mu_{\mathrm{Siegel}(n)}) \ll N\exp(-c\log^{1/10} N)$$

whenever $1 \le N' \le N$ and $1 \le a' \le q' \le \exp(2c_2 \log^{1/10} N)$ with $(a',q') = 1$.

When there is no Q-Siegel zero the claim is immediate from [35, Exercise 11.3.12] (modified slightly due to our slightly different definition of a Siegel zero). Now suppose that there is a Q-Siegel zero. The result previously cited in [35, Exercise 11.3.12] (again modified slightly to account for our slightly different notion of Siegel zero) then gives the pseudopolynomially accurate asymptotic

$$\sum_{\substack{n\le N'\\ n\equiv a'\ (q')}} \mu(n) = 1_{q_{\mathrm{Siegel}}|q'}\frac{\chi_{q'}(a')(N')^\beta}{\phi(q')L'(\beta,\chi_{q'})\beta} + O(N\exp(-c\log^{1/10} N))$$

where $\chi_{q'}(n) := \chi_{\text{Siegel}}(n)1_{(n,q')=1}$ is the character of modulus q' induced from χ_{Siegel} when q' is a multiple of q_{Siegel}. Note that

$$L(s,\chi_{q'}) = L(s,\chi_{\text{Siegel}}) \prod_{\substack{p|q' \\ p\nmid q_{\text{Siegel}}}} \left(1-\frac{\chi_{\text{Siegel}}(p)}{p^s}\right),$$

and thus by the product rule (and the fact that $L(\beta,\chi_{\text{Siegel}}) = 0$),

$$L'(\beta,\chi_q) = L'(\beta,\chi_{\text{Siegel}}) \prod_{\substack{p|q' \\ p\nmid q_{\text{Siegel}}}} \left(1-\frac{\chi_{\text{Siegel}}(p)}{p^\beta}\right).$$

We conclude that

$$\sum_{\substack{n\le N' \\ n\equiv a'\ (q')}} \mu(n) = 1_{q_{\text{Siegel}}|q'} \frac{(N')^\beta \chi_{\text{Siegel}}(a')}{\beta\phi(q)L'(\beta,\chi_{\text{Siegel}})} \prod_{\substack{p|q' \\ p\nmid q_{\text{Siegel}}}} \left(1-\frac{\chi_{\text{Siegel}}(p)}{p^\beta}\right)^{-1} + O(N\exp(-c\log^{1/10} N)).$$

It will thus suffice to establish the corresponding pseodupolynomially accurate asymptotic

$$\sum_{\substack{n\le N' \\ n\equiv a'\ (q')}} \mu_{\text{Siegel}}(n) = 1_{q_{\text{Siegel}}|q'} \frac{(N')^\beta \chi_{\text{Siegel}}(a')}{\beta\phi(q')L'(\beta,\chi_{\text{Siegel}})} \prod_{\substack{p|q' \\ p\nmid q_{\text{Siegel}}}} \left(1-\frac{\chi_{\text{Siegel}}(p)}{p^\beta}\right)^{-1} + O(N\exp(-c\log^{1/10} N)) \tag{7.6}$$

for μ_{Siegel}. It suffices to establish the variant estimate

$$\sum_{\substack{n\le N' \\ n\equiv a'\ (q')}} \mu_{\text{Siegel}}(n) = \frac{(N')^\beta \chi_{\text{Siegel}}(a')}{\beta\phi(q')L'(\beta,\chi_{\text{Siegel}})} \prod_{\substack{p|q' \\ p\nmid q_{\text{Siegel}}}} \left(1-\frac{\chi_{\text{Siegel}}(p)}{p^\beta}\right)^{-1} + O(N\exp(-c\log^{1/10} N)) \tag{7.7}$$

(say) whenever $1 \le a' \le q' \le \exp(O(\log^{1/10} N))$ with $(a',q') = 1$ and $q_{\text{Siegel}} \mid q'$. Indeed, this estimate immediately implies (7.6) when q_{Siegel} divides q', and when q_{Siegel} does not divide q', one splits up the primitive residue class $a'\ (q')$ into primitive residue classes modulo $[q', q_{\text{Siegel}}]$ on the support of μ_{Siegel}, applies (7.7) to each such class, and sums, using the orthogonality of Dirichlet characters to cancel out the main term.

We use Definition 2.1 to expand the left-hand of (7.7) as

$$\sum_{d\in\mathcal{D}} \mu(d) \sum_{\substack{n\le N'/d \\ dn\equiv a'\ (q')}} \mu'(n),$$

where $\mathcal{D}$ consists of all the factors d of $P(Q)$ with $(d,q') = 1$. As in the proof of (5.18), we can decompose $\mathcal{D}_\le \cup \mathcal{D}_>$, where $\mathcal{D}_\le$ are those $d' \in \mathcal{D}$ with $d' \le \exp(\log^{1/2} N)$ (say) and $\mathcal{D}_>$ are those $d' \in \mathcal{D}$ with $d' > \exp(\log^{1/2} N)$. The contribution of $\mathcal{D}_>$ can be

disposed of by the same argument used to prove (5.18), so it remains to show that

$$\sum_{d\in\mathcal{D}_\leq}\mu(d)\sum_{\substack{n\leq N'/d\\ dn\equiv a'\ (q')}}\mu'(n)=\frac{(N')^\beta\chi_{\mathrm{Siegel}}(a')}{\beta\phi(q')L'(\beta,\chi_{\mathrm{Siegel}})}\prod_{\substack{p\mid q'\\ p\nmid q_{\mathrm{Siegel}}}}\left(1-\frac{\chi_{\mathrm{Siegel}}(p)}{p^\beta}\right)^{-1}$$
$$+O(N\exp(-c\log^{1/10}N)).$$

By Definition 2.1, we have

$$\sum_{\substack{n\leq N'/d\\ dn\equiv a'\ (q')}}\mu'(n)=\alpha\frac{\phi(P(Q))}{P(Q)}\sum_{\substack{n\leq N'/d\\ n\equiv a'/d\ (q')}}(\Lambda_{\mathrm{Cramér},Q}(n)-\Lambda_{\mathrm{Siegel}}(n)).$$

Applying (7.4), as well as Lemma 5.5, we can write this as

$$\alpha\frac{\phi(P(Q))}{P(Q)}\frac{(N'/d)^\beta}{\beta\phi(q')}\chi_{\mathrm{Siegel}}(a')\chi_{\mathrm{Siegel}}(d)$$

up to acceptable error terms. Canceling some terms, it thus suffices to show that

$$\alpha\frac{\phi(P(Q))}{P(Q)}\sum_{d\in\mathcal{D}_\leq}\frac{\mu(d)\chi_{\mathrm{Siegel}}(d)}{d^\beta}=\frac{1}{L'(\beta,\chi_{\mathrm{Siegel}})}\prod_{\substack{p\mid q'\\ p\nmid q_{\mathrm{Siegel}}}}\left(1-\frac{\chi_{\mathrm{Siegel}}(p)}{p^\beta}\right)^{-1}$$
$$+O(\exp(-c\log^{1/10}N)).$$

A standard Euler product calculation using (2.3) gives

$$\alpha\frac{\phi(P(Q))}{P(Q)}\sum_{d\in\mathcal{D}}\frac{\mu(d)\chi_{\mathrm{Siegel}}(d)}{d^\beta}=\frac{1}{L'(\beta,\chi_{\mathrm{Siegel}})}\prod_{\substack{p\mid q'\\ p\nmid q_{\mathrm{Siegel}}}}\left(1-\frac{\chi_{\mathrm{Siegel}}(p)}{p^\beta}\right)^{-1},$$

so it suffices to show that

$$\alpha\frac{\phi(P(Q))}{P(Q)}\sum_{d\in\mathcal{D}_>}\frac{\mu(d)\chi_{\mathrm{Siegel}}(d)}{d^\beta}\ll\exp(-c\log^{1/10}N)).$$

By Lemma 5.5 and the triangle inequality it suffices to show that

$$\sum_{d\in\mathcal{D}_>}\frac{1}{d^\beta}\ll\exp(-c\log^{1/5}N)).$$

But we can bound

$$\frac{1}{d^\beta}\leq\frac{1}{d^{\beta-\log^{-1/10}N}}\exp(-c\log^{2/5}N)\leq\frac{1}{d^{1-2\log^{-1/10}N}}\exp(-c\log^{2/5}N)$$

when $d\in\mathcal{D}_>$, and from Euler products we have

$$\sum_{d\in\mathcal{D}}\frac{1}{d^{1-2\log^{-1/10}N}}\leq\prod_{p\leq Q}\left(1+\frac{1}{p^{1-2\log^{-1/10}N}}\right)\ll\exp(O(\log\log N)),$$

and the claim follows. ∎

We return now to the proof of Theorem 2.7. Throughout this section we assume that $\epsilon > 0$ is fixed and small in terms of k, and that $c_1(\epsilon) > 0$ is sufficiently small depending on k (and we reserve the right to decrease $c_1(\epsilon)$ later in the argument if necessary). We can assume that N is sufficiently large depending on $c_1(\epsilon)$, k, as the claim is trivial otherwise. Let P, G/Γ, F, g be as in that theorem. We use $m = O((\log\log N)^{c_1(\epsilon)})$ to denote the dimension of G; to avoid some minor notational issues we will assume that $m \geq 2$ (as can be achieved trivially by adding some dummy dimensions).

We repeat the arguments from [23], but now performing a more quantitative accounting of the dependence on constants (particularly on the dimension). We first use a dimension-uniform version of the factorization theorem in [24, Theorem 1.19], which we establish in Theorem A.6. We apply that theorem with $M_0 := \exp(\log^{1/10-\epsilon/2} N)$ and $A := \exp((\log\log N)^{1/2})$ to obtain a quantity

$$\exp(\log^{1/10-\epsilon/2} N) \leq M \leq \exp(\log^{1/10-\epsilon/3} N), \tag{7.8}$$

a subgroup $G' \subset G$ which is M-rational with respect to G, and a decomposition $g = \varepsilon g'\gamma$ into polynomial sequences $\varepsilon, g', \gamma\colon \mathbb{Z} \to G$ such that

(i) ε is (M, N)-smooth;

(ii) g' takes values in G' and $(g'(n)\Gamma)_{n\in[N]}$ is totally $1/M^A$-equidistributed in G'/Γ', with respect to a Mal'cev basis $\mathcal{X}'$ consisting of M-rational linear combinations of the basis elements of $\mathcal{X}$;

(iii) γ is M-rational and $\gamma(n)\Gamma$ is periodic with period at most M.

We can partition the arithmetic progression P into $O(M^{m^{O(1)}})$ components P' such that on each of these components the periodic function $\gamma(n)\Gamma$ is equal to an M-rational constant $\gamma_{P'}\Gamma$, and the smooth sequence ε differs by at most $O(M^{-m^C})$ from a constant $\varepsilon_{P'} \in G$ of distance at most M from the origin, for a large constant C. We can also normalize $\gamma_{P'}$ to be distance $O(M^{m^{O(1)}})$ from the origin. From this and the Lipschitz nature of F, we see (for C large enough) that

$$F(g(n)\Gamma) = F(\varepsilon_{P'} g'(n)\gamma_{P'}\Gamma) + O(M^{-1})$$

for $n \in P'$. By (7.8), the triangle inequality, and Lemma 2.4, it thus suffices to establish the bounds

$$\sum_{n\in P'} (\mu - \mu_{\mathrm{Siegel}})(n)\bar{F}(\varepsilon_{P'} g'(n)\gamma_{P'}\Gamma) \ll NM^{\exp(m^{O(1)})}(\exp(-\log^{1/10-\epsilon/4} N) + M^{-A/\exp(m^{O(1)})})$$

and

$$\sum_{n\in P'} (\Lambda - \Lambda_{\mathrm{Siegel}})(n)\bar{F}(\varepsilon_{P'} g'(n)\gamma_{P'}\Gamma) \ll NM^{\exp(m^{O(1)})}(\exp(-\log^{1/10-\epsilon/4} N) + M^{-A/\exp(m^{O(1)})})$$

for all of the progressions P', where the implied constants in the $O(1)$ notation on the right-hand sides of the estimates can be taken to be uniform in ϵ for ϵ sufficiently small. We introduce the conjugated group

$$G_{P'} := \gamma_{P'}^{-1} G' \gamma_{P'}$$

and conjugated polynomial

$$g_{P'} := \gamma_{P'}^{-1} g' \gamma_{P'}$$

that takes values in $G_{P'}$, and the normalized function

$$F_{P'}(x) := \bar{F}(\varepsilon_{P'}\gamma_{P'}x) - \int_{G_{P'}/(G_{P'}\cap\Gamma)} \bar{F}(\varepsilon_{P'}\gamma_{P'}\cdot),$$

where the integral is with respect to the Haar probability measure on $G_{P'}/(G_{P'} \cap \Gamma)$ (which we can view as a subnilmanifold of G/Γ). Using Proposition 2.2 to dispose of the contribution of the constant $\int_{G_{P'}/(G_{P'}\cap\Gamma)} \bar{F}(\varepsilon_{P'}\gamma_{P'}\cdot)$ (which can be viewed as the "major arc" contribution to these correlations), we are reduced to establishing the bounds

$$\sum_{n\in P'} (\mu - \mu_{\text{Siegel}})(n) F_{P'}(g_{P'}(n)\Gamma) \ll NM^{\exp(m^{O(1)})}\big(\exp(-\log^{1/10-\epsilon/4} N) + M^{-A/\exp(m^{O(1)})}\big)$$

and

$$\sum_{n\in P'} (\Lambda - \Lambda_{\text{Siegel}})(n) F_{P'}(g_{P'}(n)\Gamma) \ll NM^{\exp(m^{O(1)})}(\exp(-\log^{1/10-\epsilon/4} N) + M^{-A/\exp(m^{O(1)})}).$$

The advantages of this reduction are that the function $F_{P'}$ is not only 1-bounded and $O(M^{m^{O(1)}})$-Lipschitz (with respect to the Mal'cev basis of $G_{P'}/(G_{P'} \cap \Gamma)$, which is a filtered nilmanifold of complexity $O(M^{m^{O(1)}})$), but it also has mean zero. By repeating the arguments from [23, p. 547] and keeping track of the constants, we see that the polynomial sequence $g_{P'}$ is totally $1/M^{A/m^{O(1)}}$-equidistributed (note that multiplicative factors of $\exp(\exp(m^{O(1)}))$ can be absorbed into the $M^{A/m^{O(1)}}$ denominator, and all the $O_m(1)$ exponents appearing in this portion of [23] (and [24]) are polynomial in m).

We can use the Gowers uniformity of χ_{Siegel} to obtain the following bound on the Siegel terms which is acceptable when q_{Siegel} is large enough:

Proposition 7.1. *We have*

$$\sum_{n\in P'} \mu_{\text{Siegel}}(n) F_{P'}(g_{P'}(n)\Gamma) \ll NM^{m^{O(1)}} q_{\text{Siegel}}^{-1/m^{O(1)}}$$

and

$$\sum_{n\in P'} (\Lambda_{\text{Siegel}}(n) - \Lambda_{\text{Cramér},Q}(n)) F_{P'}(g_{P'}(n)\Gamma) \ll NM^{m^{O(1)}} q_{\text{Siegel}}^{-1/m^{O(1)}}.$$

Proof. We apply [22, Proposition 11.2], noting that all bounds[9] can be shown to be polynomial in the parameters M, ε with exponents that are polynomial in the dimension m, to decompose

$$F_{P'}(g_{P'}(n)\Gamma) = F_1(n) + F_2(n),$$

where F_1 obeys the dual norm bound

$$\mathbb{E}_{n\in[N]} f(n)F_1(n) \ll (M/\varepsilon)^{m^{O(1)}} \|f\|_{U^k[N]}$$

for any $f : [N] \to \mathbb{C}$, and F_2 obeys the pointwise bound

$$F_2(n) \ll \varepsilon$$

for all $n \in [N]$. Here $0 < \varepsilon \leq 1$ is a parameter that we are at liberty to choose. By Theorem 2.5, the functions μ_{Siegel}, $\Lambda_{\text{Siegel}} - \Lambda_{\text{Cramér},Q}$ already have a $U^k[N]$ norm of $O(q_{\text{Siegel}}^{-c})$; a standard Fourier expansion of $1_{P'}(n)$ in terms of additive characters and the triangle inequality then show that the truncated versions $1_{P'}\mu_{\text{Siegel}}$, $1_{P'}(\Lambda_{\text{Siegel}} - \Lambda_{\text{Cramér},Q})$ have a $U^k[N]$ norm of $O(M^{O(1)}q_{\text{Siegel}}^{-c})$ (note that any logarithmic factors can be easily absorbed into the $M^{O(1)}$ factor). Applying the above decomposition as well as Lemma 2.4, we see that

$$\sum_{n\in P'} \mu_{\text{Siegel}}(n)F_{P'}(g_{P'}(n)\Gamma) \ll NM^{O(1)}(M/\varepsilon)^{m^{O(1)}} q_{\text{Siegel}}^{-c} + \varepsilon N$$

and

$$\sum_{n\in P'} (\Lambda_{\text{Siegel}}(n) - \Lambda_{\text{Cramér},Q}(n))F_{P'}(g_{P'}(n)\Gamma) \ll NM^{O(1)}(M/\varepsilon)^{m^{O(1)}} q_{\text{Siegel}}^{-c} + \varepsilon N \log N,$$

and the claim then follows by a suitable choice of ε (noting that the $\log N$ factor can be absorbed into the M factor). ∎

Based on this proposition, we may now delete the Q-Siegel zero contributions except in the regime where

$$q_{\text{Siegel}} \leq M^{A/\exp(m^{C_1})} \tag{7.9}$$

[9] The argument as stated in that paper appeals to the Stone–Weierstrass theorem and the Arzelà–Ascoli theorem, but this can be replaced by more quantitative approximation results without difficulty, such as [20, Lemma A.9], combined with standard smooth partitions of unity to allow one to work on regions such as the unit cube rather than on the original nilmanifold. As pointed out to us by James Leng, the required smoothness bounds on the function P constructed in [22, Proposition 11.5] also need to be established. To do this, one can first take advantage of the fact that $\text{HK}^{s+1}(G)$ acts transitively on the graph of P to reduce to establishing smoothness bounds at the origin. Then one can lift from G/Γ to G, and reduce to establishing that one corner of a parallelepiped in $\text{HK}^{s+1}(G)$ is a smooth function of all the other corners near the origin with the required bounds. But one can express the first corner as a word in the other corners of length depending only on s, and from many applications of the Baker–Campbell–Hausdorff formula this will give the desired quantitative bounds on this corner completion function.

where C_1 is a large constant depending on k (but not on ϵ) that we are at liberty to choose; we can also assume N to be sufficiently large depending on C_1 (as well as k and ϵ). To simplify the notation we assume henceforth that the Q-Siegel zero exists and obeys (7.9); the remaining cases follow by a simplified version of the same argument that deletes all the steps and terms that treat the contribution of the Q-Siegel zero. It will now suffice to obtain estimates of the form

$$\sum_{n\in P'} (\mu-\mu_{\text{Siegel}})(n) F_{P'}(g_{P'}(n)\Gamma) \ll N(Mq_{\text{Siegel}})^{\exp(m^{O(1)})}\big(\exp(-\log^{1/10-\epsilon/4} N) + M^{-A/\exp(m^{O(1)})}\big)$$

and

$$\sum_{n\in P'} (\Lambda-\Lambda_{\text{Siegel}})(n) F_{P'}(g_{P'}(n)\Gamma) \ll N(Mq_{\text{Siegel}})^{\exp(m^{O(1)})}\big(\exp(-\log^{1/10-\epsilon/4} N) + M^{-A/\exp(m^{O(1)})}\big),$$

where the implied constants do not depend on C_1.

To treat these sums, we make the following standard Vaughan-type decompositions. Call a sequence $a_d, d \in \mathbb{N}$ of complex numbers *divisor bounded* if one has

$$a_d \ll (\log N)^{O(1)} \tau^{O(1)}(d)$$

for all $d \in [N]$, where $\tau(n) := \sum_{d|n} 1$ is the divisor function.

Proposition 7.2 (Vaughan-type decompositions). *Any of the four functions μ, μ_{Siegel}, Λ, Λ_{Siegel} on $[N]$ can be expressed as a convex linear combination of functions of one of the following four classes (with uniform constants in the bounds):*

(i) (Type I sum) *A function of the form*

$$n \mapsto \sum_{d\le N^{2/3}} a_d 1_{d|n} 1_{[N']}(n),$$

where the coefficients a_d are divisor-bounded and $1 \le N' \le N$.

(ii) (Twisted type I sum) *A function of the form*

$$n \mapsto \sum_{d\le N^{2/3}} a_d 1_{d|n} \chi_{\text{Siegel}}(n/d) 1_{[N']}(n),$$

where the coefficients a_d are divisor-bounded and $1 \le N' \le N$.

(iii) (Type II sum) *A function of the form*

$$n \mapsto \sum_{d,w>N^{1/3}} a_d b_w 1_{dw=n}$$

for some divisor-bounded coefficients a_d, b_w.

(iv) (Negligible sum) *A divisor-bounded function $n \mapsto f(n)$ with*

$$\sum_{n\in[N]} |f(n)| \ll N \exp(-\log^{1/2} N).$$

Proof. For Λ we can use the familiar Vaughan identity [43]

$$\Lambda(n) = \Lambda(n)1_{n\leq N^{1/3}} - \sum_{d\leq N^{2/3}} a_d 1_{d|n} + \sum_{d\leq N^{1/3}} \mu(d)1_{d|n} \log\frac{n}{d} + \sum_{d,w>N^{1/3}} \Lambda(d)b_w 1_{dw=n},$$

where $a_d := \sum_{bc=d:b,c\leq N^{1/3}} \mu(b)\Lambda(c)$ and $b_w := \sum_{c|w:c>N^{1/3}} \mu(c)$. The first term is negligible, the second term is a Type I sum (restricting to $[N]$), and the fourth term is a Type II sum; the third term can be converted to a convex combination of Type I sums by using the fundamental theorem of calculus to write

$$\log\frac{n}{d} = \log N - \int_1^N 1_{t>n}\frac{dt}{t} - \log d$$

and absorbing all the various logarithmic factors into the divisor-bounded coefficients. Similarly, for μ we can use the variant identity

$$\mu(n) = \sum_{d\leq N^{2/3}} a'_d 1_{d|n} - \sum_{d,w>N^{1/3}} \mu(d)b_w 1_{dw=n},$$

where $a'_d := \sum_{bc=d:b,c\leq N^{1/3}} \mu(b)\mu(c)$ and b_w is as before; see e.g. [20, Lemma 4.1].

To handle Λ_{Siegel}, it suffices (using the estimate $P(Q)/\phi(P(Q)) \ll (\log N)^{O(1)}$ coming from Mertens' theorem) to show that the functions

$$n \mapsto 1_{(n,P(Q))=1} \tag{7.10}$$

and

$$n \mapsto n^{\beta-1}1_{(n,P(Q))=1}\chi_{\text{Siegel}}(n)$$

can be expressed in the desired form (absorbing all the constant factors into the divisor-bounded coefficients). But if λ_d^+, λ_d^- are the upper and lower linear sieve coefficients, respectively, with level $D = Q^{10(\log N)^{3/5}}$ and sifting parameter Q, one can write

$$\sum_{d\leq D} \lambda_d^- 1_{d|n} \leq 1_{(n,P(Q))} \leq \sum_{d\leq D} \lambda_d^+ 1_{d|n},$$

and by the fundamental lemma [29, Lemma 6.3] (bounding the error terms $R^{\pm}$ there as $O(D)$) we have

$$\sum_{n\in[N]} \left| 1_{(n,P(Q))} - \sum_{d\leq D} \lambda_d^{\pm} 1_{d|n} \right| \ll N \exp(-10\log^{3/5} N)$$

(say). Therefore, one can express (7.10) as a Type I sum plus an error term of $L^1[N]$ norm $\ll N\exp(-10\log^{3/5} N)$, and by multiplying by χ_{Siegel} one can then express

$n \mapsto 1_{(n,P(Q))=1}\chi_{\mathrm{Siegel}}(n)$ as a twisted Type I sum plus an error term of $L^1[N]$ norm at most $\ll N\exp(-10\log^{3/5} N)$. Indeed, in these cases one can lower the $N^{2/3}$ threshold on d to something much smaller, such as $\exp(O(\log^{7/10} N))$. Finally, the $n^{\beta-1}$ weight can be handled using the fundamental theorem of calculus identity (5.12).

Now we turn to $\mu_{\mathrm{Siegel}} = \mu_{\mathrm{local}} * \mu'$. From the previous discussion and Lemma 5.5, μ' is already expressible as a convex combination of twisted Type I sums (where d can be constrained to be at most $\exp(O(\log^{7/10} N))$) plus an error term of $L^1[N]$ norm $\ll N\exp(-10\log^{3/5} N)$. We can then convolve by $\mu_{\mathrm{local}} 1_{[\exp(5\log^{3/5} N)]}$ and conclude that $\mu_{\mathrm{local}} 1_{[\exp(5\log^{3/5} N)]} * \mu'$ is also expressible as a convex combination of twisted Type I sums plus a negligible error (note that the values of d encountered stay well below the threshold $N^{2/3}$). Finally, the remaining term $\mu_{\mathrm{local}}(1 - 1_{[\exp(5\log^{3/5} N)]}) * \mu'$ can be seen to be negligible by the same arguments used to dispose of the $\mathcal{D}_>$ contributions to (5.18) (namely, using the fact that the density of Q-smooth numbers in any dyadic interval $[M, 2M]$ with $\exp(5\log^{3/5} N) \leq M \leq N$ is $\ll \exp(-5(\log^{1/2} N))$. ∎

The contributions of the negligible sums to the previous estimates are acceptable from the triangle inequality. By a further application of the triangle inequality, it thus suffices to establish the bound

$$\sum_{n\in P'} f(n)F_{P'}(g_{P'}(n)\Gamma) \ll N(Mq_{\mathrm{Siegel}})^{\exp(m^{O(1)})} M^{-A/\exp(m^{O(1)})} \tag{7.11}$$

whenever f is a Type I sum, a twisted Type I sum, or a Type II sum.

The Type I and Type II sums were already essentially treated in [23, Section 3], and it turns out that the methods also easily extend to cover the twisted Type I case. We briefly review the argument as follows. We begin with the twisted Type I case; the Type I case is treated by a simplification of the argument that deletes the role of the Q-Siegel character, and is omitted here (and in any case would follow closely the treatment in [23, Section 3]). Suppose that

$$\Big|\sum_{n\in P'} f(n)F_{P'}(g_{P'}(n)\Gamma)\Big| \geq \delta N \tag{7.12}$$

for some $0 < \delta < \frac{1}{Mq_{\mathrm{Siegel}}}$ and a twisted Type I sum f. By the definition of such sums and the triangle inequality, this implies that

$$\sum_{d\leq N^{2/3}} \tau^C(d)\Big|\sum_{n\in P''\cap d\mathbb{Z}} \chi_{\mathrm{Siegel}}(n/d)F_{P'}(g_{P'}(n)\Gamma)\Big| \gg \delta^{O(1)}N$$

for some constant $C = O(1)$, where $P'' := P' \cap [N']$ (note that all $\log^{O(1)} N$ terms can be easily absorbed into the $\delta^{O(1)}$ factor). Standard divisor sum estimates give

$$\sum_{d\leq N^{2/3}} \tau^{2C}(d)/d \ll \log^{O(1)} N$$

(with the implied constant depending on C), hence by Cauchy–Schwarz,

$$\sum_{d\leq N^{2/3}} d\Big|\sum_{n\in P''\cap d\mathbb{Z}} \chi_{\mathrm{Siegel}}(n/d)F_{P'}(g_{P'}(n)\Gamma)\Big|^2 \gg \delta^{O(1)}N^2,$$

and hence by dyadic decomposition there exists $1 \leq D \leq N^{2/3}$ such that

$$\sum_{D \leq d \leq 2D} \Big| \sum_{n \in P'' \cap d\mathbb{Z}} \chi_{\mathrm{Siegel}}(n/d) F_{P'}(g_{P'}(n)\Gamma)\Big|^2 \gg \delta^{O(1)} \frac{N^2}{D}.$$

Since the inner sum is $O(N/D)$, we conclude that

$$\Big| \sum_{n \in P'' \cap d\mathbb{Z}} \chi_{\mathrm{Siegel}}(n/d) F_{P'}(g_{P'}(n)\Gamma)\Big| \gg \delta^{O(1)} \frac{N}{D}$$

for $\gg \delta^{O(1)} D \log^{-O(1)} N$ natural numbers d in $[D, 2D]$. For such a d, we partition into residue classes modulo dq_{Siegel} and use the triangle inequality to conclude that

$$\Big| \sum_{n \in [N_d]} F_{P'}(g_{P'}(d(q_{\mathrm{Siegel}} n + a_d))\Gamma)\Big| \gg \delta^{O(1)} \frac{N}{D}$$

for some $1 \leq N_d \leq N/D$ and $1 \leq a_d \leq q_{\mathrm{Siegel}}$ (note that all q_{Siegel} factors can be absorbed into the $\delta^{O(1)}$ factor). Applying Theorem A.3, we can then find a horizontal character η_d of G' with

$$0 < |\eta_d| \ll \delta^{-\exp(m^{O(1)})} \tag{7.13}$$

such that

$$\|\eta_d \circ g_{P'}(d(q_{\mathrm{Siegel}} \cdot + a_d))\|_{C^\infty[N/D]} \ll \delta^{-\exp(m^{O(1)})},$$

where $\|\cdot\|_{C^\infty}$ is defined in [24, Definition 2.7]. The parameter a_d is annoying, but we can remove it[10] by applying [24, Lemma 8.4] to conclude that

$$\|\eta'_d \circ g_{P'}(d(q_{\mathrm{Siegel}} \cdot))\|_{C^\infty[N/D]} \ll \delta^{-\exp(m^{O(1)})}$$

for some η'_d that continues to obey (7.13). The total number of such η'_d is $O(\delta^{-\exp(m^{O(1)})})$. Thus by the pigeonhole principle, we can find one such horizontal character η such that

$$\|\eta \circ g_{P'}(d(q_{\mathrm{Siegel}} \cdot))\|_{C^\infty[N/D]} \ll \delta^{-\exp(m^{O(1)})}$$

for $\gg \delta^{\exp(m^{O(1)})} D$ values of $d \in [D, 2D]$. If we expand out the polynomial

$$\eta \circ g_{P'}(q_{\mathrm{Siegel}} n) = \beta_k n^k + \cdots + \beta_0 \bmod 1 \tag{7.14}$$

for some real numbers $\beta_0, \ldots, \beta_k$, then by applying [23, Lemma 3.2] we conclude that there is a positive integer $q = O(1)$ such that

$$\|q d^j \beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll (N/D)^{-j} \delta^{-\exp(m^{O(1)})}$$

for all $j = 0, \ldots, k$, where $\|x\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance to the nearest integer. Applying a Waring-type result from [23, Lemma 3.3], we then have, for each $j = 0, \ldots, k$,

$$\|q d' \beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll (N/D)^{-j} \delta^{-\exp(m^{O(1)})}$$

[10] We thank the anonymous referee for this suggestion, which patched a gap in a previous version of this argument.

for $\gg \delta^{\exp(m^{O(1)})} D^j$ integers d' of size $d' = O(D^j)$. Applying Vinogradov's lemma [23, Lemma 3.4], and clearing denominators, we then conclude that there is a positive integer $K \ll \delta^{\exp(m^{O(1)})}$ such that

$$\|K\beta_j\|_{\mathbb{R}/\mathbb{Z}} \ll N^{-j} \delta^{-\exp(m^{O(1)})}$$

for all $j = 0, \dots, k$, and thus by (7.14),

$$\|K q_{\mathrm{Siegel}}^k \eta \circ g_{P'}\|_{C^\infty[N]} \ll \delta^{-\exp(m^{O(1)})}.$$

On the other hand, $g_{P'}$ is totally $1/M^{A/m^{O(1)}}$-equidistributed. Arguing as in [23, Section 3] and noting that all exponents of the form $O_m(1)$ are in fact polynomial in m, these two facts are incompatible unless

$$\delta^{-\exp(m^{O(1)})} \gg M^{A/m^{O(1)}}, \tag{7.15}$$

which (when combined with the constraint $\delta \leq \frac{1}{M q_{\mathrm{Siegel}}}$) gives the desired bound (7.11).

For the Type II case, we can again start by assuming (7.12) for some $0 < \delta < 1/M$ and some Type II sum f. The contribution of those n less than $\delta^C N$ for a large absolute constant C can easily be seen to be negligible, so one can assume without loss of generality that $|P'|$ lies in the interval $[\delta^C N, N]$. One has

$$\sum_{d > N^{1/3}} \sum_{w > N^{1/3}} a_d b_w F_{P'}(g_{P'}(dw)\Gamma) 1_{P'}(dw) \gg \delta^{O(1)} N$$

for some divisor-bounded a_d, b_w, and then after some dyadic decomposition and Cauchy–Schwarz (cf. [20, Proposition 7.2]) one can find $N^{1/3} \ll D, W \ll \delta^{-O(1)} N^{2/3}$ with $DW = \delta^{O(1)} N$ such that

$$\sum_{d,d' \in [D,2D]} \sum_{w,w' \in [W,2W]} F_{P'}(g_{P'}(dw)\Gamma)\overline{F_{P'}}(g_{P'}(dw')\Gamma)\overline{F_{P'}}(g_{P'}(d'w)\Gamma) \times F_{P'}(g_{P'}(d'w')\Gamma) \gg \delta^{O(1)} N.$$

One now repeats the arguments used to treat the Type II case in [23, Section 3] more or less verbatim (noting that all exponents are of order $\exp(m^{O(1)})$ at worst) to obtain a contradiction to the total $1/M^{A/m^{O(1)}}$-equidistribution of $g_{P'}$ unless (7.15) holds, and we again obtain (7.11) as desired. This concludes the proof of Theorem 2.7.

8. Applying densification

We now use densification methods to establish a general transference principle (which seems of independent interest) that converts inverse theorems for the Gowers norms for 1-bounded functions to inverse theorems for Gowers norms for ν-bounded functions for various "pseudorandom" weights ν. Our pseudorandomness condition will be relatively mild (a U^{2k} estimate on $\nu - 1$), and the losses in the transference argument will only

be polynomial in nature. However, one drawback of the theorem is that the input inverse theorem must also have polynomial bounds.

In Section 8.2, we will use Theorem 8.1 to complete the proof of Theorem 2.6 in the von Mangoldt case.

8.1. Transferring inverse theorems

Theorem 8.1 (Transference principle for U^k inverse theorems). *Let $k \geq 2$ be fixed. Let $G = (G, +)$ be a finite abelian group. Suppose that for every $0 < \delta \leq 1/2$ there is a family Ψ_δ of 1-bounded functions $\psi : G \to \mathbb{C}$, non-increasing in δ and closed under translations and complex conjugation, obeying the following U^k inverse theorem:*

(i) *If $0 < \delta \leq 1/2$ and $f : G \to \mathbb{C}$ is 1-bounded with $\|f\|_{U^k(G)} \geq \delta$, then there exists $\psi \in \Psi_\delta$ such that $|\mathbb{E}_{x\in G} f(x)\overline{\psi(x)}| \gg \delta^B$ for some $B > 0$.*

Let C_0 be sufficiently large depending on k, let $0 < \delta \leq 1/2$, and let $\nu : G \to \mathbb{R}^+$ be a weight with

$$\|\nu - 1\|_{U^{2k}(G)} \leq \delta^{C_0}. \tag{8.1}$$

Let $f : G \to \mathbb{C}$ be ν-bounded with

$$\|f\|_{U^k(G)} \geq \delta. \tag{8.2}$$

Then there exist $\psi_1, \dots, \psi_{2^k-1} \in \Psi_{\delta^{O(1)}}$ such that

$$\left|\mathbb{E}_{x\in G} f(x) \prod_{j=1}^{2^k-1} \overline{\psi}_j(x)\right| \gg \delta^{O(1)}.$$

We remark that this theorem strengthens a similar result in [9], in that the class Ψ_δ is allowed to be more general than the space of "dual functions", and the bounds are polynomial in nature rather than qualitative.

We now begin the proof of this theorem. Let the notation and hypotheses be as in Theorem 8.1. From (8.2) we have

$$\left|\mathbb{E}_{(x,\vec{h})\in G^{k+1}} \prod_{\omega\in\{0,1\}^k} f_\omega(x + \omega \cdot \vec{h})\right| \gg \delta^{O(1)}, \tag{8.3}$$

where $f_0 = f$, and all the other $f_\omega : G \to \mathbb{C}$ are either equal to f or its complex conjugate. The key step is

Proposition 8.2 (Densification of a single factor). *Suppose that the bound (8.3) holds for some $\nu + 1$-bounded functions $f_\omega, \omega \in \{0,1\}^k$. Let $\omega_0 \in \{0,1\}^k$. Then*

$$\left|\mathbb{E}_{(x,\vec{h})\in G^{k+1}} \prod_{\omega\in\{0,1\}^k} \tilde{f}_\omega(x + \omega \cdot \vec{h})\right| \gg \delta^{O(1)},$$

where $\tilde{f}_\omega = f_\omega$ for $\omega \in \{0,1\}^k \backslash \{\omega_0\}$, and $\tilde{f}_{\omega_0} \in \Psi_{\delta^{O(1)}}$.

Indeed, after applying this proposition $2^k - 1$ times starting with (8.3), we conclude that

$$\left|\mathbb{E}_{(x,\vec{h})\in G^{k+1}} f(x) \prod_{\omega\in\{0,1\}^k\backslash\{0\}^k} \psi_\omega(x+\omega\cdot\vec{h})\right| \gg \delta^{O(1)}$$

for some $\psi_\omega \in \Psi_{\delta^{O(1)}}$ and all $\omega \in \{0,1\}^k\backslash\{0\}^k$ (one can use the non-decreasing nature of Ψ to make the implied constant in $O(1)$ uniform in ω). In particular, by the pigeonhole principle there exist $h_1,\dots,h_k \in G$ such that

$$\left|\mathbb{E}_{x\in G} f(x) \prod_{\omega\in\{0,1\}^k\backslash\{0\}^k} \psi_\omega(x+\omega\cdot\vec{h})\right| \gg \delta^{O(1)},$$

giving Theorem 8.1 thanks to the translation and conjugation invariance of $\Psi_{\delta^{O(1)}}$.

It remains to prove Proposition 8.2. By relabeling we may assume $\omega_0 = 0^k$. By replacing ν with $\frac{\nu+1}{2}$ (and adjusting C_0 if necessary), and then rescaling by various factors of 2, we may assume that the f_ω are ν-bounded rather than $\nu+1$-bounded. Now we adapt the arguments of Conlon–Fox–Zhao [5]. We have

$$|\mathbb{E}_{x\in G} f_{0^k}(x)F(x)| \gg \delta^{O(1)},$$

where $F\colon G\to\mathbb{C}$ is the dual function

$$F(x) := \mathbb{E}_{\vec{h}\in G^k} \prod_{\omega\in\{0,1\}^k\backslash\{0\}^k} f_\omega(x+\omega\cdot\vec{h}).$$

Since f_{0^k} is ν-bounded, we conclude from Cauchy–Schwarz that

$$(\mathbb{E}_{x\in G}\nu(x))(\mathbb{E}_{x\in G}\nu(x)|F(x)|^2) \gg \delta^{O(1)}.$$

Since

$$\mathbb{E}_{x\in G}\nu(x) = \|\nu\|_{U^1(G)} \le \|\nu\|_{U^k(G)} \le 1 + \|\nu-1\|_{U^k(G)} \ll 1,$$

we conclude that

$$\mathbb{E}_{x\in G}\nu(x)|F(x)|^2 \gg \delta^{O(1)}. \tag{8.4}$$

Next we claim that

$$\mathbb{E}_{x\in G}(\nu-1)(x)|F(x)|^2 \ll \delta^{C_0}. \tag{8.5}$$

We can write the left-hand side of (8.5) as

$$\mathbb{E}_{(x,\vec{h})\in G^{2k+1}} \prod_{\omega\in\{0,1\}^{2k}} f_\omega(x+\omega\cdot\vec{h}),$$

where

$$f_{0^{2k}}(x) := \nu(x) - 1, \quad f_{\vec{\omega},0^k}(x) := f_{\vec{\omega}}(x), \quad f_{0^k,\vec{\omega}}(x) := \overline{f_{\vec{\omega}}}(x)$$

for $\vec{\omega}\in\{0,1\}^k\backslash\{0\}^k$, and $f_\omega(x) := 1$ for all other $\omega\in\{0,1\}^{2k}$ not covered by the preceding definitions. By the Gowers–Cauchy–Schwarz inequality (4.1), we thus have

$$\mathbb{E}_{x\in G}(\nu-1)(x)|F(x)|^2 \le \prod_{\omega\in\{0,1\}^{2k}} \|f_\omega\|_{U^{2k}(G)} \le \|\nu-1\|_{U^{2k}(G)}\|\nu+1\|_{U^{2k}(G)}^{2^{2k}-1},$$

and the claim now follows from (8.1) and the triangle inequality.

From (8.4), (8.5) and the triangle inequality we conclude (for C_0 large enough) that

$$\mathbb{E}_{x\in G}|F(x)|^2 \gg \delta^{O(1)}. \tag{8.6}$$

The function F is not quite bounded. However, as the f_ω are all ν-bounded, we certainly have the pointwise bound $|F| \leq \mathcal{D}\nu$, where $\mathcal{D}\nu$ is the dual function

$$\mathcal{D}\nu(x) := \mathbb{E}_{h\in G^k} \prod_{\omega\in\{0,1\}^k\backslash\{0\}^k} \nu(x+\omega\cdot\vec{h}).$$

We observe the moment estimates

$$\mathbb{E}_{x\in G}\mathcal{D}\nu(x)^j = 1 + O(\delta^{C_0}) \tag{8.7}$$

for $j = 0, 1, 2$. We just prove this for $j = 2$, as the $j = 0, 1$ claims are similar (and easier). We can expand

$$\mathbb{E}_{x\in G}\mathcal{D}\nu(x)^2 = \mathbb{E}_{(x,\vec{h})\in G^{2k+1}} \prod_{\omega\in\{0,1\}^{2k}} g_\omega(x+\omega\cdot\vec{h}),$$

where

$$g_{\vec{\omega},0^k}(x) := \nu(x), \quad g_{0^k,\vec{\omega}}(x) := \nu(x)$$

for $\vec{\omega} \in \{0,1\}^k\backslash\{0\}^k$, and $g_\omega(x) := 1$ for all other $\omega \in \{0,1\}^{2k}$ not covered by the preceding definitions. We split each g_ω that is of the form ν into 1 and $\nu - 1$. Applying the triangle inequality (1.2) and the Gowers–Cauchy–Schwarz inequality (4.1), we can thus write

$$\mathbb{E}_{x\in G}\mathcal{D}\nu(x)^2 = 1 + O(\|\nu-1\|_{U^{2k}(G)}(1+\|\nu-1\|_{U^{2k}(G)})^{2^{2k}-1}),$$

and the claim follows from (8.1).

From (8.7) we have

$$\mathbb{E}_{x\in G}|\mathcal{D}\nu(x) - 1|^2 \ll \delta^{C_0}. \tag{8.8}$$

Now define the truncated version

$$\tilde{F}(x) := \min(|F(x)|, 1)\operatorname{sgn}(F(x)),$$

where $\operatorname{sgn}(F(x))$ is equal to $F(x)/|F(x)|$ when $F(x) \neq 0$ and equal to zero when $F(x) = 0$. Then $\tilde{F}$ is 1-bounded and

$$|F(x) - \tilde{F}(x)| \leq \max(|F(x)| - 1, 0) \leq |\mathcal{D}\nu(x) - 1|, \tag{8.9}$$

so from (8.8) and Cauchy–Schwarz we have

$$\mathbb{E}_{x\in G}\overline{F}(x)(F(x) - \tilde{F}(x)) \leq \mathbb{E}_{x\in G}\big(|F(x) - \tilde{F}(x)|^2 + |F(x) - \tilde{F}(x)|\big) \ll \delta^{C_0/2}.$$

Hence by (8.6) and the triangle inequality we have

$$|\mathbb{E}_{x\in G}\overline{F}(x)\tilde{F}(x)| \gg \delta^{O(1)}.$$

We rewrite the left-hand side as

$$\Big|\mathbb{E}_{\vec{h}\in G^k}\prod_{\omega\in\{0,1\}^k} f_\omega^*(x+\omega\cdot\vec{h})\Big|,$$

where

$$f_{0^k}^*(x) := \tilde{F}(x), \quad f_\omega^*(x) := \overline{f_\omega}(x)$$

for $\omega \in \{0,1\}^k\backslash\{0\}^k$. The f_ω^* all have $U^k(G)$ norm of at most $\|\nu\|_{U^k(G)} \ll 1$ thanks to (8.1), hence by the Gowers–Cauchy–Schwarz inequality (4.1) one has

$$\|\tilde{F}\|_{U^k(G)} \gg \delta^{O(1)}.$$

Applying the hypothesis in Theorem 8.1 (i), we conclude that there exists $\psi \in \Psi_{\delta^{O(1)}}$ such that

$$|\mathbb{E}_{x\in G}\tilde{F}(x)\psi(x)| \gg \delta^{O(1)}.$$

On the other hand, from Cauchy–Schwarz we have

$$\mathbb{E}_{x\in G}(F(x)-\tilde{F}(x))\psi(x) \ll (\mathbb{E}_{x\in G}|F(x)-\tilde{F}(x)|^2)^{1/2} \ll \delta^{C_0/2}$$

thanks to (8.8), (8.9). Hence by the triangle inequality (for C_0 large enough) we have

$$\mathbb{E}_{x\in G}F(x)\psi(x) \gg \delta^{O(1)}.$$

But this rearranges to give the conclusion of Proposition 8.2. The proof of Theorem 8.1 is now complete.

We now combine this theorem with Manners' inverse theorem to obtain

Theorem 8.3 (Transferred inverse theorem). *Let $0 < \delta < 1/2$, and let $\nu\colon [N] \to \mathbb{C}$ be such that*

$$\|\nu - 1\|_{U^{2k}[N]} \le \delta^{C_0}$$

for some constant C_0 that is sufficiently large depending on k. Let $f\colon [N] \to \mathbb{C}$ be a ν-bounded function such that

$$\|f\|_{U^k[N]} \ge \delta. \tag{8.10}$$

Then there exist a (filtered) nilmanifold G/Γ of degree $k-1$, dimension $O(\delta^{-O(1)})$, and complexity at most $\exp\exp(O(1/\delta^{O(1)}))$, a 1-bounded Lipschitz function $F\colon G/\Gamma \to \mathbb{C}$ of Lipschitz constant at most $\exp\exp(O(1/\delta^{O(1)}))$, and a polynomial map $g\colon \mathbb{Z} \to G$, such that

$$|\mathbb{E}_{n\in[N]}f(n)\overline{F}(g(n)\Gamma)| \gg \exp(-\exp(O(1/\delta^{O(1)}))).$$

Proof. As in the proof of Theorem 6.2, we pick a prime N' with $10N \le N' \le 20N$ and extend f by zero to $\mathbb{Z}/N'\mathbb{Z}$; we also extend ν by 1 to $\mathbb{Z}/N'\mathbb{Z}$, and observe that $\|\nu - 1\|_{U^{2k}(\mathbb{Z}/N'\mathbb{Z})} \ll \delta^{C_0}$.

To apply Theorem 8.1, we will need an inverse theorem that has polynomial correlation bounds. This is not directly provided by Theorem 6.2; however, such an inverse

theorem does appear in the work of Manners [33]. Indeed, we see from [33, Lemmas 5.4.1, 5.5.1] (applying [33, Lemma 5.5.1] inductively, as in [33, p. 102]), that if $f: \mathbb{Z}/N'\mathbb{Z} \to \mathbb{C}$ is 1-bounded with $\|f\|_{U^k(\mathbb{Z}/N'\mathbb{Z})} \geq \delta$, then there exists a 1-bounded function $\psi: \mathbb{Z}/N'\mathbb{Z} \to \mathbb{C}$ with the polynomial correlation bound

$$|\mathbb{E}_{n \in \mathbb{Z}/N'\mathbb{Z}} f(n)\overline{\psi(n)}| \gg \delta^{O(1)}$$

such that ψ is of the form

$$\psi(n) = \sum_{i=1}^{T} \alpha_i \bar{F}_i(g_i(n)\Gamma_i),$$

where $T \ll \exp(\exp(\delta^{-O(1)}))$, the α_i are complex numbers with $|\alpha_i| \leq 1$, and for each i, G_i/Γ_i is a filtered nilmanifold of degree $k-1$, dimension $O(\delta^{-O(1)})$, and complexity at most $\exp\exp(O(1/\delta^{O(1)}))$, $F_i: G_i/\Gamma_i \to \mathbb{C}$ is a 1-bounded Lipschitz function of Lipschitz constant at most $\exp\exp(O(1/\delta^{O(1)}))$, and $g_i: \mathbb{Z} \to G_i$ is a polynomial map with $g_i\Gamma$ periodic with period N'. Let us call the collection of all such ψ (with appropriate choices of implied constants) $\mathcal{F}_\delta$; note that this collection is invariant under translation and complex conjugation. We may now apply Theorem 8.1 to the ν-bounded function f in the hypotheses of this theorem, and conclude that there exist $\psi_1, \dots, \psi_{2^k-1} \in \mathcal{F}_{\delta^{O(1)}}$ such that

$$\left|\mathbb{E}_{x \in \mathbb{Z}/N'\mathbb{Z}} f(x) \prod_{j=1}^{2^k-1} \overline{\psi_j}(x)\right| \gg \delta^{O(1)}.$$

Applying the pigeonhole principle, and taking the tensor product of various nilsequences, we conclude a correlation

$$|\mathbb{E}_{n \in \mathbb{Z}/N'\mathbb{Z}} f(n)\bar{F}(g(n)\Gamma)| \gg \exp(-\exp(\delta^{-O(1)})),$$

where G/Γ is a filtered nilmanifold of degree $k-1$, dimension $O(\delta^{-O(1)})$, and complexity at most $\exp\exp(O(1/\delta^{O(1)}))$, $F: G/\Gamma \to \mathbb{C}$ is a 1-bounded Lipschitz function of Lipschitz constant at most $\exp\exp(O(1/\delta^{O(1)}))$, and $g: \mathbb{Z} \to G$ is a polynomial map with $g\Gamma$ periodic of period N'. Now argue as in the proof of Theorem 6.2 to conclude. ∎

8.2. *Completing the proof of the main theorem*

Now we can show how the bound (2.11) in Theorem 2.6 follows from the bound (2.13) given by Theorem 2.7. This will complete the proof of Theorem 2.6 and hence that of Theorem 1.4. We begin with an application of the "W-trick". Let $W := P(\log^{\varepsilon} N)$, where $\varepsilon > 0$ is a small constant depending on k to be chosen later; we may assume that N is sufficiently large depending on ε. Observe that the set $\{n \in [N] : (n, W) = 1\}$ contains the entire support of Λ_{Siegel}, as well as the support of Λ except for $O(\log^{O(1)} N)$ numbers which give a negligible contribution to the $U^k[N]$ norm. Thus it will suffice to show the doubly logarithmic decay bound

$$\|(\Lambda - \Lambda_{\text{Siegel}})1_{(\cdot,W)=1}\|_{U^k[N]} \ll (\log\log N)^{-c}.$$

By Corollary 4.2, this will follow once we show that

$$\left\| \frac{\phi(W)}{W}(\Lambda - \Lambda_{\text{Siegel}})(W\cdot + b) \right\|_{U^k[\frac{N-b}{W}]} \ll (\log\log N)^{-c} \tag{8.11}$$

for all $1 \le b \le W$ coprime to W.

Fix b. Now we use a quantitative variant of the well known fact (see [19]) that $\frac{\phi(W)}{W}\Lambda - 1$ can be bounded by a pseudorandom weight, but now observing that we can attain logarithmic accuracy in the pseudorandomness bound.

Proposition 8.4. *$\frac{\phi(W)}{W}(\Lambda - \Lambda_{\text{Siegel}})(W\cdot + b)$ is $C\nu$-bounded for some $C = O(1)$ depending only on k and some $\nu\colon [\frac{N-b}{W}] \to \mathbb{R}^+$ with $\|\nu - 1\|_{U^{2k}[\frac{N-b}{W}]} \ll \log^{-c\varepsilon} N$.*

Proof. By the triangle inequality (1.2), it suffices to establish this for $\frac{\phi(W)}{W}\Lambda(W\cdot + b)$ and $\frac{\phi(W)}{W}\Lambda_{\text{Siegel}}(W\cdot + b)$ separately. In the latter case, we see from Definition 2.1 that

$$\left| \frac{\phi(W)}{W}\Lambda_{\text{Siegel}}(Wn + b) \right| \le \frac{\phi(W)}{W}\Lambda_{\text{Cramér},Q}(Wn + b)$$

and the claim in this case follows from Corollary 5.3.

Now we turn to $\frac{\phi(W)}{W}\Lambda(W\cdot + b)$. Here we can basically follow the analysis of Goldston–Yıldırım correlation estimates from [22, Appendix D], though with a slightly more careful accounting in order to obtain suitable estimates. We choose a smooth function $\chi\colon \mathbb{R} \to \mathbb{R}_{\ge 0}$ supported on $[-2, 2]$ that equals $1/2$ on $[-1, 1]$ with $\int_1^2 \chi'(x)^2\,dx = 1$. We set $R := N^\gamma$ for some sufficiently small constant $0 < \gamma < 1/2$ depending only on k (and independent of ε). Following [22, Appendix D], we introduce the truncated divisor sum

$$\Lambda_{\chi,R,2}(n) := \log R \cdot \left(\sum_{d|n} \mu(d)\chi\left(\frac{\log d}{\log R}\right) \right)^2.$$

From [22, Lemma D.2] and the choice of χ, the sieve factor $c_{\chi,2} = \int_0^\infty |\chi'(x)|^2\,dx$ associated to this divisor sum via [22, Definition D.1] is simply

$$c_{\chi,2} = 1. \tag{8.12}$$

We then set

$$\nu(n) := \frac{\phi(W)}{W}\Lambda_{\chi,R,2}(Wn + b). \tag{8.13}$$

Let Λ' be the restriction of Λ to those primes greater than R^2. It is not difficult to see that the error $\frac{\phi(W)}{W}\Lambda(W\cdot + b) - \frac{\phi(W)}{W}\Lambda'(W\cdot + b)$ (supported on primes up to R^2, as well as powers of primes, and bounded in size by $O(\log N)$) is non-negative with $U^{2k}[\frac{N-b}{W}]$ norm as small as $O(N^{-c})$, so by (1.2) we may freely replace $\frac{\phi(W)}{W}\Lambda(W\cdot + b)$ with $\frac{\phi(W)}{W}\Lambda'(W\cdot + b)$. By the definition of ν in (8.13) and the fact that $\chi(0) = 1/2$, we easily verify the pointwise bound

$$0 \le \frac{\phi(W)}{W}\Lambda'(Wn + b) \ll_\gamma \nu(n)$$

for all n. It will thus suffice to show the logarithmic decay bound

$$\|\nu - 1\|_{U^{2k}[\frac{N-b}{W}]}^{2^{2k}} \ll \log^{-c\varepsilon} N.$$

Expanding out the left-hand side, it suffices to show that

$$\mathbb{E}_{(n,\vec{h})\in\Omega} \prod_{\omega\in S} \frac{\phi(W)}{W} \Lambda_{\chi,R,2}(W(n+\omega\cdot\vec{h})+b) = \mathrm{vol}(\Omega) + O((N/W)^{2k+1}\log^{-c\varepsilon} N) \tag{8.14}$$

for all subsets S of $\{0,1\}^{2k}$, where $\Omega \subset \mathbb{R}^{2k+1}$ is the convex body

$$\Omega := \big\{(x,\vec{y}) \in \mathbb{R}^{2k+1} : 0 < W(x+\omega\cdot\vec{y})+b \le N \ \ \forall \omega \in \{0,1\}^{2k}\big\}.$$

Suppose that we directly apply the estimate[11] in [22, Theorem D.3], using (8.12) to eliminate the role of the sieve factors. Then we can express the left-hand side of (8.14) as

$$\left(\frac{\phi(W)}{W}\right)^{\#S}\left(\mathrm{vol}(\Omega)\prod_p \beta_p + O\left(\frac{(N/W)^{2k+1}}{\log^{1/20} R} e^{O(X)}\right)\right), \tag{8.15}$$

where β_p are the usual local factors

$$\beta_p := \mathbb{E}_{(n,\vec{h})\in(\mathbb{Z}/p\mathbb{Z})^{2k+1}} \prod_{\omega\in S} \frac{p}{p-1} 1_{W(n+\omega\cdot\vec{h})+b\neq 0},$$

X is the quantity

$$X := \sum_{p\in P} p^{-1/2}$$

and P is the set of primes p which are "exceptional" in the sense that at least two of the affine forms

$$(x,\vec{y}) \mapsto W(x+\omega\cdot\vec{y})+b \tag{8.16}$$

for $\vec{\omega} \in \{0,1\}^{2k}$ are linearly dependent modulo p.

Since $W = P(\log^{\varepsilon} N)$, one has $\beta_p = (\frac{p}{p-1})^{\#S}$ for $p < \log^{\varepsilon} N$, while from the inclusion-exclusion calculation used in the proof of Proposition 5.2 one has $\beta_p = 1 + O(1/p^2)$ for $p \ge \log^{\varepsilon} N$. Thus

$$\prod_p \beta_p = \left(\frac{W}{\phi(W)}\right)^{\#S}(1 + O(\log^{-\varepsilon} N)). \tag{8.17}$$

[11] This theorem as stated requires γ to be sufficiently small depending on W (represented in [22] by the parameter L), but the bound $R \le N^{\gamma}$ is only used before [22, (D.4)] to show that an expression of the form $O(L^{O(1)}R^{O(1)}N^{d-1}\log^{t} R)$ (here we have made the dependence on L explicit) is equal to $o(N^d)$, and this can be achieved with $R \le N^{\gamma}$ and γ independent of L, so long as we also have $L \le N^{\gamma}$, which is also the case here since $L = O(W)$ and N is assumed to be sufficiently large.

Since $\mathrm{vol}(\Omega) \ll (N/W)^{2k+1}$, the main term in (8.15) is acceptable. If it were not for the $e^{O(X)}$ term, the error term in (8.15) would similarly be acceptable; unfortunately, as defined in [22, Appendix D], the exceptional primes consist precisely of all the primes p up to $\log^{\varepsilon} X$, and this would ostensibly lead to an unacceptably large error term in (8.15). But, an inspection of the proof of [22, Proposition D.4] reveals that the $e^{O(X)}$ loss arises from three sources. One is from the crude bound

$$\prod_p \beta_p \leq e^{O(X)} \tag{8.18}$$

(see [22, (D.14)]); one is from the variant

$$\prod_{p > \log^{1/10} R} \beta_p \leq 1 + O(e^{O(X)} \log^{-1/20} R) \tag{8.19}$$

(see [22, equation after (D.15)]); and the third arises from the estimate

$$\sum_{p \in P_\Psi :\, p > \log^{1/10} R} p^{-1} = O(X \log^{-1/20} R) \tag{8.20}$$

appearing in the fourth display after [22, (D.16)]. Of course, for the first estimate (8.18) we may use the superior bound (8.17) instead in our case. In our cases none of the exceptional primes exceed $\log^{\varepsilon} N < \log^{1/10} R$, and so one can replace X with 0 in (8.19), (8.20). As a consequence of these observations, the $e^{O(X)}$ factor in [22, Proposition D.4] may be replaced with $(\frac{W}{\phi(W)})^{\#S}$, and the error term in (8.15) is now also acceptable, giving the claim. ∎

Proof of Theorem 2.6 for Λ. Combining Proposition 8.4 with (the contrapositive of) Theorem 8.3, we see that it suffices to show (for a sufficiently small constant $c_1 > 0$) that one has the pseudopolynomial bound

$$\mathbb{E}_{n \in [\frac{N-b}{W}]} \frac{\phi(W)}{W} (\Lambda - \Lambda_{\mathrm{Siegel}})(Wn + b) \overline{F}(g(n)\Gamma) \ll \exp(-c \log^c N) \tag{8.21}$$

whenever G/Γ is a (filtered) nilmanifold G/Γ of degree $k-1$, dimension at most $(\log\log N)^{c_1}$ and complexity at most $\exp(\log^{c_1} N)$, $F : G/\Gamma \to \mathbb{C}$ is a 1-bounded Lipschitz function of Lipschitz constant at most $\exp(\log^{c_1} N)$, and $g : \mathbb{Z} \to G$ is a polynomial map. Using [34, Lemma 4.2], we can write $g(n) = \tilde{g}(Wn + b)$ for another polynomial map $\tilde{g} : \mathbb{Z} \to G$. But from Theorem 2.7 we have

$$\sum_{n \leq N :\, n \equiv b\ (W)} (\Lambda - \Lambda_{\mathrm{Siegel}})(n) \overline{F}(\tilde{g}(n)\Gamma) \ll N \exp(-c \log^c N)$$

for some $c > 0$ independent of ε, and the claim (8.21) then follows for ε small enough. This (finally!) completes the proof of Theorem 2.6, and hence that of Theorem 1.4. ∎

We can now quickly deduce Corollary 1.5 from our main theorem.

Proof of Corollary 1.5. Let $w = (\log\log N)^{1/2}$. By Theorem 2.5, we have

$$\|\Lambda_{\text{Siegel}} - \Lambda_{\text{Cramér},Q}\|_{U^k[N]} \ll \log^{-c} N. \tag{8.22}$$

Using the Fourier expansion $1_{n\equiv b\ (W)} = \frac{1}{W}\sum_{1\le a\le W} e(\frac{a(n-b)}{W})$, the triangle inequality for the Gowers norms, and the fact that $\|fe(\xi\cdot)\|_{U^k[N]} = \|f\|_{U^k[N]}$ for any function f and any $\xi\in\mathbb{R}$, we deduce from (8.22) that

$$\begin{aligned}\left\|\frac{\phi(W)}{W}\big(\Lambda_{\text{Siegel}}(W\cdot+b) - \Lambda_{\text{Cramér},Q}(W\cdot+b)\big)\right\|_{U^k[\frac{N-b}{W}]}&\\ \ll W^{(k+1)/2^k}\log^{-c} N \ll \log^{-c} N.&\end{aligned} \tag{8.23}$$

From Proposition 5.3, we have

$$\left\|\frac{\phi(W)}{W}\Lambda_{\text{Cramér},Q}(W\cdot+b) - 1\right\|_{U^k[\frac{N-b}{W}]} \ll w^{-c}. \tag{8.24}$$

Let $w' = \log^{\varepsilon} N$ where ε is as in Section 8.2. Also let $W' = \prod_{p\le w'} p$. Then by Corollary 4.2 and (8.11) we have

$$\begin{aligned}&\max_{(b,W)=1}\left\|\frac{\phi(W)}{W}(\Lambda - \Lambda_{\text{Siegel}})(W\cdot+b)\right\|_{U^k[\frac{N-b}{W}]}\\ &\quad\ll \max_{(b',W')=1}\left\|\frac{\phi(W')}{W'}(\Lambda - \Lambda_{\text{Siegel}})(W'\cdot+b')\right\|_{U^k[\frac{N-b'}{W'}]} \ll (\log\log N)^{-c}.\end{aligned}$$

Now the claim follows by combining this with (8.23), (8.24) and applying the triangle inequality for Gowers norms. ∎

9. Quantitative linear equations in primes result

In this section we sketch the derivation of Theorem 1.6 from Theorem 1.4. The arguments follow those in [22] extremely closely, and we will assume familiarity with those arguments in this section.

In [22, Section 4], the qualitative version of Theorem 1.6 was derived from [22, Theorem 4.5] using some elementary linear algebra and convex geometry. The same arguments, replacing all qualitative decay terms with doubly logarithmic ones instead, show that Theorem 1.6 will follow if one shows the following.

Theorem 9.1 (Primes in affine lattices in normal form). *The statement of* [22, Theorem 4.5] *continues to hold if the qualitative error term $o(N^d)$ in that theorem is replaced with the doubly logarithmic term $O_{s,d,t}((\log\log N)^{-c}N^d)$ for some $c = c_{s,d,t} > 0$ depending only on the parameters s, d, t.* (*Also one ignores the references to the now proven conjectures* GI(s), MN(s) *in that theorem.*)

Next, we apply the W-trick arguments in [22, Section 5], setting w equal[12] to $(\log\log N)^\eta$ for a sufficiently small $\eta > 0$ depending on s, d, t rather than the more conservative choice of $\log\log\log N$. These arguments then reduce matters to showing

Theorem 9.2 (W-tricked primes in affine lattices). *The statement of* [22, Theorem 5.2] *continues to hold if the qualitative error term $o(N^d)$ in that theorem is replaced with the doubly logarithmic term $O_{s,d,t}((\log\log N)^{-c} N^d)$ for some $c = c_{s,d,t} > 0$ depending only on the parameters s, d, t. (Again one ignores the references to the now proven conjectures* $\mathrm{GI}(s), \mathrm{MN}(s)$ *in that theorem.*)

The statement of [22, Theorem 5.2] involves the functions

$$\Lambda'_{b_i,W}(n) := \frac{\phi(W)}{W}\Lambda'(Wn + b_i),$$

where $W := P(w)$ and Λ' is the restriction of Λ to the primes. From Corollary 1.5, we have the doubly logarithmic bound

$$\|\Lambda'_{b_i,W} - 1\|_{U^{s+1}[\frac{N-b_i}{W}]} \ll_{s,\eta} (\log\log N)^{-c\eta}$$

for some $c > 0$ depending only on s (and assuming as we may that N is sufficiently large depending on s, d, t, η). On the other hand, a routine modification of Proposition 8.4 (see also [22, Proposition 6.4]) reveals that for any D, the function $1 + \Lambda'_{b_1,W} + \cdots + \Lambda'_{b_t,W}$ on the interval $[N^{3/5}, N]$ can be bounded by $C\nu$ for some $C = O_{D,\eta}(1)$ and some ν that obeys the (D, D, D) linear forms condition from [22, Definition 6.2] with the $o_D(1)$ term in [22, (6.2)] replaced by $O_D((\log\log N)^{-c_{D,\eta}})$ for some $c_{D,\eta} > 0$. (We will not need the now largely obsolete "correlation condition" in [22, Definition 6.3].) The claim now follows from the generalized von Neumann theorem in [22, Theorem 7.1] proven in [22, Appendix C], after replacing all $o(1)$ type terms with $O((\log\log N)^{-c})$ type terms, noting that all the functions denoted κ in that appendix can be taken to be polynomial in nature; we leave the details to the interested reader.

Remark 9.3. It seems likely that one can improve Theorem 1.6 further, by allowing the parameter L to be as large as $(\log\log N)^c$ with uniform control on error terms; one may even be able to handle significantly larger values of the linear coefficients $\dot{\psi}_i$ than this by incorporating the various methods used in this paper. We will not pursue such refinements here, however.

10. Arithmetic progressions with shifted prime difference

In this section we prove Theorem 1.8.

[12] Note that for this choice of w, the prime number theorem in arithmetic progressions of modulus $W = P(w)$ has an effective error term with good decay, as we can use the effective lower bounds on $L(1, \chi)$ in this case rather than Siegel's theorem. It should however be possible to work with larger choices of w by incorporating the contribution of a Q-Siegel zero, as is done elsewhere in this paper.

Proof of Theorem 1.8. In what follows, let Λ' stand for the von Mangoldt function restricted to the primes. Let $A \subset [N]$ be any set with $|A| \geq \delta N$ and $\delta = (\log\log\log\log N)^{-c}$ for small enough $c > 0$ depending on k. Let $w = (\log\log N)^{1/2}$, and let $W = \prod_{p \leq w} p$. By the pigeonhole principle, we can pick $1 \leq b \leq W$ such that $A' := \{n : Wn + b \in A\}$ has size $\geq \delta N/W$. Then the count of k-term arithmetic progressions in A with shifted prime difference is

$$\geq \frac{1}{\log N} \sum_{n \leq N/W} \sum_{d \leq N/W} 1_A(Wn+b)1_A(Wn+b+Wd)\cdots 1_A(Wn+b+Wd(k-1)) \times \Lambda'(Wd+1)$$

$$= \frac{1}{\log N} \sum_{n \leq N/W} \sum_{d \leq N/W} 1_{A'}(n)1_{A'}(n+d)\cdots 1_{A'}(n+d(k-1))\Lambda'(Wd+1) =: T.$$

Note that we have the trivial bound $\sum_{n \leq N} |\Lambda(n) - \Lambda'(n)| \ll N^{1/2}\log N$. Using this and our quantitative Gowers uniformity result in the form of Corollary 1.5, we have

$$\left\| \frac{\phi(W)}{W} \cdot \Lambda'(W\cdot + 1) - 1 \right\|_{U^k[N/W]} = \left\| \frac{\phi(W)}{W} \Lambda(W\cdot + 1) - 1 \right\|_{U^k[N/W]} + O(N^{-1/2+o(1)}) \ll (\log\log N)^{-c'}$$

for some $c' > 0$ depending on k. Therefore, by applying the generalized von Neumann theorem for pseudorandomly majorized functions [22, Theorem 7.1] (with similar remarks on quantitative error terms to those in the proof of Theorem 1.6), we see that T is equal to

$$\frac{W}{\varphi(W)(\log N)} \sum_{n \leq N/W} \sum_{0 \leq d \leq N/W} 1_{A'}(n)1_{A'}(n+d)\cdots 1_{A'}(n+d(k-1)) + O\big((N/W)^2(\log\log N)^{-c'}\big). \quad (10.1)$$

For $\rho > 0$, let $N_k(\rho)$ denote the smallest positive integer such that, for any $m \geq N_k(\rho)$, any subset of $[m]$ of size $\geq \rho m$ contains a non-trivial k-term arithmetic progression. Let $c(k,\delta) := \delta^2/(16N_k(\delta/2)^3)$. Then, by a well known argument of Varnavides for quantifying Szemerédi's theorem (see e.g. [41, Theorem 18, Remark 1]), for $N/W > 2N_k(\delta/2)$ the expression (10.1) is

$$\geq \frac{W}{\varphi(W)} c(k,\delta)(N/W)^2 + O\big((N/W)^2(\log\log N)^{-c'}\big).$$

We have $c(k,\delta) \gg \exp(-\exp(\delta^{-C}))$ for some $C \geq 1$ (depending on k) by Gowers's bound $N_k(\rho) \ll \exp(\exp(\rho^{-C'}))$, proved in [13]. Now, if c is chosen small enough in the definition of δ, we have $c(k,\delta) \gg (\log\log N)^{-o(1)}$, which proves the statement of the theorem for $k \geq 4$. For $k = 4$, the same argument works, except that we now use the bound $N_4(\rho) \ll \exp(\rho^{-C})$ from [25] to get $c(4,\delta) \gg \exp(-C\delta^{-C})$, which enables

taking $\delta = (\log\log\log N)^{-c}$ for some $c > 0$. Finally, for $k = 3$, using the very recent bound (see [30]) $N_3(\rho) \ll \exp((\log(1/\rho))^C)$ we have $c(3,\delta) \gg \exp(-C(\log(1/\delta))^C)$, which enables taking $\delta = \exp(-(\log\log\log N)^c)$ for some $c > 0$. ∎

Appendix A. Quantitative Leibman theory with explicit dimension dependence

In this appendix we refine the equidistribution theory on nilmanifolds from [24], tracking more carefully the dependence on dimension m (but allowing all constants to depend on the degree d, which in our context will be equal to $k-1$). The key point is that all bounds will be at most doubly exponential in this dimension parameter, basically because the arguments rely on applying the Cauchy–Schwarz inequality (or variants such as the van der Corput inequality) a number of times that is polynomial in the dimension. (Many of the estimates here require only single exponential dependence on m at worst, but the induction on dimension we use only closes if we allow double exponential dependence.) In order to improve this double exponential dependence it would seem necessary to adopt a different approach to equidistribution that is not as reliant on so many applications of the Cauchy–Schwarz inequality.

We freely use the notation from [24], and let m be a dimensional parameter. To conveniently track bounds that depend in double exponential fashion on the dimension we adopt the following notation. For any $0 < \delta < 1/2$ let $\mathrm{poly}_m(\delta)$ be any quantity $\gg \exp(-\exp(m^{O(1)}))\delta^{\exp(m^{O(1)})}$, and for any $Q > 2$ let $\mathrm{poly}_m(Q)$ be any quantity $\ll \exp(\exp(m^{O(1)}))Q^{\exp(m^{O(1)})}$. In particular, $\mathrm{poly}_m(1/\delta)$ is any quantity $\ll \exp(\exp(m^{O(1)}))\delta^{-\exp(m^{O(1)})}$.

We begin with a more quantitative version of [24, Lemma 3.1]:

Lemma A.1 (Quantitative Kronecker Theorem). *Let $m \geq 1$ and $0 < \delta < 1/2$, $\alpha \in \mathbb{R}^m$, $N \geq 1$. If $(\alpha n \bmod \mathbb{Z}^m)_{n\in[N]}$ is not δ-equidistributed in $\mathbb{R}^m/\mathbb{Z}^m$, then there exists $k \in \mathbb{Z}^m$ with $0 < |k| \ll \mathrm{poly}_m(1/\delta)$ such that $\|k\cdot\alpha\|_{\mathbb{R}/\mathbb{Z}} \ll \mathrm{poly}_m(1/\delta)/N$.*

Proof. The "simple calculation" used to establish [24, (3.3)], when done a little more carefully, gives

$$\sum_{\substack{k\in\mathbb{Z}^m\\ |k|\geq M}} |\hat{K}(k)| \ll \mathrm{poly}_m(1/\delta)M^{-1} \tag{A.1}$$

and by chasing through the argument with this bound we obtain the claim. ∎

This gives a version of [24, Lemma 3.7]:

Lemma A.2 (Vertical oscillation reduction). *Let G/Γ be a filtered nilmanifold of degree d, with vertical torus dimension m_d. Let $0 < \delta < 1/2$, and let $g\colon \mathbb{Z} \to G$ be a polynomial sequence for which $(g(n)\Gamma)_{n\in[N]}$ is not δ-equidistributed. Then there is a vertical character ξ with $|\xi| \leq \mathrm{poly}_{m_d}(1/\delta)$ such that $(g(n)\Gamma)_{n\in[N]}$ is not $\mathrm{poly}_{m_d}(\delta)$-equidistributed along the vertical oscillation ξ.*

Proof. Repeat the proof of [24, Lemma 3.7] verbatim, using the estimate (A.1) in place of [24, (3.3)]. ∎

Now we state the main technical theorem on quantitative Leibman theory (a version of [24, Theorem 7.1]):

Theorem A.3 (Variant of Main Theorem). *Let $m \geq m_* \geq 0$ be integers, $0 < \delta < 1/2$, $N \geq 1$. Let G/Γ be a filtered nilmanifold of degree d, nonlinearity dimension m_* (defined in* [24, *Section* 7]*), and complexity at most $1/\delta$. Let $g: \mathbb{Z} \to G$ be a polynomial sequence. If $(g(n)\Gamma)_{n\in[N]}$ is not δ-equidistributed then there exists a horizontal character η with $0 < |\eta| \leq \delta^{-\exp((m+m_*)^{C_d})}$ such that*

$$\|\eta \circ g\|_{C^\infty[N]} \leq \delta^{-\exp((m+m_*)^{C_d})}$$

where C_d is a sufficiently large constant depending only on d.

We now prove this theorem. We assume inductively that the claim has already been established for smaller values of d, or for the same value of d and smaller values of m_*. Henceforth we refine the poly_m notation by permitting the implied constants to depend on the constant C_{d-1}, but not on C_d.

By repeating the derivation of [24, (7.1)] (using Lemma A.2 in place of [24, Lemma 3.7]) we may find some function $F: G/\Gamma \to \mathbb{C}$ with $\|F\|_{\mathrm{Lip}} \ll \mathrm{poly}_m(1/\delta)$ and vertical frequency ξ with $|\xi| \ll \mathrm{poly}_m(1/\delta)$ such that $(g(n)\Gamma)_{n\in[N]}$ is not $\delta^{O(1)}$-equidistributed along ξ, and such that

$$\left|\mathbb{E}_{n\in[N]} F(g(n)\Gamma) - \int_{G/\Gamma} F\right| \gg \mathrm{poly}_m(\delta).$$

If $\xi = 0$ then a repetition of the arguments after [24, (7.1)] gives the claim from the induction hypothesis, so without loss of generality we assume $\xi \neq 0$, thus we now have

$$|\mathbb{E}_{n\in[N]} F(g(n)\Gamma)| \gg \mathrm{poly}_m(\delta).$$

Repeating the reductions after [24, (7.2)], we may assume that $g(0) = \mathrm{id}_G$ and $|\psi(g(1))| \leq 1$, where $\psi: G \to \mathbb{R}^m$ is the Mal'cev coordinate map. Continuing the argument down to [24, (7.8)] we conclude that

$$|\mathbb{E}_{n\in[N]} \overline{F_h^\square}(\overline{g_h^\square}(n)\,\overline{\Gamma^\square})| \gg \mathrm{poly}_m(\delta)$$

with $\overline{F_h^\square}, \overline{g_h^\square}, \overline{\Gamma^\square}$ defined as in [24].

One can rather tediously verify that all the estimates in [24, Appendix A] can be refined by replacing all estimates of the form $X \ll_m Q^{O_m(1)}Y$ with $X \ll \mathrm{poly}_m(Q)Y$. As a consequence, we can refine [24, Lemma 7.4] (by exact repetition of the proof) to

Lemma A.4 (Rationality bounds for the relative square). *There is a $\mathrm{poly}_m(1/\delta)$-rational Mal'cev basis $\mathcal{X}^\square$ for $G^\square/\Gamma^\square$ adapted to the filtration $(G^\square)_\bullet$ with the property that $\psi_{X^\square}(x, x')$ is a polynomial of degree $O(1)$ with rational coefficients of height $\mathrm{poly}_m(1/\delta)$*

in the coordinates $\psi(x), \psi(x')$. *With respect to the metric* $d_{\mathcal{X}^\square}$ *we have* $\|F_h^\square\|_{\mathrm{Lip}} \ll \mathrm{poly}_m(1/\delta)$ *uniformly in* h.

Continuing the arguments down to [24, Lemma 7.5], one can find horizontal characters $\eta_1 \colon G \to \mathbb{R}/\mathbb{Z}$, $\eta_2 \colon G_2 \to \mathbb{R}/\mathbb{Z}$ with η_2 annihilating $[G, G_2]$ and $|\eta_1|, |\eta_2| \ll \mathrm{poly}_m(1/\delta)$ such that the character $\eta \colon G^\square \to \mathbb{R}/\mathbb{Z}$ defined by

$$\eta(g', g) := \eta_1(g) + \eta_2(g'g^{-1})$$

is such that

$$\|\eta \circ g_h^\square\|_{C^\infty[N]} \ll \mathrm{poly}_m(1/\delta)$$

for $\gg \mathrm{poly}_m(\delta)N$ values of $h \in [N]$.

Continuing the argument down to [24, (7.16)], and using the induction hypothesis for Theorem A.3 (with d replaced by $d-1$, and m, m_* replaced by quantities not exceeding $2m$), we can find $1 \le q \ll \mathrm{poly}_m(1/\delta)$ such that

$$\|\eta_1(g(1)) + \zeta \cdot \{\gamma h\} + q\alpha h\|_{\mathbb{R}/\mathbb{Z}} \ll \mathrm{poly}_m(1/\delta)/N$$

for $\gg \mathrm{poly}_m(\delta)N$ values of $h \in [N]$, where

(i) $\alpha \in \mathbb{R}/\mathbb{Z}$ is the quantity $\alpha := \partial^2(\eta_2 \circ g_2)(0)$, where $g_2(n) = g(n)g(1)^{-n}$ is the nonlinear part of g;

(ii) $\gamma \in (\mathbb{R}/\mathbb{Z})^{m_{\mathrm{lin}}}$ is (the first m_{lin} components of) $\psi(g(1))$;

(iii) $\zeta \in \mathbb{R}^{m_{\mathrm{lin}}}$ is the vector such that

$$\eta_2([g(1), x]) = \zeta \cdot \psi(x) \bmod \mathbb{Z}$$

for all $x \in G$ (extending ζ by zero to $\mathbb{R}^m$).

Here it is important that the implied constants in the poly_m notation are allowed to depend on C_{d-1} (but not on C_d).

It is routine to verify that $|\zeta| \ll \mathrm{poly}_m(1/\delta)$. An inspection of the proof of [24, Proposition 5.3] and [24, Claim 7.7], using Lemma A.1 in place of [24, Lemma 3.1], shows that we may replace all bounds of the form $X \ll_m \delta^{-O_m(1)}Y$ appearing in these statements by $X \ll \mathrm{poly}_m(1/\delta)Y$, to obtain one of the following claims:

(I) there is $r \ll \mathrm{poly}_m(1/\delta)$ such that $\|r\zeta_i \bmod \mathbb{Z}\|_{\mathbb{R}/\mathbb{Z}} \ll \mathrm{poly}_m(1/\delta)/N$ for all $i = 1, \dots, m_{\mathrm{lin}}$; or

(II) there exists $k \in \mathbb{Z}^{m_{\mathrm{lin}}}$ with $0 < |k| \ll \mathrm{poly}_m(1/\delta)$ such that

$$\|k \cdot \gamma\|_{\mathbb{R}/\mathbb{Z}} \ll \mathrm{poly}_m(1/\delta)/N.$$

In case (II) we conclude exactly as in [24], so suppose we are in case (I). Arguing as in [24] we can easily close the induction except in the case when η_2 (and hence η) annihilates $[G, G]$, at which point the arguments in [24] lead to

$$\|\eta_2 \circ g_2\|_{\mathbb{R}/\mathbb{Z}} \ll \mathrm{poly}_m(1/\delta)$$

(possibly after first multiplying η_1, η_2 by a positive integer of size $\mathrm{poly}_m(1/\delta)$.

Repeating the rest of the proof of [24, Theorem 7.1] (replacing all bounds of the form $X \ll_m \delta^{-O_m(1)} Y$ with $X \ll \mathrm{poly}_m(1/\delta)Y$) and using the induction hypothesis with (d, m_*) replaced by $(d, m_* - 1)$, we see that

$$\|\eta \circ g\|_{C^\infty[N]} \leq \mathrm{poly}_m(1/\delta)^{\exp((m+m_*-1)^{C_d})}$$

for some horizontal character $\eta : G \to \mathbb{R}/\mathbb{Z}$ with $|\eta| \leq \mathrm{poly}_m(1/\delta)^{\exp((m+m_*-1)^{C_d})}$. For C_d large enough, we have

$$\mathrm{poly}_m(1/\delta)^{\exp((m+m_*-1)^{C_d})} \leq \delta^{-\exp((m+m_*)^{C_d})},$$

and Theorem A.3 follows.

Repeating the proof of [24, Proposition 9.2] (specializing to the single-parameter case $t = 1$), we then obtain

Proposition A.5 (Factorization of poorly-distributed polynomial sequences). *Let $m \geq 1$, $0 < \delta < 1/2$, $N \geq 1$, $d \geq 0$, let G/Γ be an m-dimensional filtered nilmanifold of complexity at most $1/\delta$, and let $g : \mathbb{Z} \to G$ be a polynomial sequence. Suppose that $(g(n)\Gamma)_{n\in[N]}$ is not totally δ-equidistributed. Then there is a factorization $g = \varepsilon g' \gamma$ with $\varepsilon, g', \gamma : \mathbb{Z} \to G$ polynomials such that*

(i) *$\varepsilon : \mathbb{Z} \to G$ is $(\mathrm{poly}_m(1/\delta), N)$-smooth;*

(ii) *$g' : \mathbb{Z} \to G$ takes values in a connected proper $\mathrm{poly}_m(1/\delta)$-rational subgroup G' of G;*

(iii) *$\gamma : \mathbb{Z} \to G$ is $\mathrm{poly}_m(1/\delta)$-rational.*

In [24, Lemma 10.1] with $t = 1$, one easily verifies that the bound $M^{O_m(1)}$ in the conclusion can be sharpened to $\mathrm{poly}_m(M)$. We now claim the following quantitative version of [24, Theorem 1.19]:

Theorem A.6 (Factorization theorem). *Let $m \geq 0$, $M_0 \geq 2$, $A \geq 2$, $N \geq 1$, $d \geq 0$. Let G/Γ be an m-dimensional filtered nilmanifold of degree d and complexity at most M_0, and let $g : \mathbb{Z} \to G$ be a polynomial sequence. Then there is some M with $M_0 \leq M \leq M_0^{A^{(2+m)^{O_d(1)}}}$, a subgroup $G' \subset G$ which is M-rational with respect to $\mathcal{X}$, and a decomposition $g = \varepsilon g' \gamma$ with $\varepsilon, g', \gamma : \mathbb{Z} \to G$ polynomials such that*

(i) *ε is (M, N)-smooth;*

(ii) *g' takes values in G' and $(g'(n)\Gamma)_{n\in[N]}$ is totally $1/M^A$-equidistributed in G'/Γ', with respect to a Mal'cev basis $\mathcal{X}'$ consisting of M-rational linear combinations of the basis elements of the Mal'cev basis for G;*

(iii) *γ is M-rational and $\gamma(n)\Gamma$ is periodic with period at most M.*

Proof. Repeat the proof of [24, Theorem 10.2] with $t = 1$, setting $\delta_{i+1} := \delta_i^{A^{mC}}$ for a sufficiently large constant $C = C_d$ depending only on d (in particular, $1/\delta_{i+1}$ is much larger than any quantity of the form $\mathrm{poly}_m(1/\delta_i)^A$ if C is large enough). ∎

Appendix B. Proof of Theorem 1.3

Proof of Theorem 1.3. Part (i) follows easily from the prime number theorem with Vinogradov–Korobov error terms (for the Möbius case, see [44, Satz 3 in Section V.5]). Part (ii) for the Möbius function follows from the strongly logarithmic exponential sum estimates

$$\sup_\theta |\mathbb{E}_{n\in[N]}\mu(n)e(\theta n)| \ll_A^{\text{ineff}} \log^{-A} N$$

of Davenport [7], the Plancherel estimate

$$\int_0^1 |\mathbb{E}_{n\in[N]}\mu(n)e(\theta n)|^2\, d\theta \ll N, \tag{B.1}$$

the circle method, and Cauchy–Schwarz. For the second part of (ii), observe from Proposition 1.2 and (1.2) that we may take $w = \log^{1/100} N$ (say) without loss of generality. The standard Vinogradov estimates for exponential sums over primes (see e.g. [29, Ch. 13]) eventually reveal the logarithmic bounds

$$\sup_\theta |\mathbb{E}_{n\in[N]}(\Lambda(n) - \Lambda_{\text{Cramér},w}(n))e(\theta n)| \ll^{\text{ineff}} \log^{-c} N,$$

while the Fourier restriction estimate from [17, Proposition 4.2] gives

$$\int_0^1 |\mathbb{E}_{n\in[N]}(\Lambda(n) - \Lambda_{\text{Cramér},w}(n))e(\theta n)|^q\, d\theta \ll_q 1$$

for any $2 < q < \infty$, and the claim now follows from the circle method and Hölder's inequality. Finally, for (iii), we see from Proposition 1.2 and (1.2) that we may assume that w grows sufficiently slowly in N, and then the bounds in (iii) follow easily from the main theorems in [22] as well as Corollary 4.2, after inserting the resolution of the inverse conjecture for the Gowers norms (first proven in [27]) and the strong orthogonality of the Möbius function to nilsequences (first proven in [23]). ∎

Remark B.1. An alternative approach to (1.4) proceeds by comparing $\Lambda(n) = -\sum_{d|n}\mu(d)\log d$ first with a truncated divisor sum $\Lambda^\sharp(n) := -\sum_{d|n:\, d\le N^{c_1}}\mu(d)\log d$ for some small absolute constant $c_1 > 0$, and establishing the strongly logarithmic estimate

$$\|\Lambda - \Lambda^\sharp\|_{U^2[N]} \ll_A^{\text{ineff}} \log^{-A} N$$

from the circle method (here we can use a Plancherel bound analogous to (B.1) that loses a factor of $\log N$, thus avoiding the need to invoke the restriction theory from [17]), and the logarithmic estimate

$$\|\Lambda^\sharp - \Lambda_{\text{Cramér},w}\|_{U^2[N]} \ll \log^{-c} N$$

from sieve theory with (say) $w = \log^{1/100} N$, and then applying the triangle inequality (1.2); we leave the details to the interested reader. In this paper we found the Cramér models $\Lambda_{\text{Cramér},w}$ to be slightly more convenient technically to work with than the truncated divisor sum model $\Lambda^\sharp$, and therefore made no further use of $\Lambda^\sharp$ here.

Acknowledgments. We thank the anonymous referee for a careful reading of the paper and for numerous helpful corrections. We thank Sean Prendiville for helpful discussions, and Andrew Granville, James Leng and Wataru Kai for corrections.

Funding. TT was supported by a Simons Investigator grant, the James and Carol Collins Chair, the Mathematical Analysis & Application Research Fund Endowment, and by NSF grant DMS-1764034. JT was supported by a Titchmarsh Fellowship and funding from the European Union's Horizon Europe research and innovation programme under Marie Skłodowska-Curie grant agreement no. 101058904.

References

[1] Bloom, T. F.: Quantitative inverse theory of Gowers uniformity norms (after F. Manners). In: Séminaire Bourbaki 2019/2021, exp. 1174, Astérisque **430**, 237–273 (2021) Zbl 07554181 MR 4439193

[2] Chinis, J.: Siegel zeros and Sarnak's conjecture. arXiv:2105.14653 (2021)

[3] Cladek, L., Tao, T.: Additive energy of regular measures in one and higher dimensions, and the fractal uncertainty principle. Ars Inven. Anal. **2021**, art. 1, 38 pp. (2021) Zbl 1481.28008 MR 4462469

[4] Conlon, D., Fox, J., Zhao, Y.: The Green–Tao theorem: an exposition. EMS Surv. Math. Sci. **1**, 249–282 (2014) Zbl 1311.11092 MR 3285854

[5] Conlon, D., Fox, J., Zhao, Y.: A relative Szemerédi theorem. Geom. Funct. Anal. **25**, 733–762 (2015) Zbl 1345.11008 MR 3361771

[6] Cramér, H.: On the order of magnitude of the difference between consecutive prime numbers. Acta Arith. **2**, 23–46 (1936) Zbl 62.1147.01

[7] Davenport, H.: On some infinite series involving arithmetical functions. II. Quart. J. Math. Oxford **8**, 313–320 (1937) Zbl 0017.39101

[8] Davenport, H.: Multiplicative number theory. 3rd ed., Graduate Texts in Mathematics 74, Springer, New York (2000) Zbl 1002.11001 MR 1790423

[9] Dodos, P., Kanellopoulos, V.: Uniformity norms, their weaker versions, and applications. Acta Arith. **203**, 251–270 (2022) Zbl 07547459 MR 4436255

[10] Frantzikinakis, N., Host, B., Kra, B.: Multiple recurrence and convergence for sequences related to the prime numbers. J. Reine Angew. Math. **611**, 131–144 (2007) Zbl 1126.37005 MR 2361086

[11] Friedlander, J., Iwaniec, H.: Opera de cribro. American Mathematical Society Colloquium Publications 57, American Mathematical Society, Providence, RI (2010) Zbl 1226.11099 MR 2647984

[12] Germán, L., Kátai, I.: On multiplicative functions on consecutive integers. Lith. Math. J. **50**, 43–53 (2010) Zbl 1248.11070 MR 2607678

[13] Gowers, W. T.: A new proof of Szemerédi's theorem. Geom. Funct. Anal. **11**, 465–588 (2001) Zbl 1028.11005 MR 1844079

[14] Gowers, W. T.: Decompositions, approximate structure, transference, and the Hahn–Banach theorem. Bull. London Math. Soc. **42**, 573–606 (2010) Zbl 1233.05198 MR 2669681

[15] Granville, A.: Harald Cramér and the distribution of prime numbers. Scand. Actuarial J. **1995**, no. 1, 12–28 Zbl 0833.01018 MR 1349149

[16] Green, B.: On Sárközy's theorem for shifted primes. arXiv:2206.08001 (2022), to appear in J. Amer. Math. Soc.

[17] Green, B., Tao, T.: Restriction theory of the Selberg sieve, with applications. J. Théor. Nombres Bordeaux **18**, 147–182 (2006) Zbl 1135.11049 MR 2245880

[18] Green, B., Tao, T.: An inverse theorem for the Gowers $U^3(G)$ norm. Proc. Edinburgh Math. Soc. (2) **51**, 73–153 (2008) Zbl 1202.11013 MR 2391635

[19] Green, B., Tao, T.: The primes contain arbitrarily long arithmetic progressions. Ann. of Math. (2) **167**, 481–547 (2008) Zbl 1191.11025 MR 2415379

[20] Green, B., Tao, T.: Quadratic uniformity of the Möbius function. Ann. Inst. Fourier (Grenoble) **58**, 1863–1935 (2008) Zbl 1160.11017 MR 2473624

[21] Green, B., Tao, T.: An equivalence between inverse sumset theorems and inverse conjectures for the U^3 norm. Math. Proc. Cambridge Philos. Soc. **149**, 1–19 (2010) Zbl 1229.11132 MR 2651575

[22] Green, B., Tao, T.: Linear equations in primes. Ann. of Math. (2) **171**, 1753–1850 (2010) Zbl 1242.11071 MR 2680398

[23] Green, B., Tao, T.: The Möbius function is strongly orthogonal to nilsequences. Ann. of Math. (2) **175**, 541–566 (2012) Zbl 1347.37019 MR 2877066

[24] Green, B., Tao, T.: The quantitative behaviour of polynomial orbits on nilmanifolds. Ann. of Math. (2) **175**, 465–540 (2012) Zbl 1251.37012 MR 2877065

[25] Green, B., Tao, T.: New bounds for Szemerédi's theorem, III: a polylogarithmic bound for $r_4(N)$. Mathematika **63**, 944–1040 (2017) Zbl 1434.11037 MR 3731312

[26] Green, B., Tao, T., Ziegler, T.: An inverse theorem for the Gowers U^4-norm. Glasgow Math. J. **53**, 1–50 (2011) Zbl 1262.11013 MR 2747135

[27] Green, B., Tao, T., Ziegler, T.: An inverse theorem for the Gowers $U^{s+1}[N]$-norm. Ann. of Math. (2) **176**, 1231–1372 (2012) Zbl 1282.11007 MR 2950773

[28] Hildebrand, A., Tenenbaum, G.: Integers without large prime factors. J. Théor. Nombres Bordeaux **5**, 411–484 (1993) Zbl 0797.11070 MR 1265913

[29] Iwaniec, H., Kowalski, E.: Analytic number theory. American Mathematical Society Colloquium Publications 53, American Mathematical Society, Providence, RI (2004) Zbl 1059.11001 MR 2061214

[30] Kelley, Z., Meka, R.: Strong bounds for 3-progressions. arXiv:2302.05537 (2023)

[31] Leng, J.: Improved quadratic Gowers uniformity for the Möbius function. arXiv:2212.09635 (2022)

[32] Lucier, J.: Difference sets and shifted primes. Acta Math. Hungar. **120**, 79–102 (2008) Zbl 1164.11063 MR 2431361

[33] Manners, F.: Quantitative bounds in the inverse theorem for the Gowers U^{s+1}-norms over cyclic groups. arXiv:1811.00718 (2018)

[34] Matomäki, K., Radziwiłł, M., Tao, T., Teräväinen, J., Ziegler, T.: Higher uniformity of bounded multiplicative functions in short intervals on average. Ann. of Math. (2) **197**, 739–857 (2023) Zbl 07654586 MR 4543441

[35] Montgomery, H. L., Vaughan, R. C.: Multiplicative number theory. I. Classical theory. Cambridge Studies in Advanced Mathematics 97, Cambridge University Press, Cambridge (2007) Zbl 1142.11001 MR 2378655

[36] Reingold, O., Trevisan, L., Tulsiani, M., Vadhan, S.: Dense subsets of pseudorandom sets. In: Proceedings of the 49th IEEE FOCS, 76–85 (2008)

[37] Ruzsa, I. Z., Sanders, T.: Difference sets and the primes. Acta Arith. **131**, 281–301 (2008) Zbl 1170.11023 MR 2379933

[38] Sanders, T.: On the Bogolyubov–Ruzsa lemma. Anal. PDE **5**, 627–655 (2012) MR 2994508

[39] Sárközy, A.: On difference sets of sequences of integers. III. Acta Math. Acad. Sci. Hungar. **31**, 355–386 (1978) Zbl 0413.10051 MR 487031

[40] Schmidt, W. M.: Small fractional parts of polynomials. Regional Conference Series in Mathematics 32, American Mathematical Society, Providence, RI (1977) Zbl 0362.10032 MR 457360

[41] Shkredov, I. D.: Szemerédi's theorem and problems of arithmetic progressions. Uspekhi Mat. Nauk **61**, no. 6, 111–178 (2006) (in Russian) Zbl 1178.11012 MR 2330014

[42] Tao, T., Teräväinen, J.: Odd order cases of the logarithmically averaged Chowla conjecture. J. Théor. Nombres Bordeaux **30**, 997–1015 (2018) Zbl 1441.11255 MR 3938639

[43] Vaughan, R.-C.: Sommes trigonométriques sur les nombres premiers. C. R. Acad. Sci. Paris Sér. A-B **285**, A981–A983 (1977) Zbl 0374.10025 MR 498434

[44] Walfisz, A.: Weylsche Exponentialsummen in der neueren Zahlentheorie. Mathematische Forschungsberichte 16, Deutscher Verlag der Wissenschaften, Berlin (1963) Zbl 0146. 06003 MR 220685

[45] Walker, A.: Linear inequalities in primes. J. Anal. Math. **145**, 29–127 (2021) Zbl 1497.11249 MR 4361902

[46] Wang, R.: On a theorem of Sárközy for difference sets and shifted primes. J. Number Theory **211**, 220–234 (2020) Zbl 1444.11024 MR 4074556

[47] Wooley, T. D., Ziegler, T. D.: Multiple recurrence and convergence along the primes. Amer. J. Math. **134**, 1705–1732 (2012) Zbl 1355.11005 MR 2999293