LINEAR EQUATIONS IN PRIMES

BEN GREEN AND TERENCE TAO

Abstract. Consider a system Ψ of non-constant affine-linear forms $\psi_1, \dots, \psi_t : \mathbb{Z}^d \to \mathbb{Z}$, no two of which are linearly dependent. Let N be a large integer, and let $K \subseteq [-N, N]^d$ be convex. A generalisation of a famous and difficult open conjecture of Hardy and Littlewood predicts an asymptotic, as $N \to \infty$, for the number of integer points $n \in \mathbb{Z}^d \cap K$ for which the integers $\psi_1(n), \dots, \psi_t(n)$ are simultaneously prime. This implies many other well-known conjectures, such as the twin prime conjecture and the (weak) Goldbach conjecture. It also allows one to count the number of solutions in a convex range to any simultaneous linear system of equations, in which all unknowns are required to be prime.

In this paper we (conditionally) verify this asymptotic under the assumption that no two of the affine-linear forms $\psi_1, \dots, \psi_t$ are affinely related; this excludes the important "binary" cases such as the twin prime or Goldbach conjectures, but does allow one to count "non-degenerate" configurations such as arithmetic progressions. Our result assumes two families of conjectures, which we term the *inverse Gowers-norm conjecture* (GI(s)) and the *Möbius and nilsequences conjecture* (MN(s)), where $s \in \{1, 2, \dots\}$ is the *complexity* of the system and measures the extent to which the forms ψ_i depend on each other. The case $s = 0$ is somewhat degenerate, and follows from the prime number theorem in APs.

Roughly speaking, the inverse Gowers-norm conjecture GI(s) asserts the Gowers U^{s+1}-norm of a function $f : [N] \to [-1, 1]$ is large if and only if f correlates with an s-step nilsequence, while the Möbius and nilsequences conjecture MN(s) asserts that the Möbius function μ is strongly asymptotically orthogonal to s-step nilsequences of a fixed complexity. These conjectures have long been known to be true for $s = 1$ (essentially by work of Hardy-Littlewood and Vinogradov), and were established for $s = 2$ in two papers of the authors. Thus our results in the case of complexity $s \leqslant 2$ are unconditional.

In particular we can obtain the expected asymptotics for the number of 4-term progressions $p_1 < p_2 < p_3 < p_4 \leqslant N$ of primes, and more generally for any (non-degenerate) problem involving two linear equations in four prime unknowns.

Contents

While this work was carried out the first author was a Clay Research Fellow, and is pleased to acknowledge the support of the Clay Mathematics Institute. Some of this work was carried out while he was on a long-term visit to MIT. The second author was supported by a grant from the Packard Foundation.

1. Introduction

A Generalised Hardy-Littlewood Conjecture. Let $P := \{2, 3, 5, \ldots\} \subset \mathbb{Z}$ denote the prime numbers. We refer to the lattice points $(p_1, \ldots, p_t) \in P^t$ as *prime points* in $\mathbb{Z}^t$. A basic problem in additive number theory is to count the number of prime points on a given affine sublattice of $\mathbb{Z}^t$ in a given range. For instance, the twin prime conjecture asserts that the number of prime points in $\{(n, n+2) : n \in \mathbb{Z}\} \subset \mathbb{Z}^2$ is infinite. When the affine lattice is formed by intersecting $\mathbb{Z}^t$ with an affine subspace,

this problem is equivalent to finding solutions to simultaneous linear equations in which all unknowns are prime. To formalise these types of problems more concretely, it is convenient to parameterise this lattice by d affine-linear forms, as follows.

Definition 1.1 (Affine-linear forms)**.** Let $d, t \geqslant 1$ be integers. An *affine-linear form* on $\mathbb{Z}^d$ is a function $\psi : \mathbb{Z}^d \to \mathbb{Z}$ which is the sum $\psi = \dot{\psi} + \psi(0)$ of a linear form $\dot{\psi} : \mathbb{Z}^d \to \mathbb{Z}$ and a constant $\psi(0) \in \mathbb{Z}$. A *system of affine-linear forms* on $\mathbb{Z}^d$ is a collection $\Psi = (\psi_1, \dots, \psi_t)$ of affine-linear forms on $\mathbb{Z}^d$. To avoid trivial degeneracies we shall require that all the affine-linear forms are non-constant and no two forms are rational multiples of each other. The entire system Ψ can be thought of as an affine-linear map from $\mathbb{Z}^d$ to $\mathbb{Z}^t$, which is the sum $\Psi = \dot{\Psi} + \Psi(0)$ of a linear map $\dot{\Psi} : \mathbb{Z}^d \to \mathbb{Z}^t$ and a constant $\Psi(0) \in \mathbb{Z}^t$; we refer to the range $\Psi(\mathbb{Z}^d)$ of this map as an *affine sublattice* of $\mathbb{Z}^t$. We extend Ψ (and $\dot{\Psi}$) in the obvious manner to an affine-linear map from $\mathbb{R}^d$ to $\mathbb{R}^t$. If $N > 0$, we define the *size* $\|\Psi\|_N$ of Ψ relative to the scale N to be the quantity

$$\|\Psi\|_N := \sum_{i=1}^{t} \sum_{j=1}^{d} |\dot{\psi}_i(e_j)| + \sum_{i=1}^{t} \left| \frac{\psi_i(0)}{N} \right| \tag{1.1}$$

where $e_1, \dots, e_d$ is the standard basis for $\mathbb{Z}^d$.

Example 1. The line $\{(n, n+2) : n \in \mathbb{Z}\}$ is the affine lattice associated to the system $\Psi : n \mapsto (n, n+2)$ with $d = 1$ and $t = 2$. This example has bounded size for any $N \geqslant 1$. The system $\Psi : n \mapsto (n, N - n)$ counts pairs of primes which sum to N, and has bounded size at scale N.

In order to count the number of prime points on an affine lattice, it is convenient to use the *von Mangoldt function* $\Lambda : \mathbb{Z} \to \mathbb{R}^+$, defined by setting $\Lambda(n) := \log p$ when $n > 1$ is a power of a prime p, and $\Lambda(n) = 0$ otherwise (in particular, $\Lambda(n) = 0$ whenever $n \leqslant 0$). We are then interested in estimating the sum

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda(\psi_i(n)) \tag{1.2}$$

where K is a convex subset of $\mathbb{R}^d$ and $[t] := \{1, \dots, t\}$.

Remark. We do not necessarily assume that Ψ is injective, that is to say we allow the sum in (1.2) to count a single prime point repeatedly. This freedom will be convenient for us at a later stage of the argument when we increase the number d of parameters in order to place Ψ in a certain normal form. However, in most applications of interest it will indeed be the case that Ψ is injective, and so the prime points are counted without multiplicity.

The prime number theorem asserts that the average value of $\Lambda(n)$ is 1 for positive n and 0 for negative n, so it is first natural (cf. Cramer's model for the primes) to consider the much simpler sum

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} 1_{\mathbb{R}^+}(\psi_i(n))$$

where we use 1_E to denote the indicator of a set E (thus $1_E(x) = 1$ when $x \in E$ and $1_E(x) = 0$ otherwise). Let us assume that the convex body K is contained in the box

$[-N,N]^d$ for some large integer N, and let us also assume the size bounds $\|\Psi\|_N \leqslant L$ for some $L > 0$. Then a simple volume packing argument (see Appendix A) yields the asymptotic

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} 1_{\mathbb{R}^+}(\psi_i(n)) = \beta_\infty + O_{d,t,L}(N^{d-1}) = \beta_\infty + o_{d,t,L}(N^d) \tag{1.3}$$

where the *archimedean factor* β_∞ is defined by

$$\beta_\infty := \mathrm{vol}_d\left(K \cap \Psi^{-1}((\mathbb{R}^+)^t)\right) \tag{1.4}$$

(see §3 for our conventions concerning asymptotic notation). Note that the main term β_∞ is typically of size N^d or so. One can be much more precise about the nature of the error term, but we will not be concerned with quantitative decay rates here. Indeed the rates provided by our later arguments will be poor and often ineffective, and will dominate whatever gains one could extract from the error term in (1.3).

In view of (1.3) and the prime number theorem, one might naïvely conjecture that the expression (1.2) also enjoys the asymptotic $\beta_\infty + o_{d,t,L}(N^d)$. However this is not the case due to local obstructions at small moduli. For instance, we have

$$\sum_{n=1}^{N} \Lambda(qn+b) = \Lambda_{\mathbb{Z}_q}(b)N + o_q(N) \tag{1.5}$$

whenever $q \geqslant 1$ and $|b| \leqslant q$, where $\Lambda_{\mathbb{Z}_q} : \mathbb{Z} \to \mathbb{R}^+$ is the *local von Mangoldt function*, that is the q-periodic function defined by setting $\Lambda_{\mathbb{Z}_q}(b) := \frac{q}{\phi(q)}$ when b is coprime to q and $\Lambda_{\mathbb{Z}_q}(b) = 0$ otherwise. Here $\mathbb{Z}_q := \mathbb{Z}/q\mathbb{Z}$ is the cyclic group of order q and $\phi(q) := |\mathbb{Z}_q^\times|$ is the Euler totient function. We shall refer to (1.5) as the *prime number theorem in APs*. A well-known quantitative version of this result is the Siegel-Walfisz theorem, which establishes the asymptotic (1.5) uniformly in the range $q \leqslant \log^A N$ for any fixed A. In this range, the o-term is ineffective, and if one wishes for an effective error term it is necessary to restrict to $q \leqslant \log^{1-\delta} N$ for some $\delta > 0$. See [11, p. 123] for details.

More generally, given a system $\Psi = (\psi_1, \dots, \psi_t)$ of affine-linear forms, one can define the *local factor* β_q for any integer $q \geqslant 1$ by the formula

$$\beta_q := \mathbb{E}_{n \in \mathbb{Z}_q^d} \prod_{i \in [t]} \Lambda_{\mathbb{Z}_q}(\psi_i(n)). \tag{1.6}$$

The symbol $\mathbb{E}$ denotes expectation or averaging; see §3 for more details. From the Chinese remainder theorem we see that this factor is multiplicative, indeed we have $\beta_q = \prod_{p|q} \beta_p$, where the product is over all primes[1] p dividing q. We then have

Conjecture 1.2 (Generalised Hardy-Littlewood conjecture)**.** *Let N, d, t, L be positive integers, and let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms with size $\|\Psi\|_N \leqslant L$. Let $K \subset [-N,N]^d$ be a convex body. Then we have*

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda(\psi_i(n)) = \beta_\infty \prod_p \beta_p + o_{t,d,L}(N^d) \tag{1.7}$$

[1]More generally, we adopt the convention that whenever a product ranges over p, that p is understood to be restricted to the primes.

where the archimedean factor β_∞ and the local factors β_p for each prime p were defined in (1.4), (1.6).

Roughly speaking, this conjecture asserts that Λ "behaves like" the independent product of $1_{\mathbb{R}^+}$ and $\Lambda_{\mathbb{Z}_p}$, as p ranges over primes. In typical applications, the quantities β_∞ and β_p are quite easy to compute explicitly: see Examples 5-9 below. We shall refer to the quantity $\prod_p \beta_p$ as the *singular product*. The local factors β_p can be easily estimated:

Lemma 1.3 (Local factor bounds)**.** *With the hypotheses of Conjecture 1.2, we have $\beta_p = 1 + O_{t,d,L}(p^{-1})$. If furthermore no two of the forms $\psi_1, \ldots, \psi_t$ are affinely related (i.e. no two of the forms $\dot\psi_1, \ldots, \dot\psi_t$ are parallel), or if $p > C(t,d,L)N$ for some sufficiently large constant $C(t,d,L)$, then we have $\beta_p = 1 + O_{t,d,L}(p^{-2})$.*

Proof. Without loss of generality we may assume p to be large compared to t, d, L, as the claim is trivial otherwise. Let n be selected uniformly at random from $\mathbb{Z}_p^d$. Since the ψ_i are non-constant, we easily see that $\Lambda_{\mathbb{Z}_p}(\psi_i(n))$ will equal $\frac{p}{p-1}$ with probability $1 - \frac{1}{p}$, and 0 otherwise. In particular the product in (1.6) is equal to $(\frac{p}{p-1})^t = 1 + O_t(\frac{1}{p})$ with probability $1 - O_t(\frac{1}{p})$ and zero otherwise, which gives the first bound on β_p. Now suppose that either no two of $\psi_1, \ldots, \psi_t$ are affinely related, or that $p > C(t,d,L)N$ for some sufficiently large $C(t,d,L)$. Then for any $1 \leqslant i < j \leqslant t$, we see from elementary linear algebra that $\psi_i(n)$ and $\psi_j(n)$ will simultaneously be divisible by p with probability $O(\frac{1}{p^2})$; the point is that the hypotheses imply that[2] ψ_i and ψ_j cannot be linear multiples of each other modulo p. The desired bound on β_p then follows from a simple application of the Bonferroni inequalities (that is, the fact that truncations of the inclusion-exclusion formula give upper and lower bounds alternately). □

In particular we see that the singular series $\prod_p \beta_p$ is always convergent (though it could vanish, thanks to the presence of the small primes $p = O_{t,d,L}(1)$).

A straightforward argument shows that Conjecture 1.2 implies a conjecture which counts primes more explicitly:

Conjecture 1.4 (Generalised Hardy-Littlewood conjecture, again)**.** *Let N, d, t, L, Ψ, K be as in Conjecture 1.2. Then*

$$\begin{aligned} |K \cap \mathbb{Z}^d \cap \Psi^{-1}(P^t)| &= \#\{n \in K \cap \mathbb{Z}^d : \psi_1(n), \ldots, \psi_t(n) \text{ prime}\} \\ &= (1 + o_{t,d,L}(1)) \frac{\beta_\infty}{\log^t N} \prod_p \beta_p + o_{t,d,L}\left(\frac{N^d}{\log^t N}\right). \end{aligned} \tag{1.8}$$

Remarks. It would be slightly more accurate to replace $\frac{\beta_\infty}{\log^t N}$ with the more precise expression

$$\int_K \prod_{j \in [t]} \frac{1_{\psi_j(x) > 2}}{\log \psi_j(x)}\, dx,$$

[2]One could view this as a (very simple) manifestation of the Lefschetz principle.

but the difference between these two expressions can be absorbed into the qualitative $o_{t,d,L}()$ error terms. In most (though not quite all) cases, the singular series $\prod_p \beta_p$ is bounded by $O_{t,d,L}(1)$, which allows one to absorb the first error term into the second. Informally speaking, this conjecture asserts that the probability that a randomly selected point in $\Psi(\mathbb{Z}^d) \cap \mathbb{Z}_+^t$ of magnitude N is a prime point is asymptotically $\frac{1}{\log^t N} \prod_p \beta_p$.

Sketch proof of Conjecture 1.4 assuming Conjecture 1.2. Let $0 < \varepsilon < 1$ be a small quantity (depending on N, d, t, L) to be chosen later. The contribution to (1.8) where $\min_{1 \leqslant i \leqslant t} |\psi_i(n)| \leqslant N^{1-\varepsilon}$ can easily be shown to be $o_{t,d,L,\varepsilon}(N^{d-\varepsilon/2})$ by crude estimates; the analogous contribution to (1.7) can similarly be shown to be $o_{t,d,L,\varepsilon}(N^d)$. The contribution to (1.7) where at least one of the $\psi_i(n)$ is a power of a prime $p^2, p^3, \dots$ can similarly be shown to be $o_{t,d,L}(N^d)$. Finally, for the remaining non-zero contributions to (1.7), the quantity $\prod_{i \in [t]} \Lambda(\psi_i(n))$ is equal to $(1 + O(t\varepsilon)) \log^t N$. Putting all this together, we see that the left-hand side of (1.8) is

$$(1 + O(t\varepsilon)) \frac{\beta_\infty}{\log^t N} \prod_p \beta_p + o_{t,d,L,\varepsilon}(\frac{N^d}{\log^t N}).$$

Setting ε to be a sufficiently slowly decaying function of N (for fixed t, d, L) we obtain the claim. □

Note that the case $d = t = 1$ of the generalised Hardy-Littlewood conjecture is essentially the prime number theorem in APs (1.5). We have been referring to the *generalised* Hardy-Littlewood conjecture because Hardy and Littlewood [28] in fact only conjectured an asympotic for the number of $n \leqslant N$ for which the forms $n + b_1, \dots, n + b_t$ are all prime. If this were generalised to deal with the case of forms $a_1 n + b_1, \dots, a_t n + b_t$ – the case $d = 1$ of Conjecture 1.2 – then a d-parameter version along the lines we have been discussing would follow easily by holding $d - 1$ of the variables fixed and summing in the remaining one. One has the impression that, had they thought to ask the question, Hardy and Littlewood would easily have produced a conjecture for the asymptotic formula. The name of Dickson is sometimes associated to this circle of ideas. In the 1904 paper [12], he noted the obvious necessary condition on the a_i, b_i in order that the forms $a_1 n + b_1, \dots, a_t n + b_t$ might all be prime infinitely often and suggested that this condition might also be sufficient.

Dickson also suggested that the "experts in the new Dirichlet theory" try their hand at establishing this. His hope has yet to be realised, however, since the $d = 1$, $t > 1$ case of Conjecture 1.2 seems to be extremely difficult. The twin prime, Sophie Germain, and weak[3] even Goldbach conjectures, for instance, follow easily from the $d = 1$, $t = 2$ case of the conjecture. These cases are probably well beyond the reach of current technology, although we remark that if one replaces the von Mangoldt function Λ with substantially simpler weight functions arising from the Selberg Λ^2 sieve then such asymptotics can be obtained by standard sieve theory methods (see Theorem D.3). This in turn leads to *upper* bounds on (1.2) which differ from (1.7) only by a multiplicative constant depending only on d, t, L.

[3]That is, the conjecture that every *sufficiently large* even number is the sum of two primes.

Note also that it is possible to establish the case $d = 1$, $t > 1$ of the Hardy-Littlewood conjecture *on average* over the choice of forms $\psi_1, \dots, \psi_t$ in a certain sense: see [3]. This essentially amounts to increasing d, which can place one back in the "finite complexity" regime discussed below.

COMPLEXITY. We will not make any progress on the $d = 1$, $t > 1$ case here, but instead focus on the substantially simpler cases when $d > 1$ and the system is "finite complexity" in the following sense.

Definition 1.5 (Complexity)**.** Let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms. If $1 \leqslant i \leqslant t$ and $s \geqslant 0$, we say that Ψ has *i-complexity at most s* if one can cover the $t-1$ forms $\{\psi_j : j \in [t]\backslash\{i\}\}$ by $s+1$ classes, such that ψ_i does not lie in the affine-linear span of any of these classes. The *complexity* of the Ψ is defined to be the least s for which the system has i-complexity at most s for all $1 \leqslant i \leqslant t$, or ∞ if no such s exists.

Remark. It is easy to see that one can replace "cover ... by" by "partition ... into" in the above definition without affecting the definition of i-complexity or complexity. While partitions are slightly more natural here than covers, we prefer to use covers as it makes it a little easier to compute the complexity in some cases.

Examples 1. The system $\Psi(n_1, \dots, n_d) := (n_1, \dots, n_d)$, which counts d-tuples of independent primes, has complexity 0, because no form u_i lies in the affine span of all the other forms. For any $k \geqslant 2$, the system $\Psi(n_1, n_2) := (n_1, n_1 + n_2, \dots, n_1 + (k-1)n_2)$, which counts arithmetic progressions of primes of length k, has complexity $k-2$, because each form does not lie in the affine span of any other *individual* form, though it is in the affine span of any two other forms. The system $\Psi(n_1, n_2) := (n_1, n_2, N - n_1 - n_2)$, which counts triples of primes that sum to a fixed number N, has complexity 1. The system $\Psi(n_1, n_2) := (n_1, n_2, n_1 + n_2 - 1, n_1 + 2n_2 - 2)$, which counts progressions of primes of length three, whose difference $n_2 - 1$ is one less than a prime, has complexity 2. The system $\Psi(n_1) := (n_1, n_1 + 2)$, which counts twin primes, has infinite complexity. So too does the system $\Psi(n_1) := (n_1, N - n_1)$, which counts pairs of primes which sum to a fixed number N, as well as $\Psi(n_1) = (n_1, 2n_1 + 1)$, which counts Sophie Germain primes. More generally, any system with $d = 1$ and $t > 1$ has infinite complexity.

Example 2 (Cubes). Let $d \geqslant 2$ and $t := 2^{d-1}$. Then the system

$$\Psi(n_1, \dots, n_d) := \big(n_1 + \sum_{j \in A} n_j\big)_{A \subseteq \{2,\dots,d\}},$$

(which counts $(d-1)$-dimensional cubes whose vertices are all prime) has a very large value of t, but has complexity at most $d-2$. For instance, if one considers the form n_1, then one can cover the other $t-1$ forms by $d-1$ classes, with the i^{th} class consisting of those forms which involve n_{i+1}, then n_1 is not in the affine span of any of these classes because the i^{th} class always assigns the same coefficient to both n_1 and n_{i+1}. The other forms can be treated similarly after "reflecting" the cube appropriately.

Example 3 (IP_0 cubes). Let $d \geqslant 1$ and $t := 2^d - 1$. Then the system

$$\Psi(n_1, \dots, n_d) := \big(1 + \sum_{j \in A} n_j\big)_{A \subseteq [d]; A \neq \emptyset},$$

which counts d-dimensional cubes pinned at the origin whose remaining vertices are one less than a prime, also has a large value of t but has complexity at most $d-1$, for reasons similar to the previous example.

In fact in Example 2 the complexity is *exactly* $d-2$, whilst in Example 3 it is *exactly* $d-1$. We leave the proofs to the reader.

Example 4 (Balog's example). Let $d \geqslant 2$ and $t := \frac{d(d+1)}{2}$. Then the system

$$\Psi(n_1, \dots, n_d) := (n_i + n_j + 1)_{1 \leqslant i \leqslant j \leqslant d},$$

which counts d-tuples of odd primes $p_1, \dots, p_d$, all of whose midpoints $\frac{p_i+p_j}{2}$ are also prime, has complexity 1, even though t is quite large. Indeed, if one considers the form $n_i + n_j + 1$ with $i < j$, one can partition the other $t-1$ forms into two classes, those which do not involve n_i, and those which do involve n_i (and hence do not involve n_j), and $n_i + n_j + 1$ is an affine-linear combination of neither of these two classes. If instead one considers the form $n_i + n_i + 1 = 2n_i + 1$, one can partition the other $t-1$ forms into two classes, those which involve n_i (and one other n_j), and those which do not involve n_i at all, and again $2n_i + 1$ is an affine-linear combination of neither of these two classes.

The complexity is a little difficult to compute directly, but the following lemma gives some easy bounds on this quantity.

Lemma 1.6 (Complexity bounded by codimension)**.** *Let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms. Then this system has finite complexity if and only if no two of the ψ_i are affinely dependent. Furthermore, in this case the complexity of the system is less than or equal to $t - \dim(\dot{\Psi})$.*

Proof. If two of the forms ψ_i and ψ_j are affinely related, then it is not possible for the i-complexity to be finite, as ψ_i will lie in the affine span of any collection of forms which contain ψ_j. Conversely, if no two of the ψ_i are affinely related, then the i-complexity is at most $t-2$, as we can partition the $t-1$ forms $\{\psi_j : j \in [t] \backslash \{i\}\}$ into singletons. This gives the first claim of the lemma.

Now suppose that no two of the ψ_i are affinely dependent. Write $r := \dim(\dot{\Psi})$. Choose any homogeneous form, say $\dot{\psi}_1$; this will be nonzero. Relabelling if necessary, we may suppose that $\{\dot{\psi}_1, \dots, \dot{\psi}_r\}$ is a basis for $\dot{\Psi}$. Consider the set $\{\psi_2, \dots, \psi_r\}$ along with the singleton sets $\{\psi_{r+1}\}, \dots, \{\psi_t\}$. Clearly ψ_1 is not in the affine-linear span of any such set, and so the system has 1-complexity at most $t-r$. Since this is true with any ψ_i in place of ψ_1, the claim follows. □

Remark. This lemma is sharp in all the cases treated in Examples 1, but is very far from sharp in Examples 2-4. It asserts that the infinite complexity systems are precisely those which encode a "binary" problem such as the twin prime, Goldbach, Sophie Germain, or prime tuples conjectures. Observe from Lemma 1.6 and Lemma 1.3 that if the system has finite complexity, then $\beta_p = 1 + O_{t,d,L}(\frac{1}{p^2})$ and so the singular series $\prod_p \beta_p$ is either zero, or is bounded above and below by constants depending only on t, d, L. In particular we can eliminate the first error term in (1.8) in this setting.

For systems of complexity 0, The generalised Hardy-Littlewood conjecture follows easily from the prime number theorem in APs (1.5). For systems of complexity 1, the conjecture can be treated by the Hardy-Littlewood circle method (see e.g. [3, 4]). Systems of complexity 2 or higher, on the other hand, are largely out of reach of the circle method and the conjecture has remained open in these cases.

We mention two directions in which a partial approach to high complexity cases of the generalised Hardy-Littlewood conjecture has been made. The first is that a version of the conjecture remains true if one is willing to enlarge sufficiently many of the Λ factors, replacing primes with some notion of an *almost prime*, and adjust the singular series appropriately; see for instance Theorem D.3 for a simplified version of this result. One consequence of this is that *upper bounds* in (1.7) (or (1.8)) are known which are only off by a multiplicative constant of $O_{t,d,L}(1)$.

For certain special systems a *lower bound* of the correct order of magnitude is available. For some systems such as the cube systems in Example 2 this is rather simple, involving nothing more than a few applications of the Cauchy-Schwarz inequality, despite the fact that such systems can have arbitrarily high complexity. However, the task of obtaining asymptotics here is just as difficult as obtaining asymptotics for other systems; see [32] for some related discussion of this phenomenon.

There is also the system $\Psi(n_1, n_2) := (n_1, n_1 + n_2, \dots, n_1 + (k-1)n_2)$ of arithmetic progressions of length k, for which the powerful tool of *Szemerédi's theorem* [39] was available. Despite the fact that these systems can have arbitrarily high complexity, a *lower bound* for (1.7) and (1.8) was established which was again only off by a multiplicative constant. In particular this implied that the primes contain arbitrarily long arithmetic progressions; see [24].

Our arguments in this paper borrow many ideas and results from [24], in particular drawing heavily on the *transference principle* developed in that paper. However we shall not use Szemerédi's theorem in this paper, as it does not apply to the general systems of affine-linear forms studied here. Roughly speaking, one only expects Szemerédi-type theorems for systems which are *homogeneous* (so $\Psi(0) = 0$) and *translation invariant*, that is the lattice $\dot{\Psi}(\mathbb{Z}^d)$ contains the diagonal generator $(1, \dots, 1)$. In any case Szemerédi's theorem only provides lower bounds and not asymptotics.

Main result. Our main result settles the generalised Hardy-Littlewood conjecture for any system of affine-linear forms of finite complexity, conditional on two simpler, partially resolved, conjectures.

Main Theorem (Generalised Hardy-Littlewood conjecture, finite complexity case)**.** *Suppose that the inverse Gowers-norm conjecture* $\mathrm{GI}(s)$ *and the Möbius and nilsequences conjecture* $\mathrm{MN}(s)$ *are true for some finite* $s \geqslant 1$*. Both of these conjectures will be stated formally in §8. Then the generalised Hardy-Littlewood conjecture is true for all systems of affine-linear forms of complexity at most* s*.*

We have deferred the precise statement of the conjectures GI(s) and MN(s) to §8 on account of the fact that both of them are somewhat technical to state formally. The impatient reader may wish to jump to that section to view these conjectures, but for now we settle for informal one-line statements of them.

The inverse Gowers-norm conjecture GI(s) gives an explicit criterion as to when a bounded sequence of complex numbers is "Gowers uniform of order s", this being a measure of pseudorandomness of the sequence; namely, this Gowers uniformity holds whenever the sequence fails to be correlated with any s-step nilsequence.

The Möbius and nilsequences conjecture MN(s) asserts that the Möbius function $\mu(n)$ (which is of course closely related to $\Lambda(n)$) does indeed have negligible correlation with all s-step nilsequences.

Neither of these two conjectures are fully resolved at present. However, the case $s = 1$ is classical and was essentially already present in the work of Hardy-Littlewood and Vinogradov, though not in this language. The conjecture GI(2) was settled more recently in [26], while the conjecture MN(2) was settled in [27]. Because of this, we have the following unconditional result:

Corollary 1.7. *The generalised Hardy-Littlewood conjecture is true for all systems of affine-linear forms of complexity at most* 2*. In particular, thanks to Lemma 1.6, the generalised Hardy-Littlewood conjecture is true for any system* $\Psi = (\psi_1, \dots, \psi_t)$ *in which no two* ψ_i, ψ_j *are affinely dependent, and such that* $\mathrm{codim}(\dot{\Psi}(\mathbb{R}^d)) \leqslant 2$.

We expect both GI(s) and MN(s) to be settled shortly for general s, and hope to report on progress on both of these conjectures in the not-too-distant future[4]. We therefore expect to settle the generalised Hardy-Littlewood conjecture entirely in the finite complexity case, or in other words we should be able to remove the last hypothesis in Corollary 1.7. The only unresolved case of the generalised Hardy-Littlewood conjecture would then be the presumably very hard "binary" or "infinite complexity" case in which two or more of the forms are affinely related.

Let us now state some particular new consequences of our results. The first three are unconditional, while the last two require further progress on the inverse Gowers-norm and Möbius and nilsequences conjectures.

Example 5 (APs of length 4). The number of 4-tuples of primes $p_1 < p_2 < p_3 < p_4 \leqslant N$ which lie in arithmetic progression is $(1 + o(1))\mathfrak{S}_1 \frac{N^2}{\log^4 N}$, where

$$\mathfrak{S}_1 := \frac{3}{4} \prod_{p \geqslant 5} \big(1 - \frac{3p-1}{(p-1)^3}\big) \approx 0.4764.$$

This follows from Corollary 1.7 with the system $\Psi(n_1, n_2) := (n_1, n_1 + n_2, n_1 + 2n_2, n_1 + 3n_2)$, with K being the convex region $\{(n_1, n_2) : 1 \leqslant n_1 \leqslant n_1 + 3n_2 \leqslant N\}$; one has $\beta_\infty = N^2/6$, $\beta_2 = 4$, $\beta_3 = 9/8$, and $\beta_p = 1 - \frac{3p-1}{(p-1)^3}$ for $p \geqslant 5$. Note that the results in

[4]Note added in April 2008: in a recent preprint, the authors have fully resolved the MN(s) conjecture for every s.

[24] do not give this asymptotic, instead yielding a *lower* bound of $(c+o(1))\frac{N^2}{\log^4 N}$ for some explicitly computable but rather small constant $c>0$.

Example 6 (APs of length 3 with common difference $p\pm 1$). The number of triples of primes $p_1<p_2<p_3\leqslant N$ in arithmetic progression, in which the common difference p_2-p_1 is equal to a prime plus 1, is $(1+o(1))\mathfrak{S}_2 N^2\log^{-4}N$, where

$$\mathfrak{S}_2:=\prod_{p\geqslant 3}\big(1-\frac{p^2-4p+1}{(p-1)^4}\big)\approx 1.0481.$$

The same asymptotic holds for progressions in which p_2-p_1 is a prime minus 1. This follows from a similar application of Corollary 1.7 as in Example 5.

Example 7 (Vinogradov 3-primes theorem with a constraint). Let N be a large odd integer. Then the number of distinct representations of N as $p_1+p_2+p_3$ in which p_1-p_2 is equal to a prime minus 1 is $(\mathfrak{S}_3(N)+o(1))\frac{N^2}{\log^4 N}$, where

$$\mathfrak{S}_3(N):=\frac{1}{3}\prod_{\substack{p\geqslant 3\\ p\mid N^3-N}}\big(1-\frac{p^2-4p+1}{(p-1)^4}\big)\prod_{\substack{p\geqslant 3\\ p\nmid N^3-N}}\big(1+\frac{4p-1}{(p-1)^4}\big).$$

Thanks to Lemma 1.3, we see that $\mathfrak{S}_3(N)$ is bounded above and below by absolute positive constants independently of N. Again, this result follows from a specific application of Corollary 1.7.

Example 8 (APs of length k). Let $k\geqslant 2$ be a fixed integer. Assume the GI$(k-2)$ conjecture and the MN$(k-2)$ conjecture. Then the number of k-tuples of primes $p_1<p_2<\cdots<p_k\leqslant N$ which lie in arithmetic progression is

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

where

$$\beta_p:=\begin{cases}\frac{1}{p}\left(\frac{p}{p-1}\right)^{k-1} & \text{if } p\leqslant k\\ \left(1-\frac{k-1}{p}\right)\left(\frac{p}{p-1}\right)^{k-1} & \text{if } p\geqslant k.\end{cases}$$

The $k=4$ case of this is Example 5; the $k=3$ case is due to van der Corput [47]; and the $k=1,2$ cases are equivalent to the prime number theorem. For comparison, the arguments in [24] give an unconditional lower bound of $(c_k+o(1))\frac{N^2}{\log^k N}$ for some $c_k>0$.

Example 9 ($P-1$ and $P+1$ are IP_0-sets). Assume $s\geqslant 0$ is such that the GI(s) and MN(s) conjectures are true. Then (thanks to Example 3) there exist infinitely many $s+1$-tuples $(n_1,\dots,n_{s+1})$ of distinct positive integers such that all of the sums $\{\sum_{i\in A}n_i : A\subseteq[s+1], A\neq\emptyset\}$, are equal to a prime minus 1. Similarly for the primes plus 1. In particular, we unconditionally have the new result that there are infinitely many distinct n_1,n_2,n_3 such that $n_1,n_2,n_3,n_1+n_2,n_1+n_3,n_2+n_3,n_1+n_2+n_3$ are all one less than a prime.

Another consequence of the Main Theorem concerns counting the number of solutions in a given range to a system of linear equations, in which all unknowns are required to be prime:

Theorem 1.8 (Linear equations in primes)**.** *Assume the* GI(s) *and* MN(s) *conjectures. Let $A = (a_{ij})$ be an $s \times t$ matrix of integers, where $s \leqslant t$. Assume the* non-degeneracy conditions *that A has full rank s, and that the only element of the row-space of A over $\mathbb{Q}$ with two or fewer non-zero entries is the zero vector. Let $N > 1$, let $b = (b_1, \dots, b_s) \in \mathbb{Z}^s$ be a vector in $A\mathbb{Z}^t = \{Ax : x \in \mathbb{Z}^t\}$, and suppose that the coefficients $|a_{ij}|$ and the quantities $|b_i/N|$ are uniformly bounded by some constant L. Let $K \subseteq [-N, N]^t$ be convex. Then we have*

$$\sum_{\substack{x \in K \cap \mathbb{Z}^t \\ Ax = b}} \prod_{i \in [t]} \Lambda(x_i) = \alpha_\infty \prod_p \alpha_p + o_{t,L,s}(N^{t-s}), \tag{1.9}$$

where the local densities α_p *are given by*

$$\alpha_p := \lim_{M \to \infty} \mathbb{E}_{x \in [-M,M]^t, Ax = b} \prod_{i \in [t]} \Lambda_{\mathbb{Z}_p}(x_i) \tag{1.10}$$

and the global factor α_∞ *is given by*

$$\alpha_\infty := \#\{x \in \mathbb{Z}^t : x \in K, Ax = b, x_i \geqslant 0\}. \tag{1.11}$$

Theorem 1.8 follows easily from the Main Theorem and some elementary linear algebra: the details may be found in §4. The quantities α_p and α_∞ can be easily computed in practice. One can also formulate an analogue of Theorem 1.8 which counts prime solutions to $Ax = b$, just as Conjecture 1.4 could be deduced from Conjecture 1.2. We leave the details to the reader. Theorem 1.8 is not the most general consequence of the Main Theorem, but it is rather representative. For instance, it already implies Examples 5–8 (and also implies Example 9 if GI(s) and MN(s) are known for all s).

Another simple "qualitative" consequence of the Main Theorem is the following.

Corollary 1.9 (Qualitative generalised H-L conjecture for finite complexity systems)**.** *Suppose that* GI(s) *and* MN(s) *are true for some $s \geqslant 1$. Let $\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of complexity at most s, and let $K \subset \mathbb{R}^d$ be an open convex cone, that is to say an open convex set which is closed under dilations. Suppose that we have the following two local solvability conditions:*

- (Solvability at p) *For each prime p, there exists $n \in \mathbb{Z}^d$ such that the forms $\psi_1(n), \dots, \psi_t(n)$ are all coprime to p.*
- (Solvability at ∞) *There exists $n \in K \cap \mathbb{Z}^d$ such that $\dot{\psi}_1(n), \dots, \dot{\psi}_t(n) > 0$.*

Then there exist infinitely many $n \in K \cap \mathbb{Z}^d$ such that $\psi_1(n), \dots, \psi_t(n)$ are all prime.

Remark. This significantly generalises the main theorem in [24] that the primes contain infinitely many progressions of length k, though for progressions of length $k > 4$ the argument here is conditional on the conjectures GI($k-2$) and MN($k-2$).

Proof. If we truncate K to $[-N, N]^d$, then the hypotheses ensure that $\beta_\infty \gg_{K,d} N^d$ and $\beta_p \neq 0$ for all p. From Lemma 1.3 we conclude that $\beta_\infty \prod_p \beta_p \gg_{K,\Psi,d,t} N^d$, and the claim now follows by letting $N \to \infty$. □

ACKNOWLEDGEMENT. The authors would like to thank the two referees, who both produced extremely careful and helpful reports which have improved the presentation of this paper.

2. Overview of the paper

This section is a kind of roadmap for the rest of the paper, and is somewhat informal in nature. Also, it employs some terminology which will only be rigorously defined in later sections.

The bulk of the paper will be concerned with the proof of the Main Theorem. A substantial portion of our argument consists of reprising the transference principle machinery from [24]. This allows us to model certain unbounded functions, such as Λ, by bounded ones. Another large component of this paper consists of some facts on nilmanifolds which are essentially contained in papers in the ergodic literature, particularly that of Host and Kra [32]. Unfortunately, as our situation here is slightly different from that in [24] we cannot simply cite the results we need directly from that paper, and for similar reasons we cannot cite the nilmanifold material directly. Thus we have placed a large number of appendices in this paper in which we slightly modify the arguments from these sources to suit our present needs.

In §4 we use linear algebra to deduce Theorem 1.8 from the Main Theorem, and also to reduce the Main Theorem to a simplified form, Theorem 4.5, in which the archimedean factor β_∞ is not present and the system Ψ is in a certain "normal form". Then we use the "W-trick" from [24] to eliminate the local factors β_p and reduce matters to establishing a discorrelation estimate, Theorem 5.2, for certain variants $\Lambda'_{b,W} - 1$ of the von Mangoldt function.

In §6, we recall one of the main ingredients of [24]. This is the idea that the von Mangoldt function Λ, or more precisely the variants $\Lambda'_{b,W} - 1$, are dominated by a certain *enveloping sieve* ν which obeys some good pseudorandomness properties. The verification of these properties is essentially given in [24, Ch. 9,10]. We take the opportunity, in Appendix D, to give a simpler variant along the lines of unpublished notes of the second author [43].

In §7 we recall the *generalised von Neumann theorem* from [24], which allows us to use the pseudorandom enveloping sieve ν to deduce the desired discorrelation estimate, Theorem 5.2, from a Gowers uniformity estimate on $\Lambda'_{b,W} - 1$. This latter estimate is the content of Theorem 7.2. We in fact provide a more general type of generalised von Neumann theorem: the one in [24] was specific to the case of arithmetic progressions, and did not allow one to count points inside an arbitrary convex body K. The basic theory of Gowers uniformity norms is reviewed in Appendix B, whilst the generalised von Neumann theorem itself is proved in Appendix C, following some preliminaries on convex geometry in Appendix A.

To prove the Gowers uniformity estimate, we begin by stating in §8 the two conjectures we need, namely the inverse Gowers-norm conjecture GI(s) and the Möbius and nilsequences conjecture MN(s). At this point we pause to present some easy consequences of these conjectures, deducing in §9 some results concerning the behaviour of the Möbius and Liouville functions along systems of linear forms. These functions have an advantage over Λ, in that they are bounded by 1.

In §10 we apply the transference principle technology from [24] to extend the inverse Gowers-norm conjecture GI(s) to cover functions which are bounded only by a pseudorandom measure. This result, Proposition 10.1, is in a sense the conceptual heart of the paper. Once this is done the matter is reduced to the task of showing that $\Lambda'_{b,W} - 1$ is asymptotically orthogonal to nilsequences. The precise statement of such a result is Proposition 10.2.

At this point we need a technical reduction, replacing a nilsequence by a slightly better behaved *averaged nilsequence*. This reduction is carried out in §11, and uses some basic structural facts about nilmanifolds and the cubes within them. These facts are somewhat difficult to extract from the literature, so we give them in Appendix E. In preparing this appendix we benefitted much from conversations with Sasha Leibman.

Finally, to show that $\Lambda'_{b,W} - 1$ is asymptotically orthogonal to an averaged nilsequence, we split Λ into a "smooth" part $\Lambda^\sharp$ and a "rough" part $\Lambda^\flat$. This is a fairly standard construction in analytic number theory which we learnt from [34]. The contribution of the smooth part $\Lambda^\sharp$ can be handled by the Gowers-Cauchy-Schwarz inequality (B.12), combined with correlation estimates for truncated divisor sums. The latter type of estimates are given in Appendix D – the technology is that we used to build the enveloping sieve. The rough part $\Lambda^\flat$ can be handled by the Möbius and nilsequences conjecture MN(s), thus concluding the proof.

In §13 we gather some concluding remarks concerning possible extensions of our results, as well as possibilities for making our estimates effective. We also indicate a proof of (say) the asymptotic in Example 5 which is somewhat shorter than the one given here, but is harder to motivate from the conceptual point of view.

In §14 we gather some remarks concerning bounds for the error terms in our main results. The most interesting part of this discussion focusses on what can be said assuming GRH, since unconditionally all error terms are at present completely ineffective.

The remainder of the paper consists of appendices which supply proofs for various results that we need, but which require techniques which are either standard or somewhat outside the line of the main portion of the paper.

3. General notation

Our conventions for asymptotic notation are as follows. We use $O_{a_1,\dots,a_k}(X)$ to denote a quantity which is bounded in magnitude by $C_{a_1,\dots,a_k}X$ for some finite positive quantity $C_{a_1,\dots,a_k}$ depending only on $a_1,\dots,a_k$; we also write $Y \ll_{a_1,\dots,a_k} X$ or $X \gg_{a_1,\dots,a_k} Y$ for the estimate $|Y| \leqslant O_{a_1,\dots,a_k}(X)$.

In this paper we always think of the parameter N as "large" or "tending to infinity". Thus we use $o_{a_1,\dots,a_k}(X)$ to denote a quantity bounded by $c_{a_1,\dots,a_k}(N)X$, where $c_{a_1,\dots,a_k}(N)$ is a quantity which goes to zero as $N \to \infty$ for each fixed $a_1,\dots,a_k$. We do not assume that the convergence is uniform in these parameters $a_1,\dots,a_k$.

We do not require the implied constants $C_{a_1,\dots,a_k}$, $c_{a_1,\dots,a_k}(N)$ to be effective. While the arguments presented in this paper are entirely effective, the bounds that arise in the Möbius and nilsequences conjecture MN(s), Conjecture 8.5, inevitably involve Siegel zeroes and are thus ineffective with current technology. They are, however, effective if the GRH is assumed.

The o-notation being reserved for functions which become small as $N \to \infty$, we introduce a further notation, the κ-notation, for functions which tend to zero as their parameters become *small*. Thus $\kappa(\delta)$ denotes a quantity which tends to 0 as $\delta \to 0$. Once again the κ may be subscripted by other parameters, indicating a rate of decay which depends on those parameters.

We will frequently take advantage of the fact that two errors involving different parameters can often be concatenated by choosing one of the parameters properly. To give a typical example, suppose we have a quantity $Q(N)$ for which we have established the bound

$$Q(N) \leqslant o_\epsilon(1) + \kappa(\epsilon) \tag{3.1}$$

where $\epsilon \in (0,1)$ is a parameter at our disposal and $Q(N)$ does not depend on ϵ. Then we can concatenate the two error terms by optimising in ϵ and conclude that

$$Q(N) = o(1). \tag{3.2}$$

Indeed for fixed ϵ one may choose N so large that the $o_\epsilon(1)$ term in (3.1) is at most ϵ. This means that $Q(N) = \epsilon + \kappa(\epsilon)$, still a function of the form $\kappa(\epsilon)$. Since ϵ can be as small as one likes, one obtains $Q(N) = o(1)$. Note that this kind of trick was already used to deduce Conjecture 1.4 from Conjecture 1.2.

If A is a finite non-empty set and $f : A \to \mathbb{C}$ is a function, we write $|A|$ for the cardinality of A and $\mathbb{E}_{x\in A} f(x) := \frac{1}{|A|}\sum_{x\in A} f(x)$ for the average of f on A. We extend this notation to functions of several variables in the obvious manner, thus for instance $\mathbb{E}_{x\in A, y\in B} f(x,y) := \frac{1}{|A||B|}\sum_{x\in A}\sum_{y\in B} f(x,y)$.

For any integer $N \geqslant 1$, we use $[N]$ to denote the discrete interval $[N] := \{1,\dots,N\}$, while $\mathbb{Z}_N$ denotes the cyclic group $\mathbb{Z}_N := \mathbb{Z}/N\mathbb{Z}$. At some places in the argument it will

be convenient to pass from intervals $[N]$ to cyclic groups $\mathbb{Z}_N$, possibly after modifying N by a constant multiplicative factor.

The letter i is too important for use only as the square-root of minus one. Occasionally it will be used in this capacity and as an index in the same formula. This ought not to cause any confusion; an earlier attempt to write $\sqrt{-1}$ throughout made several of our formulae rather difficult to read.

In an earlier version of the paper we used vector notation such as $\vec{x}$ to indicate that certain elements lay in product spaces such as $\mathbb{Z}^d$. It was discovered that consistent use of this notation rendered certain of our expressions rather difficult to read, and so we have abandoned this practice. The reader may, at certain times, need to carefully remind herself of the spaces in which certain variables take values.

Important convention. For the rest of the paper, the parameters t, d, s, L (which control the size and complexity of our system $\Psi = (\psi_i)_{i\in[t]}$ of linear forms). All implied constants in the $\ll$, $O(\ )$, or $o(\ )$ notation are understood to be dependent on these parameters t, d, s, L, even if we do not subscript them explicitly. In particular, any quantity depending just on t, d, s, L is automatically $O(1)$. Note however that we do allow our system Ψ to vary (for instance, in order to encompass Vinogradov's three-primes theorem, Ψ must depend on N), and our estimates will be uniform in the choice of Ψ so long as the parameters t, d, s, L remain fixed.

4. Linear algebra reductions

In this section we show how the Main Theorem implies Theorem 1.8, and also reduce the Main Theorem to the case in which the system Ψ is placed in a suitable "normal form". More precisely, in this section we reduce both the Main Theorem and Theorem 1.8 to the simpler Theorem 4.5. Our methods here use only elementary linear algebra. In particular we do not require precise knowledge of exactly what the conjectures GI(s), MN(s) are at this point. We will however restrict to the case $s \geqslant 1$, because the case $s = 0$ follows from the $s = 1$ case (note that the conjectures GI(1), MN(1) are known to be true) and in any event the $s = 0$ case can be easily deduced from (1.5). This allows us to avoid some degeneracies later on.

Derivation of Theorem 1.8 from the Main Theorem. Suppose that we are in the situation of the Main Theorem. Because A has full rank, and b lies in the set $A\mathbb{Z}^t$, the set $\Gamma := \{x \in \mathbb{Z}^t : Ax = b\}$ is a non-empty affine sublattice of $\mathbb{Z}^t$ of rank $d := t - s$. Since $b = O(N)$ and A have bounded integer coordinates, it is not hard to see that Γ must contain at least one point of magnitude $O(N)$. For instance, one could apply any standard linear algebra algorithm to produce an element of Γ, which will then necessarily have magnitude $O(N)$ from inspection of the algorithm. Furthermore, the generators of this lattice can also be chosen to have magnitude $O(1)$, again by applying standard linear algebra algorithms. Thus we have a multiplicity-free parameterisation $\Gamma = \Psi(\mathbb{Z}^{t-s})$ for some system of affine-linear forms $\Psi = (\psi_1, \dots, \psi_t)$ with $\|\Psi\|_N = O(1)$.

The full rank of A ensures that the codimension of $\Psi(\mathbb{Z}^d)$ is the minimal value, namely s. We can then write the left-hand side of (1.9) as

$$\sum_{n \in K' \cap \mathbb{Z}^{t-s}} \prod_{i \in [t]} \Lambda(\psi_i(n))$$

where $K' \subset \mathbb{R}^{t-s}$ is the convex body

$$K' := \{y \in \mathbb{R}^{t-s} : \Psi(y) \in K\}.$$

Note that K' is contained in the box $[-N', N']^{t-s}$ for some $N' = O(N)$.

If two of the ψ_i were affinely dependent then two of the coordinates of lattice points in Γ would obey an affine-linear constraint. This is equivalent to the row space of A containing a non-trivial vector with at most two non-zero entries, which is contrary to assumption. From Lemma 1.6 we conclude that Ψ has complexity at most s. We now invoke the Main Theorem. Comparing (1.7) with (1.9) we see that we will be done as soon as we show that $\alpha_\infty \prod_p \alpha_p = \beta_\infty \prod_p \beta_p + o(N^d)$. For any fixed prime p, the set $\{n \in \mathbb{Z}^{t-s} : \Psi(n) \in [-M, M]^t\}$ is asymptotically uniformly distributed in residue classes in $\mathbb{Z}_p^{t-s}$ in the limit $M \to \infty$ and hence $\alpha_p = \beta_p$. Since the product $\prod_p \beta_p$ is either zero or comparable to 1, it thus suffices to show that $\alpha_\infty = \beta_\infty + o(N^d)$. But this follows from (1.3). □

Elimination of the archimedean factor. We now return to the task of proving the Main Theorem, using some simple linear algebra to obtain some reductions.

First of all, we can use the following easy trick to hide the "archimedean factor" β_∞ from view. Clearly we may intersect K with the convex set $\Psi^{-1}((\mathbb{R}^+)^t)$ and reduce to the case where $\psi_i > 0$ on K; in this case β_∞ is simply the volume of K. In light of (1.3) and the boundedness of the product $\prod_p \beta_p$, we can then rewrite (1.7) as

$$\sum_{n \in K \cap \mathbb{Z}^d} \big(\prod_{i \in [t]} \Lambda(\psi_i(n)) - \prod_p \beta_p \big) = o(N^d). \tag{4.1}$$

Remark. One can easily verify the "local" version of this formula,

$$\sum_{n \in K \cap \mathbb{Z}^d} \big(\prod_{i \in [t]} \Lambda_{\mathbb{Z}_p}(\psi_i(n)) - \beta_p \big) = o_p(N^d);$$

indeed this is a variant of the identity $\alpha_p = \beta_p$ discussed previously.

It turns out to be convenient to strengthen the condition $\psi_i > 0$ slightly, say to $\psi_i > N^{9/10}$. The exact power of N is not important so long as it lies between 0 and 1. One can easily verify, by estimating Λ crudely by $\log N$, that for each i the contribution of the case $0 \leqslant \psi_i(n) \leqslant N^{9/10}$ to (4.1) is $o(N^d)$. We have thus reduced to showing

Theorem 4.1 (Finite complexity generalised H-L conjecture, again)**.** *Let $s \geqslant 1$, and let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms of complexity s. Suppose that the inverse Gowers-norm conjecture* GI(s) *and the Möbius and nilsequences conjecture* MN(s) *are true. Let $N > 1$ and suppose that $\|\Psi\|_N = O(1)$. Let $K \subset [-N, N]^t$ be a convex body such that $\psi_1, \dots, \psi_t > N^{9/10}$ on K. Then* (4.1) *holds.*

Normal form reduction of the Main Theorem. We now reduce Theorem 4.1 further by placing the system Ψ in a convenient "normal form". We denote the standard basis of $\mathbb{Z}^d$ by $e_1, \dots, e_d$.

Definition 4.2 (Normal form)**.** Let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms on $\mathbb{Z}^d$, and let $s \geqslant 0$. We say that Ψ is in *s-normal form* if for every $i \in [t]$, there exists a collection $J_i \subseteq \{e_1, \dots, e_d\}$ of basis vectors of cardinality $|J_i| \leqslant s+1$ such that $\prod_{e \in J_i} \dot{\psi}_{i'}(e)$ is non-zero for $i' = i$ and vanishes otherwise.

If a system is in s-normal form, then we can explicitly see that for each $i \in [t]$ the i-complexity of the system is at most s. Indeed, we can cover the $t-1$ forms $\{\psi_j : j \in [t]\backslash\{i\}\}$ by $|J_i|$ classes, where the class associated to a basis vector $e \in J_i$ is simply the collection of all the forms $\psi_{i'}$ for which $\dot{\psi}_{i'}(e) = 0$; since $\dot{\psi}_i(e) \neq 0$, we see that ψ_i cannot lie in the affine span of such a class. It is, therefore, necessary that a system be of a finite complexity s before admitting an s-normal form. We now investigate the converse relationship, beginning with some illustrative examples.

Example 10. The system of affine-linear forms $\Psi(n_1, n_2) := (n_1, n_1 + n_2, n_1 + 2n_2, n_1 + 3n_2)$, which counts progressions of length four, has complexity 2 but is not in s-normal form for any s. However the system of affine-linear forms

$$\Psi'(n_1, n_2, n_3, n_4) := (n_2 + 2n_3 + 3n_4, -n_1 + n_3 + 2n_4, -2n_1 - n_2 + n_4, -3n_1 - 2n_2 - n_3),$$

which also counts progressions of length four, is also of complexity 2 and is now in 2-normal form.

Example 11. The system in Example 2, which counts $(d-1)$-dimensional cubes, has complexity $d-2$ but is not in s-normal form for any s. However the system

$$\Psi'(n_1, \dots, n_{d-1}, n'_1, \dots, n'_{d-1}) = \big(\sum_{i \in A} n_i + \sum_{i \in [d-1]\backslash A} n'_i\big)_{A \subset [d-1]},$$

which also counts $(d-1)$-dimensional cubes, is also of complexity at most $d-2$ and is now in $(d-2)$-normal form.

Example 12. Let $t := \frac{d(d+1)}{2}$, and consider the system of affine-linear forms

$$\Psi(n_1, \dots, n_d) := (n_i + n_j + 1)_{1 \leqslant i \leqslant j \leqslant d}$$

from Example 4. This system has complexity 1 but is not in s-normal form for any s. However, if we increase the number of parameters from d to $2d$, and consider the system

$$\Psi'(n_1, \dots, n_d, n_{d+1}, \dots, n_{2d}) := \big(n_i + n_j + 1 + n_{d+i} + n_{d+j} - \sum_{k=d+1}^{2d} n_k\big)_{1 \leqslant i \leqslant j \leqslant d},$$

which count the same type of pattern, then this system still has complexity 1 and is now in 1-normal form. Indeed for the off-diagonal forms $i < j$ we may use the basis vectors e_i, e_j, while for the diagonal forms $i = j$ we may use the basis vectors e_i, e_{d+i}.

Remark. Informally speaking, if $(\psi_1, \dots, \psi_t)$ is in s-normal form, then for each form ψ_i there exist a set of at most $s+1$ variables $(n_j)_{j \in J_i}$, such that ψ_i is the only form which truly utilises all the variables at once. As we shall see later, this property will be convenient for establishing a "generalised von Neumann theorem" (Proposition 7.1),

which roughly speaking controls averages such as (4.1) in terms of Gowers uniformity norms, which we shall recall in Appendix B.

Now we investigate the converse question, namely whether every system of complexity s has a normal form representation. To formalise this we first need the concept of *extending* a system of affine-linear forms by adding some "dummy" parameters:

Definition 4.3 (Extensions)**.** Let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms. An *extension* of this system is a system $\Psi' : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ with $d' \geqslant d$, such that

$$\Psi'(\mathbb{Z}^{d'}) = \Psi(\mathbb{Z}^d) \tag{4.2}$$

and furthermore if we identify $\mathbb{Z}^d$ with the subset $\mathbb{Z}^d \times \{0\}^{d'-d}$ of $\mathbb{Z}^{d'}$ in the obvious manner, then Ψ is the restriction of Ψ' to $\mathbb{Z}^d$.

We note that if Ψ is in s-normal form at i, and if Ψ' is an extension of Ψ, then Ψ' is also in s-normal form at i. By the same token, we note also that if $\Psi = (\psi_i)_{i=1}^d$ is in s-normal form, then so is any subsystem $(\psi_i)_{i\in I}$, $I \subset \{1, \ldots, d\}$.

Example 13. In Example 4/Example 12, Ψ' is an extension of Ψ. This is not quite the case in Examples 10, 11, because Ψ is not a restriction of Ψ'. However in these two examples, the direct sum $\Psi \oplus \Psi'$ of the two systems is both an extension of Ψ *and* in normal form; for instance, in Example 10 the system

$$\begin{aligned}\Psi \oplus \Psi'(n_1, n_2, n_1', n_2', n_3', n_4') := (&n_1 + n_2' + 2n_3' + 3n_4', n_1 + n_2 - n_1' + n_3' + 2n_4',\\ &n_1 + 2n_2 - 2n_1' - n_2' + n_4', n_1 + 3n_2 - 3n_1' - 2n_2' - n_3')\end{aligned}$$

is an extension of Ψ which is in 2-normal form.

Lemma 4.4 (Existence of normal forms)**.** *Let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms of some finite complexity s. Then there exists an extension $\Psi' : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ of Ψ which is in s-normal form, where $d' = O(1)$. Furthermore if the original system Ψ had size $\|\Psi\|_N = O(1)$, then the same is true of the extended system Ψ'.*

Proof. Let us fix $i \in [t]$. We shall obtain an extension $\Psi' : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ of Ψ which in s-normal form at i, by which we mean that there is a collection $J_i \subseteq \{e_1, \ldots, e_{d'}\}$ of basis vectors of cardinality $|J_i| \leqslant s+1$ such that $\prod_{e\in J_i} \dot{\psi}_{i'}(e)$ is non-zero for $i' = i$ and vanishes otherwise. Applying this extension procedure once for each value of i we shall obtain the result.

By hypothesis, Ψ has i-complexity at most s, and so we can cover $[t]\backslash\{i\}$ by $s+1$ classes $A_1, \ldots, A_{s+1}$, such that ψ_i is not in the affine-linear span of $\{\psi_j : j \in A_k\}$ for $k \in [s+1]$. In particular, this implies that one can find vectors $f_1, \ldots, f_{s+1} \in \mathbb{Q}^d$ which "witness this fact", that is to say such that $\dot{\psi}_j(f_k) = 0$ and $\dot{\psi}_i(f_k) \neq 0$ all $k \in [s+1]$ and $j \in A_k$. By clearing denominators we can take $f_1, \ldots, f_{s+1} \in \mathbb{Z}^d$. Since $\dot{\Psi}$ has bounded integer coefficients we also see that $f_1, \ldots, f_{s+1} = O(1)$. If we now let $d' := d + s + 1$ and let $\Psi' : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ be the system

$$\Psi'(n, m_1, \ldots, m_{s+1}) := \Psi(n + m_1 f_1 + \ldots + m_{s+1} f_{s+1})$$

for all $n \in \mathbb{Z}^d$ and $m_1, \dots, m_{s+1} \in \mathbb{Z}$, we easily verify that Ψ' satisfies the desired s-normal form property at i, as well as the size bounds on Ψ'. By repeating this procedure once for each i we obtain the claim. □

Using this lemma it is not hard to show that, in order to prove the Main Theorem, it suffices to prove the following result for s-independent systems.

Theorem 4.5 (Primes in affine lattices in normal form)**.** *Let $s \geqslant 1$, and let $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms of complexity s in s-normal form. Suppose that the inverse Gowers-norm conjecture* $\mathrm{GI}(s)$ *and the Möbius and nilsequences conjecture* $\mathrm{MN}(s)$ *are true. Let $N > 1$ and suppose that $\|\Psi\|_N = O(1)$. Let $K \subseteq [-N, N]^t$ be a convex body such that $\psi_1, \dots, \psi_t > N^{8/10}$ on K. Then* (4.1) *holds, that is to say*

$$\sum_{n \in K \cap \mathbb{Z}^d} \big(\prod_{i \in [t]} \Lambda(\psi_i(n)) - \prod_p \beta_p \big) = o(N^d).$$

Proof of the Main Theorem assuming Theorem 4.5. By our earlier reduction it suffices to show that Theorem 4.1 holds. Let Ψ, K, N be as in Theorem 4.1. We may assume N large as the claim is trivial for N small.

Let $\Psi' : \mathbb{Z}^{d'} \to \mathbb{Z}^t$ be the s-normal form extension given by Lemma 4.4. An inspection of the proof of that lemma allows us to find vectors $f_{d+1}, \dots, f_{d'} \in \mathbb{Z}^d$ of magnitude $O(1)$ such that

$$\Psi'(n, m_{d+1}, \dots, m_{d'}) := \Psi(n + m_{d+1} f_{d+1} + \dots + m_{d'} f_{d'}).$$

(One can also deduce the existence of these vectors directly from the conclusions of Lemma 4.4.) We observe that the local factors β'_p associated to the system Ψ' are precisely the same as the local factors β_p associated to Ψ; this is ultimately due to the translation-invariance of $\mathbb{Z}_p$. Now let $K' \subseteq \mathbb{R}^{d'}$ be the convex body

$$K' := \{(n, m_{d+1}, \dots, m_{d'}) \in \mathbb{R}^d \times [-N, N]^{d'-d} : n + m_{d+1} f_{d+1} + \dots + m_{d'} f_{d'} \in K\}.$$

This is contained in $[-N', N']^{d'}$ for some $N' = O(N)$. Applying Theorem 4.5 we conclude

$$\sum_{(n,m) \in K' \cap \mathbb{Z}^{d'}} \big(\prod_{i \in [t]} \Lambda(\psi'_i(n, m)) - \prod_p \beta_p \big) = o(N^{d'}).$$

Making the change of variables $r := n + m_{d+1} f_{d+1} + \dots + m_{d'} f_{d'}$, the left-hand side can be simplified to

$$\Big| [-N, N]^{d'-d} \cap \mathbb{Z}^{d'-d} \Big| \sum_{r \in K \cap \mathbb{Z}^d} \big(\prod_{i \in [t]} \Lambda(\psi_i(r)) - \prod_p \beta_p \big)$$

and (4.1) follows upon dividing out by $(2N+1)^{d'-d}$. □

This completes our linear algebra manipulations. It now remains to prove Theorem 4.5, a task which will occupy the remainder of the paper.

5. The W-trick

In the preceding section we were able to eliminate the archimedean factor β_∞ by assuming that $\psi_1, \dots, \psi_t$ were non-negative on K, and using the formulation (4.1). Now we use a somewhat similar trick, which we term the "W-trick". This was a vital trick in [22, 24, 25], where it was used in similar fashion to eliminate the local factors β_p. Once again, the reductions here will not actually require any knowledge of the two conjectures GI(s) and MN(s), which we shall finally introduce in §8.

Important convention. From now on in the paper, fix some slowly growing function $w = w(N)$. Any function such that $w(N) \leqslant \frac{1}{2}\log\log N$ and $\lim_{N\to\infty} w(N) = \infty$ would suffice; for sake of definiteness we shall conservatively set $w := \log\log\log N$. The exact choice of w is only relevant for determining the decay rate of the $o()$ terms, but as our final decay bounds are ineffective we will not attempt to optimise in w.

We define the quantity $W = W(w)$ by

$$W := \prod_{p \leqslant w} p;$$

since $w \leqslant \frac{1}{2}\log\log N$ we have $W = O(\log^{1/2} N)$. For each $b \in [W]$ with $\gcd(b, W) = 1$, let $\Lambda_{b,W} : \mathbb{Z}^+ \to \mathbb{R}^+$ be the function

$$\Lambda_{b,W}(n) := \frac{\phi(W)}{W}\Lambda(Wn + b) \tag{5.1}$$

where we recall that $\phi(W) = \#\{b \in [W] : \gcd(b, W) = 1\}$ is the Euler totient function of W. Thus for instance the prime number theorem in APs (1.5) asserts[5] that $\Lambda_{b,W}(n)$ has average value 1 as $n \to \infty$. Actually it will be slightly more convenient to work with the variant

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

where Λ' is the restriction of Λ to the primes, i.e. $\Lambda'(p) = \log p$ for all primes p and $\Lambda'(n) = 0$ for non-prime p. Thus Λ' only differs from Λ on the (negligible) set of prime powers $p^2, p^3, \dots$.

Recall that we reduced the task of proving the Main Theorem to that of proving Theorem 4.5. We now make a further reduction, showing that it suffices to prove the following.

Theorem 5.1 (W-tricked primes in affine lattices)**.** *Let $s \geqslant 1$, and suppose that $\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ is a system of affine-linear forms in s-normal form and with $\|\Psi\|_N = O(1)$. Suppose that the inverse Gowers-norm conjecture* GI(s) *and the Möbius and nilsequences conjecture* MN(s) *are true. Let $K \subseteq [-N, N]^t$ be any convex body on which $\psi_1, \dots, \psi_t > N^{7/10}$. Then for any $b_1, \dots, b_t \in [W]$ which are coprime to W, we have*

$$\sum_{n \in K \cap \mathbb{Z}^d} \big(\prod_{i \in [t]} \Lambda'_{b_i,W}(\psi_i(n)) - 1\big) = o(N^d).$$

[5] In order to obtain this statement for w as large as $\frac{1}{2}\log\log N$, one needs a more quantitative version of (1.5) such as the Siegel-Walfisz theorem.

Remark. Note that the bounds on the right do not depend on $b_1, \dots, b_t$. The philosophy here is that the functions $\Lambda'_{b,W}$ should behave "pseudorandomly" with average value one; this is in contrast with Λ, which has many local irregularities with respect to small moduli which necessitate the introduction of the local factors β_p. This philosophy of passing from Λ to the more uniformly distributed $\Lambda'_{b,W}$ underlies the arguments in [24]. In §12 we will have to invert the W-trick and deduce some correlation estimates on $\Lambda'_{b,W}$ from that on Λ.

Proof of the Main Theorem assuming Theorem 5.1. By previous reductions, it suffices to establish Theorem 4.5. Let Ψ, K be as in Theorem 4.5. We may then replace Λ by Λ' as the contribution of the prime powers is easily seen to be negligible. To prove (4.1), it then suffices by (1.3) to show that

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda'(\psi_i(n)) = \mathrm{vol}_d(K) \prod_p \beta_p + o(N^d). \tag{5.2}$$

We may take N to be large, since the claim is trivial otherwise.

Now the upper bound on w ensures that $W \leqslant \log N$. From Lemma 1.3 followed by the multiplicativity of the local factors β we have

$$\prod_p \beta_p = \prod_{p \leqslant w} \beta_p + o(1) = \beta_W + o(1);$$

since $\mathrm{vol}_d(K) = O(N^d)$, we conclude that

$$\mathrm{vol}_d(K) \prod_p \beta_p = \mathrm{vol}_d(K)\beta_W + o(N^d).$$

Now let A be the set

$$A := \{a \in [W]^d : \gcd(\psi_i(a), W) = 1 \text{ for all } i \in [t]\}.$$

Then from (1.6) we have $\beta_W = \left(\frac{W}{\phi(W)}\right)^t |A|/W^d$, which implies that

$$\mathrm{vol}_d(K) \prod_p \beta_p = \sum_{a \in A} \left(\frac{W}{\phi(W)}\right)^t W^{-d} \mathrm{vol}_d(K) + o(N^d). \tag{5.3}$$

Also, from Lemma 1.3 we know that β_W is comparable to 1, and so

$$|A| \ll \left(\frac{\phi(W)}{W}\right)^t W^d. \tag{5.4}$$

Next, note that by a simple expansion we have

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda'(\psi_i(n)) = \sum_{a \in [W]^d} \sum_{\substack{n \in \mathbb{Z}^d \\ Wn + a \in K}} \prod_{i \in [t]} \Lambda'(\psi_i(Wn + a)). \tag{5.5}$$

If a does not lie in A, then $\psi_i(Wn + a)$ will not be coprime to W for some $i \in [t]$. Since $\psi(Wn + a) > N^{7/10}$ by hypothesis, and W is so small compared to N, we see that $\Lambda'(\psi_i(Wn + a)) = 0$. Thus we may restrict a to A. Now for each $a \in A$ and $i \in [t]$, we can write

$$\psi_i(Wn + a) = W\tilde{\psi}_{i,a}(n) + b_i(a)$$

where $b_i(a)$ lies in $[W]$ and is coprime to W, while $\tilde{\psi}_{i,a}$ is a translate of ψ_i whose constant term $\tilde{\psi}_{i,a}(0)$ is $O(N/W)$. Indeed $b_i(a)$ is simply the remainder formed when dividing $\psi_i(a)$ by W. We then have

$$\Lambda'(\psi_i(Wn+a)) = \frac{W}{\phi(W)}\Lambda'_{b_i(a),W}(\tilde{\psi}_{i,a}(n)).$$

It follows from (5.5) that

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]}\Lambda'(\psi_i(n)) = \sum_{a\in A}\left(\frac{W}{\phi(W)}\right)^t \sum_{\substack{n\in\mathbb{Z}^d\\ Wn+a\in K}}\prod_{i\in[t]}\Lambda'_{b_i(a),W}(\tilde{\psi}_{i,a}(n)). \tag{5.6}$$

However from Theorem 5.1 (with N replaced by $\tilde{N} = O(N/W)$ and $\tilde{K} := (K-a)/W$: note that $\|\tilde{\Psi}\|_{\tilde{N}} = O(1)$) we have

$$\sum_{\substack{n\in\mathbb{Z}^d\\ Wn+a\in K}}\Big(\prod_{i\in[t]}\Lambda'_{b_i(a),W}(\tilde{\psi}_{i,a}(n)) - 1\Big) = o\left(\frac{N}{W}\right)^d.$$

Recalling (5.4), this together with (5.6) implies that

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]}\Lambda'(\psi_i(n)) = \sum_{a\in A}\left(\frac{W}{\phi(W)}\right)^t \sum_{\substack{n\in\mathbb{Z}^d\\ Wn+a\in K}} 1 \quad + o(N^d). \tag{5.7}$$

On the other hand a simple volume-packing argument (cf. Appendix A) yields

$$\sum_{\substack{n\in\mathbb{Z}^d\\ Wn+a\in K}} 1 = W^{-d}\operatorname{vol}_d(K) + o\left(\frac{N}{W}\right)^d$$

and so, using (5.4) once more together with (5.7), we see that

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]}\Lambda'(\psi_i(n)) = \sum_{a\in A}\left(\frac{W}{\phi(W)}\right)^t W^{-d}\operatorname{vol}_d(K) + o(N^d).$$

Subtracting this against (5.3) we see that the left-hand side of (5.2) is $o(N^d)$. This proves the claim. □

Theorem 5.1, as we have just seen, implies the Main Theorem. Before moving on to the more substantial arguments in this paper, we give one further simple reduction, deducing Theorem 5.1 from the following variant.

Theorem 5.2 (Final technical reduction)**.** *Let $s \geqslant 1$, and let $\Psi = (\psi_1, \dots, \psi_t) : \mathbb{Z}^d \to \mathbb{Z}^t$ be a system of affine-linear forms in s-normal form. Suppose that the inverse Gowers-norm conjecture* $\mathrm{GI}(s)$ *and the Möbius and nilsequences conjecture* $\mathrm{MN}(s)$ *are true. Let $K \subseteq [-N,N]^t$ be any convex body on which $\psi_1, \dots, \psi_t > N^{7/10}$. Then for any $b_1, \dots, b_t \in [W]$ which are coprime to W, we have*

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]}(\Lambda'_{b_i,W}(\psi_i(n)) - 1) = o(N^d).$$

Indeed, Theorem 5.1 follows immediately from Theorem 5.2 by splitting each $\Lambda'_{b_i,W}$ as $(\Lambda'_{b_i,W} - 1) + 1$, expanding out the product in Theorem 5.1, and using Theorem 5.2 repeatedly, noting that any subsystem of Ψ will still be in s-normal form. □

The remainder of the paper shall be devoted to establishing Theorem 5.2.

6. The enveloping sieve

In previous sections we have reduced matters to establishing a certain discorrelation estimate, Theorem 5.2, for the functions $\Lambda'_{b_i,W} - 1$. A major difficulty in the analysis here is that these functions are not bounded uniformly in N. However, as in [24, 25], we shall be able to import tools from sieve theory. In particular, we use the principle of the "enveloping sieve". This is a well-behaved function ν, some constant multiple of which provides a pointwise bound for the functions $\Lambda'_{b_i,W} - 1$. Of course, the function ν will not be bounded as $N \to \infty$; however it does obey a number of very good correlation or *pseudorandomness* estimates which assert, roughly speaking, that ν "effectively behaves like" the bounded function 1.

To define the notion of pseudorandomness properly we recall the *linear forms condition* and *correlation condition* from [24], modified slightly for the application at hand. In the following three definitions we assume that N is a large positive integer, and that $N' = N'(N)$ is a prime number of size $N < N' \leqslant O_{s,t,d,L}(N)$.

Definition 6.1 (Measures)**.** A *measure* on $\mathbb{Z}_{N'}$ is a function $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ (depending of course on N' and hence on N) with

$$\mathbb{E}_{n \in \mathbb{Z}_{N'}} \nu(n) = 1 + o(1). \tag{6.1}$$

Definition 6.2 (Linear forms condition)**.** Let ν be a measure on $\mathbb{Z}_{N'}$, and let m_0, d_0 and L_0 be positive integer parameters. Then we say that ν satisfies the (m_0, d_0, L_0)-linear forms condition if the following holds: given $1 \leqslant d \leqslant d_0$, $1 \leqslant t \leqslant m_0$, and any finite complexity system $\Psi = (\psi_1, \dots, \psi_t)$ of affine-linear forms on $\mathbb{Z}^d$ with all coefficients of $\dot{\Psi}$ bounded in magnitude by L_0, we have

$$\mathbb{E}_{n \in \mathbb{Z}^d_{N'}} \prod_{i \in [t]} \nu(\psi_i(n)) = 1 + o_{m_0,d_0,L_0}(1). \tag{6.2}$$

In this expression we induce the affine-linear forms $\psi_j : \mathbb{Z}^d_{N'} \to \mathbb{Z}_{N'}$ from their global counterparts $\psi_j : \mathbb{Z}^d \to \mathbb{Z}$ in the obvious manner.

Remarks. Note that (6.2) includes (6.1) as a special case. Strictly speaking, it would be more accurate to call measures "probability densities", and the linear forms condition is really an "affine-linear forms condition", but we will keep the notation as above for brevity and compatibility with [24]. In [24] the coefficients of the affine-linear forms were allowed to be rational with bounded numerator and denominator. Since N' is a large prime, it is always possible in practice to clear denominators and deal only with forms having integer coefficients. Note that Theorem 5.1 is a (conditional) assertion that the $\Lambda_{b,W}$ essentially obey the linear-forms condition. Thus trying to establishing

the linear forms condition for $\Lambda_{b,W}$ would essentially be as hard as trying to prove the Main Theorem. The point of the definition, however, is that it will suffice to achieve the much simpler task of *majorising* $\Lambda_{b,W}$ by constant multiples of measures ν which obey this condition. Finally, we note that the error term in (6.2) is uniform over all choices of constant term $\Psi(0)$.

Definition 6.3 (Correlation condition)**.** Let $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ be a measure, and let m_0 be a positive integer parameter. We say that ν satisfies the *m_0-correlation condition* if for every $1 < m \leqslant m_0$ there exists a weight function $\tau = \tau_m : \mathbb{Z}_{N'} \to \mathbb{R}^+$ which obeys the moment conditions

$$\mathbb{E}_{n \in \mathbb{Z}_{N'}} \tau^q(n) \ll_{m,q} 1 \tag{6.3}$$

for all $1 \leqslant q < \infty$ and such that

$$\mathbb{E}_{n \in \mathbb{Z}_{N'}} \prod_{i \in [m]} \nu(n + h_i) \leqslant \sum_{1 \leqslant i < j \leqslant m} \tau(h_i - h_j) \tag{6.4}$$

for all $h_1, \ldots, h_m \in \mathbb{Z}_{N'}$, not necessarily distinct.

Remarks. Because we are only seeking upper bounds here rather than asymptotics, this condition would follow from a standard upper bound sieve such as Selberg's sieve. One should compare this condition with the much more difficult prime tuples conjecture, which is part of the "infinite complexity" case $d = 1$, $t > 1$ of the generalised Hardy-Littlewood conjecture. The correlation condition will only be used implicitly in this paper, as it is needed in the proof of [24, Proposition 8.1], which is in turn used in the proof of Proposition 10.3.

Let D be a positive integer. We call a measure *D-pseudorandom* if it obeys the (D, D, D)-linear forms and D-correlation conditions. In practice, we shall work with measures which are D-pseudorandom where D is a sufficiently large function of s, d, t, L. The exact value will not be terribly important for our arguments and, whilst it could be specified explicitly, we shall not do so.

Our next task is to show that the functions $\Lambda'_{b_1,W}, \ldots, \Lambda'_{b_t,W}$ can be dominated by a D-pseudorandom measure for any fixed D that we choose, providing we are willing to concede multiplicative constants that depend on D.

Proposition 6.4 (Domination by a pseudorandom measure)**.** *Let $D > 1$ be arbitrary. Then there is a constant $C_0 := C_0(D)$ such that the following is true. Let $C \geqslant C_0$, and suppose that $N' \in [CN, 2CN]$. Let $b_1, \ldots, b_t \in \{0, 1, \ldots, W - 1\}$ be coprime to $W := \prod_{p \leqslant w} p$. Then there exists a D-pseudorandom measure $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ which obeys the pointwise bounds*

$$1 + \Lambda'_{b_1,W}(n) + \ldots + \Lambda'_{b_t,W}(n) \ll_{D,C} \nu(n)$$

for all $n \in [N^{3/5}, N]$, where we identify n with an element of $\mathbb{Z}_{N'}$ in the obvious manner.

The proof of this proposition is a minor variant of that in [24]. For the sake of completeness we present a proof in Appendix D. The constant C is a technicality needed to avoid certain "wraparound" issues when passing from $[N]$ to $\mathbb{Z}_{N'}$ and can be largely ignored.

The philosophy of the *transference principle* developed in [24] is that functions which are dominated by pseudorandom measures behave almost as if they were bounded, for the purposes of computing correlations and other multilinear averages. We shall see examples of this in later sections. For now, we turn to the first significant step in the paper, namely the reduction of matters to establishing a Gowers uniformity norm estimate for $\Lambda'_{b,W} - 1$.

7. Reduction to a Gowers norm estimate

We shall informally refer to a function $f : [N] \to \mathbb{C}$ as being *Gowers uniform of order* s if its Gowers uniformity norm $\|f\|_{U^{s+1}[N]}$ is small; see Appendix B for definitions and basic properties of this norm. A basic principle is that Gowers uniform functions of order s have a negligible impact on multilinear averages of complexity s or less. An example of this is [24, Proposition 5.3], but we will prove a much more general result of this type here. We refer to such statements as *generalised von Neumann theorems*. The name originally came from results in ergodic theory such as [32, Theorem 11.1], but it has been convenient to use the name to describe a large number of contexts in additive combinatorics in which some kind of expression is bounded using Gowers norms[6].

A crucial observation in [24] is that this type of principle also applies to *unbounded* functions, so long as these unbounded functions are in turn dominated pointwise by a suitably pseudorandom measure.

Proposition 7.1 (Generalised von Neumann theorem)**.** *Let s, t, d, L be positive integer parameters as usual. Then there are constants C_1 and D, depending on s, t, d and L, such that the following is true. Let $C_1 \leqslant C \leqslant O_{s,t,d,L}(1)$ be arbitrary and suppose that $N' \in [CN, 2CN]$ is a prime. Let $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ be a D-pseudorandom measure, and suppose that $f_1, \dots, f_t : [N] \to \mathbb{R}$ are functions with $|f_i(x)| \leqslant \nu(x)$ for all $i \in [t]$ and $x \in [N]$. Suppose that $\Psi = (\psi_1, \dots, \psi_t)$ is a system of affine-linear forms in s-normal form with $\|\Psi\|_N \leqslant L$. Let $K \subseteq [-N, N]^d$ be a convex body such that $\Psi(K) \subseteq [N]^t$. Suppose also that*

$$\min_{1 \leqslant j \leqslant t} \|f_j\|_{U^{s+1}[N]} \leqslant \delta$$

for some $\delta > 0$. Then we have

$$\sum_{n \in K} \prod_{i \in [t]} f_i(\psi_i(n)) = o_{\delta,C}(N^d) + \kappa_C(\delta) N^d. \tag{7.1}$$

Remarks. For an explanation of the κ-notation, we refer the reader to §3. One could specify explicit values for C_1, D, but we have not done so. In applications to the primes we will always take $C \geqslant C_0(D)$, where C_0 is the function defined in Proposition 6.4.

This proposition is a variant of [24, Proposition 5.3]. It is somewhat more elaborate than that result in that it applies to a general system of affine linear forms, and one has the flexibility of summing over an arbitrary convex body. Once the convex body is handled by standard techniques, however, the only real tool that is needed is several

[6]Another example of this is the *Koopman von Neumann theorem,* which we will introduce in §10.

applications of the Cauchy-Schwarz inequality. This is a common feature of generalised von Neumann theorems. We give a proof of Proposition 7.1 in Appendix C, which uses some preliminaries in Appendices A, B but is otherwise self-contained. Using Propositions 6.4 and 7.1 we reduce Theorem 5.2, and hence the Main Theorem, to the following Gowers uniformity estimate.

Theorem 7.2 (Gowers uniformity estimate)**.** *Let $N, w > 1$, and let $b \in [W]$ be coprime to $W = \prod_{p \leqslant w} p$. Suppose that the inverse Gowers-norm conjecture* $\mathrm{GI}(s)$ *and the Möbius and nilsequences conjecture* $\mathrm{MN}(s)$ *are true for some $s \geqslant 1$. Then we have*

$$\|\Lambda'_{b,W} - 1\|_{U^{s+1}[N]} = o(1).$$

Remark. Observe (cf. Examples 2 and 11) that this theorem is a special case of Theorem 5.2. Thus the generalised von Neumann theorem, Proposition 7.1, can be viewed as an assertion that the U^{s+1} average is "universal" or "characteristic" among all multilinear averages of complexity s, even when dealing with functions that are bounded only by a pseudorandom measure.

Proof of Main Theorem assuming Theorem 7.2. By previous reductions, is suffices to prove Theorem 5.2. Let the notation and assumptions be as in that theorem. By enlarging N by a multiplicative factor of $O(1)$ if necessary we may assume that $\Psi(K) \subseteq [N]^t$. Let $D = D_{s,t,d,L}$ be the constant in Proposition 7.1, and set $C := \max(C_1, C_0(D))$, where C_0 is the function appearing in Proposition 6.4 and C_1 is the one appearing in Proposition 7.1. Applying Bertrand's postulate, we may select a prime N' such that $CN \leqslant N' \leqslant 2CN$. Let ν be the D-pseudorandom measure given by (6.4). Then the functions $f_i(n) := c \cdot (\Lambda'_{b_i,W} - 1)$ will be pointwise dominated in magnitude by ν for some suitably small constant $c = c_{s,t,d,L} > 0$. Applying Theorem 7.2 and Proposition 7.1, we obtain the desired estimate after dividing out the factors of c. □

We have now completed yet another reduction, and it remains to prove Theorem 7.2. Note that we have eliminated the system Ψ of affine-linear forms, as well as the convex body K, replacing them both with the Gowers norm $U^{s+1}[N]$; the parameters d, t have also disappeared. In order to proceed further, we need to exploit some deeper facts and conjectures concerning the Gowers norm. In particular we shall shortly need the *inverse Gowers-norm conjecture* $\mathrm{GI}(s)$, to which we now turn.

8. The inverse Gowers-norm and Möbius and nilsequences conjectures

Nilsequences. The purpose of this section is to state the two conjectures $\mathrm{GI}(s)$ and $\mathrm{MN}(s)$ which have appeared in many of the above theorems, most recently in Theorem 7.2. Both conjectures revolve around the concept of a *nilsequence*, which we now pause to recall.

Definition 8.1 (Nilmanifolds and nilsequences)**.** Let G be a connected, simply connected, Lie group. We define the *central series* $G_0 \supseteq G_1 \supseteq G_2 \supseteq \dots$ by defining $G_0 = G_1 = G$, and $G_{i+1} = [G, G_i]$ for $i \geqslant 2$, where the commutator group $[G, G_i]$ is the group generated by $\{ghg^{-1}h^{-1} : g \in G, h \in G_i\}$. We say that G is *s-step nilpotent*

if $G_{s+1} = 1$. Let $\Gamma \subseteq G$ be a discrete, cocompact subgroup. Then the quotient G/Γ is called an s-step nilmanifold. If $g \in G$ then g acts on G/Γ by left multiplication, $x \mapsto g \cdot x$. By a *an s-step nilsequence*, we mean a sequence of the form $(F(g^n x))_{n\in\mathbb{N}}$, where $x \in G/\Gamma$ is a point and $F : G/\Gamma \to \mathbb{R}$ is a continuous function. We say that the nilsequence is *1-bounded* if F takes values in $[-1, 1]$.

Remark. For a full technical treatment of nilsequences, see [9]. The reader might consult [6, 32, 35] for the ergodic theory perspective, or other papers of the authors [23, 26, 27] for various discussions more-or-less in the spirit of additive combinatorics.

As remarked above, the exact definition of a nilsequence will not be terribly important to our arguments here. In the $s = 2$ case, representative examples of nilsequences are those associated to the *Heisenberg nilmanifold*, which is discussed in detail in [6, 23, 26, 27]. See also the proof of Proposition 8.4.

Remark. Note that we are requiring our nilpotent groups to be connected and simply connected. The latter hypothesis is not overly restrictive, since if G *is* connected, then it may be assumed to be simply connected by passing to a universal cover. The connectedness assumption however is more substantial; the nilpotent groups constructed in the ergodic theory literature (e.g. in [32]) are not always shown to be connected. However, Sasha Leibman [36] has indicated to us that it suffices, in the context of the GI(s) conjecture, to deal with connected G. We will elaborate on this point in a future paper if necessary, but the issue does not need to be addressed here. This is because the arguments used in proving the cases $s \leqslant 2$, which are the only cases of the conjectures established so far, give connectedness as a byproduct.

As we shall need to be rather quantitative regarding these nilmanifolds, we shall arbitrarily endow[7] each nilmanifold G/Γ with a smooth Riemannian metric $d_{G/\Gamma}$. We then define the *Lipschitz constant* of a nilsequence $F(g^n x)$ to be the Lipschitz constant of F.

Remark. Note that the Lipschitz constant of a nilsequence depends on the choice of metric $d_{G/\Gamma}$ one places on the nilmanifold; there is no obvious canonical metric to assign to any given nilmanifold, and so the Lipschitz constant is a somewhat arbitrary quantity. However if one replaces the metric with another smooth Riemannian metric then from the compactness of G/Γ we see that the Lipschitz constant is only affected by at most a multiplicative constant. One could replace the Lipschitz constant here by other quantitative measures of regularity, such as Hölder continuity norms or C^k norms, but this will not significantly affect the statements of the conjectures here, basically because a function which is controlled in one of these norms can be approximated in a quantitative manner as the uniform limit of functions controlled in any other of these norms.

[7]Strictly speaking, we are abusing notation here; a nilmanifold should not be represented solely by the quotient space G/Γ, but rather as a quadruplet $(G, \Gamma, G/\Gamma, d_{G/\Gamma})$ (and the Lie group G should in turn be expanded to explicitly mention the group operations, coordinate charts, etc.). Similarly, the nilsequence should not be represented solely as $F(g^n x)$, but should really be the octuplet $(G, \Gamma, G/\Gamma, d_{G/\Gamma}, g, x, F, (n \mapsto F(g^n x)))$. However we shall continue to abuse notation in order to simplify the exposition.

Remark. The Lipschitz nilsequences form an algebra in the following sense: if $f(n)$ is an s-step nilsequence on G/Γ with Lipschitz constant M, and $\widetilde{f}(n)$ is an s-step nilsequence on $\widetilde{G}/\widetilde{\Gamma}$ with Lipschitz constant $\widetilde{M}$, and both nilsequences are bounded by $O(1)$, then $f(n) \pm \widetilde{f}(n)$ or $f(n)\widetilde{f}(n)$ is an s-step nilsequence on the product nilmanifold $(G/\Gamma) \times (\widetilde{G}/\widetilde{\Gamma})$ with Lipschitz constant $O_{M,\widetilde{M}}(1)$. However, nilsequences as we have defined them are not closed under uniform limits. This leads to a slight conflict between the nomenclature of the present paper and that of (for example) [6]. In that paper the objects we have called nilsequences are referred to as *basic* nilsequences; a nilsequence is then a uniform limit of basic nilsequences. Since our analysis is essentially finitary in nature we will not make any further mention of this distinction.

The inverse Gowers-norm conjecture. An important feature of s-step nilmanifolds is that they have significant "constraints" connecting arithmetic progressions of length $s+2$, or cubes of dimension $s+1$. Roughly speaking, given the first $s+1$ elements $x, g \cdot x, g^2 \cdot x, \ldots, g^s \cdot x$ of a progression in an s-step nilmanifold G/Γ, the next element $g^{s+1} \cdot x$ of the progression and all further elements are essentially completely determined as continuous functions of these first $s+1$ elements. For a precise formulation of this assertion see [26, Lemma 12.7]. Similarly, when considering an s-dimensional "cube" $\{g_1^{\omega_1} \ldots g_s^{\omega_s} \cdot x : (\omega_1, \ldots, \omega_s) \in \{0,1\}^s\}$ in G/Γ, the final vertex $g_1 \ldots g_s \cdot x$ of this cube is essentially a continuous function of the other 2^{s-1} elements of this cube. See Appendix E for more precise formulations of this statement, which we will make heavy use of in this paper. As a consequence of either of these facts, we can relate nilsequences to the U^{s+1} norm. The next result is in this direction, but it is not sufficiently general for our later applications. We state it now to introduce the concept of nilsequences obstructing uniformity, and because it can be proved using earlier results.

Proposition 8.2 (Nilsequences obstruct uniformity)**.** *Let $s \geqslant 1$ be an integer and let $\delta \in (0,1)$ be real. Let $G/\Gamma = (G/\Gamma, d_{G/\Gamma})$ be an s-step nilmanifold with some fixed smooth metric $d_{G/\Gamma}$, and let $(F(g^n x))_{n \in \mathbb{N}}$ be a bounded s-step nilsequence with Lipschitz constant at most M. Let $f : [N] \to [-1,1]$ be a function for which*

$$\mathbb{E}_{n \in [N]} f(n) F(g^n \cdot x) \geqslant \delta.$$

Then we have

$$\|f\|_{U^{s+1}[N]} \gg_{s,\delta,M,G/\Gamma} 1.$$

Proof. See [26, Prop. 12.6]. The lower bound arising in that proposition was stated to depend on the continuous function $F : G/\Gamma \to \mathbb{C}$, and not just on $\|F\|_{\mathrm{Lip}}$. However, an examination of the proof reveals that the argument can be made uniform in F, for a given value of $\|F\|_{\mathrm{Lip}}$. □

Remark. It turns out that one can relax the assumption that f be uniformly bounded, requiring only that f be bounded in L^1 norm; see Corollary 11.6.

The inverse Gowers-norm conjecture is an assertion in the converse direction, that nilsequences are the *only* obstruction to uniformity. More precisely, we have for each $s \geqslant 1$ the following conjecture:

Conjecture 8.3 (GI(s) conjecture)**.** *Suppose that $0 < \delta \leqslant 1$. Then there exists a finite collection $\mathcal{M}_{s,\delta}$ of s-step nilmanifolds $G/\Gamma = (G/\Gamma, d_{G/\Gamma})$ with the following property. Given any N and any $f : [N] \to [-1,1]$ such that*

$$\|f\|_{U^{s+1}[N]} \geqslant \delta,$$

there is a nilmanifold $G/\Gamma \in \mathcal{M}_{s,\delta}$ and a 1-bounded s-step nilsequence $(F(g^n x))_{n\in\mathbb{N}}$ on it with Lipschitz constant $O_{s,\delta}(1)$, such that

$$|\mathbb{E}_{n\in[N]} f(n)F(g^n x)| \gg_{s,\delta} 1.$$

This conjecture in this form is due to the authors. It was hinted at in [26, §13] and is being stated formally for the first time here. The evidence in favour of it is strong. First of all we know that the cases $s = 1, 2$ are true. The case $s = 1$ is an exercise in harmonic analysis. Indeed in this case one can take G/Γ to just be the standard unit circle $\mathbb{R}/\mathbb{Z}$, so that $\mathcal{M}_{1,\delta}$ is a singleton set independent of δ. The case $s = 2$ was established, with some effort, in [26] and is stated in Proposition 8.4 below. Note that things are not so simple when $s > 1$, and it is known that as δ decreases to zero, the collection $\mathcal{M}_{s,\delta}$ of nilmanifolds G/Γ that one must employ must have cardinality going to infinity[8].

Proposition 8.4 (The GI(2) conjecture, [26])**.** *The* GI(2) *conjecture holds in the form stated above. In fact the group G may be taken to be a product of $O(\delta^{-O(1)})$ Heisenberg groups $\begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}$, and the discrete cocompact subgroup Γ may be taken to be a product of copies of $\begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$.*

Proof. This is almost [26, Thm. 12.8]. In that theorem, a nilsequence was constructed in a somewhat *ad hoc* manner from another type of object, a generalised quadratic phase. In the argument of that paper, however, the nilpotent groups constructed were not all Heisenberg groups. Some of them were isomorphic to $\mathbb{R}^2 \times \mathbb{Z}$, which is not connected and hence, with our definition, cannot be used to construct a nilmanifold.

More precisely, in the proof of [26, Thm. 12.8] it is shown that if $\|f\|_{U^3} \geqslant \delta$ then

$$\left|\mathbb{E}_{n\in[N]} f(n)F_1(g^n x)e(n^2\theta)\right| \gg \exp(-\delta^{-O(1)}),$$

where $F_1(g^n x)$ is a product of nilsequences coming from $O(\delta^{-O(1)})$ Heisenberg groups (which are all connected and simply-connected), $\theta \in \mathbb{R}/\mathbb{Z}$, and $e(x) := e^{2\pi i x}$. In [26, Thm. 12.8] we proceeded by constructing $e(n^2\theta)$ as a nilsequence coming from a skew torus which, being a quotient of the disconnected nilpotent Lie group $\begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$, is not immediately helpful in the present context. However we might just as easily have observed that

$$\begin{pmatrix} 1 & -\theta & -\theta \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}^n = \begin{pmatrix} 1 & -n\theta & -n^2\theta \\ 0 & 1 & 2n \\ 0 & 0 & 1 \end{pmatrix}$$

which, upon quotienting by the right action of $\begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$, leads to

$$\left[\begin{pmatrix} 1 & -\theta & -\theta \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}^n\right] = \left[\begin{pmatrix} 1 & \{-n\theta\} & \{n^2\theta\} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}\right].$$

[8]This seems to be related to the fact, known to the ergodic theorists, that the inverse limit of 1-step nilsystems is a 1-step nilsystem, but the same is not true for s-step nilsystems, $s \geqslant 2$.

Here we have moved our matrix under the right action of Γ so that it lies in the fundamental domain

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

see [27] for further discussion. The fractional parts $\{t\}$ are chosen to lie in $(-\frac{1}{2}, \frac{1}{2}]$.

This almost exhibits $e(n^2\theta)$ as a nilsequence coming from the Heisenberg group, but there is one small problem: the function

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

from $\mathcal{F}$ to $\mathbb{C}$ does not extend to a continuous function on G/Γ, since there are discontinuities on the boundary $\partial\mathcal{F}$.

To get around this one may introduce a smooth partition of unity $(\chi_j)_{j\in J}$ on $(\mathbb{R}/\mathbb{Z})^2$, where each function χ_j is supported on (say) a square of width $1/100$. Each function

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

does extend to a Lipschitz function on G/Γ. This makes it clear that $e(n^2\theta)$ may, after all, be realised as a nilsequence coming from a product of $O(1)$ Heisenberg groups. □

For higher values of s, the conjecture $\mathrm{GI}(s)$ remains open. However, significant support in favour of this conjecture arises from the combinatorial and Fourier-analytic work of Gowers [21], in which a "local" form of this conjecture was established in order to provide a new proof of Szemerédi's theorem. Further substantial support for the conjecture comes from the ergodic-theoretic work of Host-Kra [32].

The Möbius and nilsequences conjecture. Our main results are concerned with the von Mangoldt function $\Lambda(n)$ and with functions derived from Λ, such as $\Lambda'_{b,W}$. It turns out, however, to be convenient to rewrite this function in terms of the closely related *Möbius function* $\mu : \mathbb{Z} \to \{-1, 0, +1\}$, defined by setting $\mu(n) := (-1)^d$ when n is the product of d distinct primes, and $\mu(n) = 0$ otherwise. The main advantage of doing so is that μ is a 1-bounded function, whereas Λ patently is not. As is well known, Λ and μ are related by the identity

$$\Lambda(n) = \sum_{d|n} \mu(d) \log \frac{n}{d} = -\sum_{d|n} \mu(d) \log d \tag{8.1}$$

for all $n \geqslant 1$. In principle this allows us to reduce the task of estimating correlations involving Λ to that of estimating correlations involving μ, although when doing so the unbounded weight $\log \frac{n}{d}$ and the summation over d will introduce some dangerous factors of $O(\log N)$ which must be handled with some caution.

Suppose we formally apply Conjecture 8.3 to the task of proving Theorem 7.2, ignoring for now the significant issue that $\Lambda'_{b,W} - 1$ is not uniformly bounded. Then we expect to reduce this theorem to the assertion that $\Lambda'_{b,W} - 1$ has small correlation with any s-step nilsequence. In the light of (8.1), we expect this statement to be related to the

corresponding assertion for the Möbius function μ. We formalise this latter statement as the following conjecture.

Conjecture 8.5 (MN(s) conjecture)**.** *Let $G/\Gamma = (G/\Gamma, d_{G/\Gamma})$ be an s-step nilmanifold with smooth metric $d_{G/\Gamma}$, and let $(F(g^n x))_{n\in[N]}$ be a bounded s-step nilsequence with Lipschitz constant M. Then we have the bound*

$$\left|\mathbb{E}_{n\leqslant N}\mu(n)F(g^n x)\right| \ll_{A,M,G/\Gamma,s} \log^{-A} N$$

for any real number $A > 0$.

Remark. It is important to note that the implied constant is *not* allowed to depend on g and x. The case $s = 1$ can be reduced to a classical result of Davenport [10]; see [27, §6] for details. The case $s = 2$ was the main result of [27]. The case $s > 2$ remains open; however, we certainly expect MN(s) to be true in this case because of the *Möbius randomness* heuristic from analytic number theory, which states that μ exhibits a substantial degree of orthogonality to any suitably "Lipschitz" function. Moreover, it seems likely that the techniques we developed to prove MN(2) will eventually extend to cover MN(s), $s \geqslant 3$, as well. This is another ongoing area of research. As is well known, even when $s = 1$ the current technology for establishing this conjecture yields ineffective implied constants in the $\ll_{A,M,G/\Gamma}$ due to our lack of knowledge regarding the existence of Siegel zeroes. This ultimately makes the decay rates in the Main Theorem (and its corollaries) similarly ineffective. If the GRH is assumed, the estimates do become effective. However they are still somewhat poor for $s \geqslant 2$, largely because the bounds in the GI(2) conjecture obtained in [26] are a little weaker than one might hope for.

9. Correlation estimates for Möbius and Liouville

Perhaps the heart of the present paper is §10, in which it is shown how, in certain circumstances, the requirement of 1-boundedness can be dropped in the GI(s) conjecture. This section is an aside to the main line of our argument, in which we use what we already have to obtain estimates similar to the generalised Hardy-Littlewood conjecture for the Möbius function and the related *Liouville function* $\lambda : \mathbb{N} \to \{-1, +1\}$, defined to be the unique completely multiplicative function such that $\lambda(p) = -1$ for all primes p.

Proposition 9.1 (Correlation estimates for μ and λ)**.** *Let d, t, L be positive integers, let N be a large positive integer parameter, and let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms with size $\|\Psi\|_N \leqslant L$ and complexity at most s. Assume the* GI(s) *and* MN(s) *conjectures. Let $K \subset [-N, N]^d$ be a convex body. Then we have*

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]}\mu(\psi_i(n)) = o_{s,t,d,L}(N^d) \tag{9.1}$$

and

$$\sum_{n\in K\cap\mathbb{Z}^d}\prod_{i\in[t]}\lambda(\psi_i(n)) = o_{s,t,d,L}(N^d). \tag{9.2}$$

Remark. Note the lack of any local factors β_p, β_∞. This makes Proposition 9.1 rather appealing from a certain point of view. It also provides an instance of the "Möbius randomness heuristic" alluded to above.

Proof. We begin by applying Proposition 7.1, the generalised von Neumann theorem. Since μ and λ are 1-bounded, this may be applied with the pseudorandom measure ν set equal to the constant function 1, which is obviously D-pseudorandom for all D. We note that in this case the proof of Proposition 7.1 that we give in Appendix C is rather simpler than in the case of a more general ν; specifically, one can use Corollary B.3 in place of Corollary B.4, while the verification of (C.10), (C.11) is trivial when $\nu = 1$.

The application of Proposition 7.1 reduces (9.1) to the statement

$$\|\mu\|_{U^{s+1}[N]} = o_s(1). \tag{9.3}$$

Applying the GI(s) conjecture, it is sufficient to establish that

$$\mathbb{E}_{n\leqslant N}\mu(n)F(g^n x) = o_{s,M,\delta}(1) \tag{9.4}$$

unifromly over all $G/\Gamma \in \mathcal{M}_{s,\delta}$ and all 1-bounded M-Lipschitz nilsequences $(F(g^n x))_{n\leqslant N}$ on G/Γ. Indeed the truth of such a statement implies, by the GI(s) conjecture, that $\|\mu\|_{U^{s+1}[N]} \leqslant \delta$, and one may then take δ arbitrarily small to deduce (9.3). Recalling that $|\mathcal{M}_{s,\delta}| = O_{\delta,s}(1)$, we see that (9.4) follows immediately from (a weak form of) the MN(s) conjecture. This proves (9.1).

The proof of (9.2) proceeds similarly. It suffices to establish the analogue of (9.4), that is to say the bound

$$\mathbb{E}_{n\leqslant N}\lambda(n)F(g^n x) = o_{s,M,\delta}(1) \tag{9.5}$$

uniformly over all $G/\Gamma \in \mathcal{M}_{s,\delta}$ and all 1-bounded M-Lipschitz nilsequences $(F(g^n x))_{n\leqslant N}$ on G/Γ. We begin by noting the identity

$$\lambda(n) := \sum_{d^2|n} \mu(\frac{n}{d^2}).$$

This implies that for any positive real X, any fixed $G/\Gamma \in \mathcal{M}_{s,\delta}$ and any 1-bounded M-Lipschitz nilsequence $(F(g^n x))_{n\leqslant N}$ on G/Γ we have

$$\begin{aligned}\mathbb{E}_{n\leqslant N}\lambda(n)F(g^n x) &= \mathbb{E}_{n\leqslant N}\sum_{d^2|n}\mu(\frac{n}{d^2})F(g^n x)\\ &= \sum_{d\leqslant X}\mathbb{E}_{n\leqslant N}1_{d^2|n}\mu(\frac{n}{d^2})F(g^n x) + \sum_{d>X}\mathbb{E}_{n\leqslant N}1_{d^2|n}\mu(\frac{n}{d^2})F(g^n x)\\ &= \sum_{d\leqslant X}\mathbb{E}_{k\leqslant N/d^2}\mu(k)F(g^{d^2 k}x) + O(X^{-1}).\end{aligned} \tag{9.6}$$

By replacing g by g^{d^2} in the MN(s) conjecture we obtain the bound

$$\mathbb{E}_{k\leqslant N/d^2}\mu(k)F(g^{d^2 k}x) = o_{G/\Gamma,M,d}(1).$$

Substituting into (9.6) we obtain

$$\mathbb{E}_{n\leqslant N}\lambda(n)F(g^n x) = o_{G/\Gamma,M,X}(1) + O(X^{-1}).$$

Let $\varepsilon > 0$ be arbitrary. Taking $X := 1/\varepsilon$, we may make this expression smaller than a constant times ε by taking N sufficiently large. This implies that

$$\mathbb{E}_{n\leqslant N}\lambda(n)F(g^n x) = o_{G/\Gamma,M}(1).$$

Recalling once more that $|\mathcal{M}_{s,\delta}| = O_{s,\delta}(1)$, we therefore obtain (9.5) and hence (9.2).□

Let us remark that, as with the Main Theorem, Proposition 9.1 is unconditional in the cases $s = 1, 2$.

We conclude with a mention of a conjecture of Chowla [8], which asserts that λ is uniformly distributed on any polynomial, thus for instance

$$\mathbb{E}_{y_1,y_2 \leqslant N} \lambda(P(y_1, y_2)) = o_P(1) \tag{9.7}$$

for any polynomial $P : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ of two variables. Our results imply (for instance) the following case of Chowla's conjecture.

Proposition 9.2. *Let $P : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be a polynomial of degree at most 4 which is the product of homogeneous linear factors over $\mathbb{Q}$, and which is not a rational multiple of a perfect square. Then we have*

$$\mathbb{E}_{y_1,y_2 \leqslant N} \lambda(P(y_1, y_2)) = o_P(1).$$

The proof is immediate from (9.2) and the complete multiplicativity of λ; note that we can easily eliminate any repeated factors in P and so the system of linear forms associated to P will be non-degenerate. We remark that this conjecture was also recently verified for all homogeneous polynomials of degree at most three in [29, 30]. Removing the homogeneity assumption looks hopeless with current technology; the case $P(y_1, y_2) = y_1(y_1 + 2)$ is already roughly of the same order of difficulty as the twin prime conjecture.

10. Transferring the inverse Gowers-norm conjecture

Recall that we are trying to use the inverse Gowers-norm and Möbius and nilsequences conjectures to prove Theorem 7.2. We cannot apply the Gowers Inverse conjecture directly to prove Theorem 7.2, because $\Lambda'_{b,W} - 1$ is not bounded uniformly in N. The difficulty here is similar to that encountered in [24], in which Szemerédi's theorem, which ostensibly only establishes multiple recurrence bounds for bounded functions, needed to be extended to an unbounded function such as $\Lambda'_{1,W}$. We will use a similar resolution to that in [24], namely to *transfer* the inverse Gowers-norm conjecture to the situation of a function bounded by a pseudorandom measure. More precisely, the purpose of this section is to prove the following result.

Proposition 10.1 (Relative inverse Gowers-norm conjecture)**.** *Assume the* GI(s) *conjecture. For any $0 < \delta \leqslant 1$ and any $C \geqslant 20$, there exists a finite collection $\mathcal{M}_{s,\delta,C}$ of nilmanifolds $G/\Gamma = (G/\Gamma, d_{G/\Gamma})$ with the following property. Let $N \geqslant 1$, suppose that $N' \in [CN, 2CN]$ is a prime, that $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ is an $(s+2)2^{s+1}$-pseudorandom measure, that $f : [N] \to \mathbb{R}$ is a function with $|f(n)| \leqslant \nu(n)$ for all $n \in [N]$ and that $\|f\|_{U^{s+1}[N]} \geqslant \delta$. Then there exists $G/\Gamma \in \mathcal{M}_{s,\delta,C}$ together with a 1-bounded s-step nilsequence $(F(g^n x))_{n\in\mathbb{Z}}$ with Lipschitz constant $O_{s,\delta,C}(1)$, such that*

$$|\mathbb{E}_{n \leqslant N} f(n) F(g^n x)| \gg_{s,C,\delta} 1.$$

Remarks. This looks significantly more complicated than the ordinary GI(s) conjecture, but this is something of an illusion. Most of the complexity comes from the need for

the additional dependence on C. A largeish value of C might be required in order to construct an appropriate pseudorandom measure ν on $\mathbb{Z}_{N'}$ (cf. Proposition 6.4) and so we leave C unspecified in this proposition.

In view of Proposition 10.1 and Proposition 6.4, it is not hard to see that Theorem 7.2, and hence the Main Theorem, follows from the next proposition. All one need do is choose $C := \max(C_0((s+2)2^{s+1}), 20)$, where C_0 is the function appearing in Proposition 6.4. This ensures that an appropriate pseudorandom measure ν can be constructed.

Proposition 10.2 (W-tricked von Mangoldt orthogonal to nilsequences)**.** *Let $s \geqslant 1$, and assume the* MN(s) *conjecture. Let $G/\Gamma = (G/\Gamma, d_{G/\Gamma})$ be an s-step nilmanifold with smooth metric $d_{G/\Gamma}$, and let $(F(g^n x))_{n\in[N]}$ be a bounded s-step nilsequence with Lipschitz constant M. Let $b \in [W]$ be coprime to W. Then we have the bound*

$$\mathbb{E}_{n\in[N]}(\Lambda'_{b,W}(n) - 1)F(g^n x) = o_{M,G/\Gamma,s}(1).$$

Remark. In principle, Proposition 10.2 is substantially easier to establish than the preceding reductions of the Main Theorem, such as Theorem 7.2. This is because we are now computing the correlation of Λ (or $\Lambda'_{b,W} - 1$) with respect to a "low complexity" sequence $F(g^n x)$, rather than the more complicated task of computing a multilinear correlation of Λ with *itself*. In particular one can now hope to use tools such as Vinogradov's method to establish this proposition. Indeed, the computation of exponential sums such as $\sum_{n\in[N]} \Lambda(n)e(\alpha n)$, or more generally $\sum_{n\in[N]} \Lambda(n)e(\alpha n^k)$, are essentially model cases of Proposition 10.2 and are well-known to be treatable by Vinogradov's method. However, Proposition 10.2 is somewhat more general as it also (for example) asserts some control on generalised polynomial exponential sums such as $\sum_{n\in[N]} \Lambda(n)e(\alpha n\lfloor\beta n\rfloor)$, where $\lfloor\cdot\rfloor$ is the greatest integer function. See [27] for further discussion of the link between such functions and 2-step nilsequences. Thus we see that the inverse Gowers-norm conjecture GI(s) is a powerful tool for establishing bounds on the Gowers norms U^{s+1}, and thence to all multilinear averages of complexity at most s.

We prove Proposition 10.2 in later sections. For the remainder of this section we derive Proposition 10.1 from the inverse Gowers-norm conjecture.

A Koopman-von Neumann theorem.[9] The primary tool in deducing Proposition 10.1 from the Gowers Inverse conjecture is the following structure theorem, which allows us to decompose an arbitrary function f which is bounded pointwise by ν into a bounded function and a Gowers-uniform function.

Proposition 10.3 (Koopman – von Neumann theorem)**.** *Let $s \geqslant 1$ and let $N' \geqslant N \geqslant 1$ be an integer. Suppose that ν is an $(s+2)2^{s+1}$-pseudorandom measure on $\mathbb{Z}_{N'}$, and that $f : \mathbb{Z}_{N'} \to \mathbb{R}$ is a function such that $|f(n)| \leqslant \nu(n)$ pointwise. Then we may decompose $f = f_1 + f_2$, where*

$$\sup_{n\in\mathbb{Z}_{N'}} |f_1(n)| \leqslant 1 \tag{10.1}$$

[9]This term has something in common with the term "generalised von Neumann theorem" in that it originally came from analogies with ergodic theory. We now use it in our work to describe a range of theorems whose general aim is to decompose a given function f into the sum of a function f_1 which is somehow less complicated than f, together with an error f_2 which is small in some Gowers norm.

and

$$\|f_2\|_{U^{s+1}(\mathbb{Z}_{N'})} = o(1). \tag{10.2}$$

If furthermore f is supported in $\{-N, \dots, N\}$ for some $N < N'/10$, then we may arrange matters so that f_1 and f_2 are both supported on $\{-2N, \dots, 2N\}$.

Remark. Informally, this theorem asserts that in the U^{s+1} topology, bounded functions are dense in the class of functions bounded by ν. This fact (and refinements thereof), in conjunction with generalised von Neumann theorems such as Proposition 7.1, underlie the "transference principle" from [24] which allow one to convert results for multilinear averages of 1-bounded functions to results for multilinear averages of functions bounded by a pseudorandom measure. This principle is essential for our arguments here, as it allows us in many cases to manipulate functions such as $\Lambda_{b,W}$ as if they were uniformly bounded.

Proof. Let us first make the observation that we can weaken (10.1) to

$$\sup_{n \in \mathbb{Z}_{N'}} |f_1(n)| \leqslant 1 + o(1) \tag{10.3}$$

since one could simply transfer the $o(1)$ error in (10.3) to the f_2 component afterwards, using the triangle inequality on (10.2).

We shall rely heavily on a similar result from [24, Proposition 8.1]. Before we give this result we need some notation.

Definition 10.4 (Conditional expectation)**.** If $f : \mathbb{Z}_{N'} \to \mathbb{R}$ is a function and $1 \leqslant p \leqslant \infty$, we denote $\|f\|_{L^p(\mathbb{Z}_{N'})} := (\mathbb{E}_{n \in \mathbb{Z}_{N'}} |f(n)|^p)^{1/p}$, with the usual convention that $\|f\|_{L^\infty(\mathbb{Z}_{N'})} := \sup_{n \in \mathbb{Z}_{N'}} |f(n)|$. If $\mathcal{B}$ is a σ-algebra on $\mathbb{Z}_{N'}$, that is to say the Boolean algebra generated by the atoms of a partition of $\mathbb{Z}_{N'}$, we define the *conditional expectation* $\mathbb{E}(f|\mathcal{B})$ of f relative to $\mathcal{B}$ to be the orthogonal projection in $L^2(\mathbb{Z}_{N'})$ from f to the $\mathcal{B}$-measurable functions.

In our current notation, Proposition 8.1 from [24] asserts[10] the following.

Proposition 8.1 of [24]**.** *Suppose that $N' \geqslant N$ and that $\nu : \mathbb{Z}_{N'} \to \mathbb{R}_{\geqslant 0}$ is an $(s+2)2^{s+1}$-pseudorandom measure. Let $f : \mathbb{Z}_{N'} \to \mathbb{R}$ be such that $|f(n)| \leqslant \nu(n)$ for all $n \in \mathbb{Z}_{N'}$. Let $\varepsilon \in (0, 1)$ be a small parameter, and assume N' is sufficiently large depending on ε. Then there exists a σ-algebra $\mathcal{B}$ and an exceptional set $\Omega \in \mathcal{B}$ such that*

- (smallness condition)
$$\mathbb{E}_{\mathbb{Z}_{N'}}(\nu 1_\Omega) = o_\varepsilon(1); \tag{10.4}$$
- (ν is uniformly distributed outside of Ω)
$$\|(1 - 1_\Omega)\mathbb{E}(\nu - 1|\mathcal{B})\|_{L^\infty(\mathbb{Z}_{N'})} = o_\varepsilon(1) \tag{10.5}$$

and

[10] In [24] the result is only stated when $0 \leqslant f(n) \leqslant \nu(n)$, but exactly the same proof applies under the more general assumption that $|f(n)| \leqslant \nu(n)$. In any case, in order to prove Proposition 10.3 one could always decompose f into non-negative and negative parts $f^+ + f^-$ and follow the proof for each part separately. The key point to note is that the function f_1^+ is non-negative, whilst $f_1^- \leqslant 0$. Thus $f_1 = f_1^+ + f_1^-$ satisfies the requisite L^∞ bound (10.3).

- (Gowers uniformity estimate)

$$\|(1-1_\Omega)(f-\mathbb{E}(f|\mathcal{B}))\|_{U^{s+1}(\mathbb{Z}_{N'})} \leqslant \varepsilon^{1/2^{s+2}} = \kappa_s(\epsilon). \tag{10.6}$$

Let ε be chosen later (it will eventually be a slowly decaying function of N). If N is sufficiently large depending on ε, we can invoke the above theorem. Write

$$f = f_1 + f_2 = f_1 + f_2^{(1)} + f_2^{(2)},$$

where

$$f_1 := (1-1_\Omega)\mathbb{E}(f|\mathcal{B}),$$
$$f_2^{(1)} := (1-1_\Omega)(f-\mathbb{E}(f|\mathcal{B}))$$

and

$$f_2^{(2)} := 1_\Omega f.$$

Then by (10.5) we have

$$\|f_1\|_{L^\infty(\mathbb{Z}_{N'})} \leqslant 1 + o_\varepsilon(1). \tag{10.7}$$

Also, by (10.6) we have

$$\|f_2^{(1)}\|_{U^{s+1}(\mathbb{Z}_{N'})} = \kappa_s(\epsilon). \tag{10.8}$$

Next, we claim that

$$\|f_2^{(2)}\|_{U^{s+1}(\mathbb{Z}_{N'})} = o_\varepsilon(1). \tag{10.9}$$

To see this, first note that from (10.4) we have

$$\|f_2^{(2)}\|_{L^1(\mathbb{Z}_{N'})} = o_\varepsilon(1). \tag{10.10}$$

Secondly, we prove that for functions g for which $|g|$ is bounded pointwise by a pseudorandom measure ν, the $L^1(\mathbb{Z}_{N'})$ norm controls the $U^{s+1}(\mathbb{Z}_{N'})$-norm. Indeed for such a function we have

$$\begin{aligned}\|g\|_{U^{s+1}(\mathbb{Z}_{N'})}^{2^{s+1}} &= \mathbb{E}_{n\in\mathbb{Z}_{N'},h\in\mathbb{Z}_{N'}^{s+1}} g(n)\prod_{\substack{\omega\in\{0,1\}^{s+1}\\ \omega\neq 0}} g(n+\omega\cdot h)\\ &\leqslant \mathbb{E}_{n\in\mathbb{Z}_{N'}}|g(n)|\sup_n \mathbb{E}_{h\in\mathbb{Z}_{N'}^{s+1}}\prod_{\substack{\omega\in\{0,1\}^{s+1}\\ \omega\neq 0}}\nu(n+\omega\cdot h)\\ &= \|\mathcal{D}\nu\|_{L^\infty(\mathbb{Z}_{N'})}\|g\|_{L^1(\mathbb{Z}_{N'})},\end{aligned}$$

where

$$\mathcal{D}\nu(n) := \prod_{\substack{\omega\in\{0,1\}^{s+1}\\ \omega\neq 0}}\nu(n+\omega\cdot h)$$

is the *dual function* associated to ν. However a simple application of the linear forms condition, given in detail in [24, Lemma 6.1], confirms that

$$\|\mathcal{D}\nu\|_{L^\infty(\mathbb{Z}_{N'})} \leqslant 1 + o(1).$$

This concludes the proof of (10.9). From this, (10.8), and the triangle inequality for the $U^{s+1}(\mathbb{Z}_{N'})$ norm we conclude that

$$\|f_2\|_{U^{s+1}(\mathbb{Z}_{N'})} \leqslant o_\varepsilon(1) + \kappa(\epsilon).$$

Choosing ε to be a sufficiently slowly decaying function of N we obtain the first part of Proposition 10.3.

It remains to deal with the situation where f is supported[11] in $\{-N,\dots,N\}$. We can write $f(n) = f(n)\psi(n)$, where $\psi : \mathbb{Z}_{N'} \to [0,1]$ equals 1 on $\{-N,\dots,N\}$, vanishes outside of $\{-2N,\dots,2N\}$ and interpolates smoothly in the range $N \leqslant |n| \leqslant 2N$. One could, for example, take ψ to be a de la Vallée Poussin kernel. If $f = f_1 + f_2$ is the previous decomposition, then upon multiplying by ψ we obtain $f = \tilde{f}_1 + \tilde{f}_2$, where $\tilde{f}_1 := f_1\psi$ and $\tilde{f}_2 := f_2\psi$. The function $\tilde{f}_1$ continues to enjoy the bound (10.3) but now also has the desired support property. To confirm that $\tilde{f}_2$ enjoys the bound (10.2), simply use Fourier series to break ψ up as a rapidly convergent linear combination of linear phases $e(n\xi/N)$, and use the triangle inequality combined with the phase invariance (B.11) of the U^{s+1} norm. This concludes the proof of Proposition 10.3. □

Proof of Proposition 10.1. Suppose that $N' \in [CN, 2CN]$ is prime, that $\nu : \mathbb{Z}_{N'} \to \mathbb{R}$ is an $(s+2)2^{s+1}$-pseudorandom measure, that $f : [N] \to \mathbb{R}$ is a function with $|f(n)| \leqslant \nu(n)$ for all $n \in [N]$ and that $\|f\|_{U^{s+1}[N]} \geqslant \delta$. Applying Proposition 10.3 we may decompose

$$f = f_1 + f_2,$$

where $\|f_1\|_{L^\infty(\mathbb{Z}_{N'})} \leqslant 1$ and $\|f_2\|_{U^{s+1}(\mathbb{Z}_{N'})} = o(1)$. Since $C > 10$, we may further assume that both f_1 and f_2 are supported in $\{-2N,\dots,2N\}$. By Lemma B.5 the assumption that $\|f\|_{U^{s+1}[N]} \geqslant \delta$ implies that $\|f\|_{U^{s+1}(\mathbb{Z}_{N'})} \gg_{C,s} \delta$, and hence that $\|f_1\|_{U^{s+1}(\mathbb{Z}_{N'})} \gg_{C,s} \delta$. Applying Lemma B.5 once more, we conclude that $\|f_1\|_{U^{s+1}(\{-2N,\dots,2N\})} \gg_{C,s} \delta$.

We now apply the inverse Gowers-norm conjecture GI(s), translating $\{-2N,\dots,2N\}$ to the interval $[4N+1]$, to conclude that there exists an s-step nilmanfold $G/\Gamma = (G/\Gamma, d_{G/\Gamma})$ from a fixed finite collection $G/\Gamma \in \mathcal{M}_{s,\delta,C}$, together with a bounded s-step nilsequence $(F(g^n x))_{n\in\mathbb{N}}$ generated by this nilmanifold and with Lipschitz constant $O_{s,\delta,C}(1)$, such that

$$|\mathbb{E}_{-2N\leqslant n\leqslant 2N} f_1(n)F(g^n x)| \gg_{s,\delta,C} 1.$$

On the other hand, from (10.2) and the contrapositive of Proposition 8.2 we have

$$|\mathbb{E}_{-2N\leqslant n\leqslant 2N} f_2(n)F(g^n x)| = o_{G/\Gamma,s,\delta,C}(1).$$

If $N \geqslant N_0(s,\delta,C)$ is large depending on s, δ and C, we conclude that

$$|\mathbb{E}_{-2N\leqslant n\leqslant 2N} f(n)F(g^n x)| \gg_{s,\delta,C} 1$$

and the claim follows (since f is supported on $[N]$).

If by contrast $N = O_{s,\delta,C}(1)$ then the claim is trivial, since all norms on $[N]$ are then equivalent up to factors of $O_N(1) = O_{s,\delta,C}(1)$, and all functions on $[N]$ can be expressed as nilsequences (say on the torus $\mathbb{R}/\mathbb{Z}$) with Lipschitz constant $O_N(1) = O_{s,\delta,C}(1)$. □

[11] An alternate way to proceed at this point is to modify the proof of [24, Proposition 8.1], where the σ-algebra $\mathcal{B}$ is initialised not at the trivial factor, but rather at the factor generated by $\{-N,\dots,N\}$.

11. Averaging the nilsequence

To summarise so far, we have reduced the task of showing that the GI(s) conjecture implies the Main Theorem to the much easier task of establishing Proposition 10.2. This, recall, is an estimate on the correlation between the number-theoretic function $\Lambda'_{b,W}(n) - 1$ and the nilsequence $F(g^n x)$.

The purpose of this section is to perform a rather technical modification to the nilsequence $F(g^n x)$, which is necessary for the following reason. At a later stage in the proof we would like to discard certain "small" components of the function $\Lambda'_{b,W}(n) - 1$ from this correlation. Some of these components will be easy to discard; for instance, any error which is small in L^1 norm will be easily removed since the nilsequence is bounded. However, there will be one component of $\Lambda'_{b,W}(n) - 1$ that we shall encounter (namely, the term arising from the "smooth" component $\Lambda^\sharp$ of the von Mangoldt function) which will not be small in L^1, but is instead small in the Gowers norm $U^{s+1}[N]$. In principle, Proposition 8.2 or Corollary 11.6 would allow us to safely drop such terms. Unfortunately, a problem arises because the component of $\Lambda'_{b,W}(n) - 1$ that we are trying to discard is not bounded, and we have also not been able to dominate this component by a pseudorandom measure or even to establish a bound for it in L^1. To get around this problem, we need to improve the "regularity" of the nilsequence $F(g^n x)$. In particular we must convert it to an object which we can bound in the dual norm $U^{s+1}[N]^*$, defined as usual by the formula

$$\|F\|_{U^{s+1}[N]^*} := \sup\{|\mathbb{E}_{n\in[N]} f(n)F(n)| : \|f\|_{U^{s+1}[N]} \leqslant 1\}.$$

This dual norm also appeared in [24], and plays a similar rôle there as it does here.

It would be very pleasant if every s-step nilsequence was automatically bounded in the $U^{s+1}[N]^*$ norm. Unfortunately, this statement is false even in the $s = 1$ case, as in that case it amounts to a certain $l^{4/3}$ summability estimate on the Fourier coefficients of Lipschitz functions on a compact abelian group. There is no such estimate if the group is of sufficiently high dimension. Of course one can rectify this by replacing the Lipschitz functions with smooth functions. It seems likely that a similar claim is true for higher s, but it also seems likely that a proof would involve a finer analysis of the structure of nilmanifolds than we need for the rest of our argument.

Fortunately, however, we can achieve an adequate substitute result by replacing the concept of a nilsequence by its convex hull. Definition 11.1 provides a precise definition.

Definition 11.1 (Averaged nilsequences)**.** Let $G/\Gamma = (G/\Gamma, d_{G/\Gamma})$ be an s-step nilmanifold, and let $M > 0$. An s-step *averaged nilsequence* on G/Γ with Lipschitz constant at most M is a function $F(n)$ having the form

$$F(n) = \mathbb{E}_{i\in I} F_i(g_i^n x_i),$$

where I is some finite index set, and for each i, $F_i(g_i^n x_i)$ is a bounded s-step nilsequence on G/Γ with Lipschitz constant at most M.

Remark. An averaged nilsequence of the type just described is a genuine nilsequence on the nilmanifold $(G/\Gamma)^I$. However the averaging set I will, in applications, have size

comparable to N and so in our finitary world these averaged nilsequences should be thought of as a strict generalisation of the notion of a nilsequence. Were it not for the desire to avoid issues of measurablility, we might even have replaced the finite averaging operator $\mathbb{E}_{i\in I}$ by an integration over a suitable probability space.

We now state the crucial technical lemma we need, which allows us to replace a nilsequence by an averaged nilsequence with a good $U^{s+1}[N]^*$ bound.

Proposition 11.2 (Decomposition of nilsequences)**.** *Let $G/\Gamma = (G/\Gamma, d_{G/\Gamma})$ be an s-step nilmanifold, and let $M > 0$. Suppose that $(F(g^n x))_{n\in\mathbb{N}}$ is a bounded s-step nilsequence on G/Γ with Lipschitz constant at most M. Let $\varepsilon \in (0,1)$ and suppose that $N \geqslant 1$. Then we may effect the decomposition*

$$F(g^n x) = F_1(n) + F_2(n), \tag{11.1}$$

where $F_1 : \mathbb{N} \to [-1,1]$ is an averaged nilsequence on $(G/\Gamma)^{2^{s+1}-1}$ with Lipschitz constant $O_{M,\varepsilon,G/\Gamma}(1)$ and obeying the dual norm bound

$$\|F_1\|_{U^{s+1}[N]^*} \ll_{M,\varepsilon,G/\Gamma} 1, \tag{11.2}$$

while $F_2 : \mathbb{N} \to \mathbb{R}$ obeys the uniform bound

$$\|F_2\|_\infty = O(\varepsilon). \tag{11.3}$$

Remark. At present, our decomposition (11.1) depends on the parameter N. It is possible to modify the argument below in such a way that the decomposition is independent of N, but this requires generalising the notion of an averaged nilsequence by replacing the averaging over a finite set I with an integral over a continuous probability measure. As this introduces some minor technical issues such as measurability, we shall settle for the slightly weaker formulation of Proposition 11.2 given above, as it still suffices for our application.

We shall prove Proposition 11.2 shortly. Assuming it for the moment, we may make yet another reduction of the Main Theorem. This we do by reducing Proposition 10.2 (which, as we have already shown, implies the Main Theorem) to the following result.

Proposition 11.3 (W-tricked Λ orthogonal to averaged nilsequences)**.** *Let $s \geqslant 1$, and assume the* MN(s) *conjecture. Let $G/\Gamma = (G/\Gamma, d_{G/\Gamma})$ be an s-step nilmanifold with smooth metric $d_{G/\Gamma}$, and let $F_1(n)$ be an averaged s-step nilsequence with Lipschitz constant M. Let $b \in [W]$ be coprime to W. Suppose we also have the dual norm bound*

$$\|F_1\|_{U^{s+1}[N]^*} \leqslant M'. \tag{11.4}$$

Then we have the bound

$$\mathbb{E}_{n\in[N]}(\Lambda'_{b,W}(n) - 1)F_1(n) = o_{M,M',G/\Gamma,s}(1).$$

Indeed, to deduce Proposition 10.2 from Proposition 11.3, let $\varepsilon \in (0,1)$ be arbitrary and apply Proposition 11.2. The contribution of F_2 will be bounded by $O(\varepsilon) + o_\varepsilon(1)$ thanks to (1.5) and (11.3). The contribution of F_1 can be controlled using Proposition 11.3. Putting these estimates together leads to the bound

$$\mathbb{E}_{n\in[N]}(\Lambda'_{b,W}(n) - 1)F(g^n x) = o_{M,G/\Gamma,s,\varepsilon}(1) + o_\varepsilon(1) + O(\varepsilon).$$

Letting ε go to zero sufficiently slowly, we obtain the claim.

In later sections we shall prove Proposition 11.3. For now we turn to the task of proving Proposition 11.2.

Proof of Proposition 11.2. Fix G/Γ, s, M. Observe that if we have proven the proposition for a single Lipschitz function F, then if we perturb F in the L^∞ norm by ε then the statement is still true for the perturbed function (with slightly worse implied constants in the $O()$ notation). On the other hand, since G/Γ is a compact metric space, we know from the Arzelà-Ascoli theorem that the space of Lipschitz functions F on G/Γ with Lipschitz constant at most M is equicontinuous and hence compact in the uniform topology. In particular, it can be covered by finitely many balls in the uniform metric of radius $\varepsilon/2$, say. In view of this compactness[12], we see that it will suffice to establish the *qualitative* version of the Proposition, namely given any continuous function F (not necessarily Lipschitz) and any $\varepsilon > 0$, we have a decomposition (11.1) for all $N \geqslant 1$, $g \in G$ and $x \in G/\Gamma$, where F_1 is an averaged nilsequence on G/Γ with Lipschitz constant uniform in g, x, N, and with dual norm $\|F_1\|_{U^{s+1}[N]^*}$ bounded uniformly in N, g, x, and F_2 obeys the bound (11.3).

Fix F and ε. To proceed further we need to detect some "constraints" on the orbit $n \mapsto g^n x$ in G/Γ. The most convenient framework for giving such constraints will be the $(s+1)$-dimensional parallelepipeds in G/Γ, as studied in [32].

Definition 11.4 (Parallelepipeds in nilmanifolds)**.** Let $(G/\Gamma)^{\{0,1\}^{s+1}}$ denote the space of all 2^{s+1}-tuples $(x_\omega)_{\omega\in\{0,1\}^{s+1}}$. An $(s+1)$-*dimensional parallelepiped* is any element of $(G/\Gamma)^{\{0,1\}^{s+1}}$ having the form

$$(g^{n+\omega\cdot h}x)_{\omega\in\{0,1\}^{s+1}}$$

for some $g \in G$, $x \in G/\Gamma$, $n \in \mathbb{Z}$, and $h \in \mathbb{Z}^{s+1}$. Here, and for the remainder of the paper, we write $\omega \cdot h := \omega_1 h_1 + \cdots + \omega_{s+1} h_{s+1}$ where $\omega = (\omega_1, \ldots, \omega_{s+1})$ and $h = (h_1, \ldots, h_{s+1})$.

A fundamental property of s-step nilmanifolds is that the value of any one vertex of a parallelepiped (say, the zero vertex $x_{0^{s+1}}$, where $0^{s+1} := (0, \ldots, 0)$) is determined "continuously" by all the other vertices. In the following proposition, and for the remainder of the paper, write $\{0,1\}^{s+1}_* := \{0,1\}^{s+1} \setminus \{0^{s+1}\}$.

Proposition 11.5 (Parallelepiped constraint)**.** *There exists a compact set*

$$\Sigma \subseteq (G/\Gamma)^{\{0,1\}^{s+1}_*}$$

and a continuous function $P : \Sigma \to G/\Gamma$ such that, for any $(s+1)$-dimensional parallelepiped $(x_\omega)_{\omega\in\{0,1\}^{s+1}}$, we have $(x_\omega)_{\omega\in\{0,1\}^{s+1}_*} \in \Sigma$ and the constraint

$$x_{0^{s+1}} = P((x_\omega)_{\omega\in\{0,1\}^{s+1}_*}).$$

[12]One could also use the compactness of G/Γ to remove the requirement that all bounds be uniform in x. However the parameter g ranges over the non-compact group G and cannot be eliminated so easily; the range of the parameter n is similarly non-compact. Thus we will be forced to look for constraints in the orbit $g^n x$ which are *independent* of g and n. This helps motivate our introduction of cubes below.

This proposition is a topological and algebraic statement about the structure of nilmanifolds, and it was essentially proved in [32]. We supply a complete and self-contained proof in Appendix E, taking the opportunity to introduce the *Host-Kra cube groups*. A closely related statement regarding arithmetic progressions in nilmanifolds appeared in [26, Lemma 12.7]; results of this latter type seem to have been around in the ergodic theory community for some time and feature, for instance, in the papers of Furstenberg [13, 14].

For now, we shall simply illustrate this proposition with two model examples before continuing with the proof of Proposition 11.2.

Example 14 (Abelian shift). Take $s = 1$, let G be an abelian Lie group, and let Γ be a cocompact lattice in G. Thus G/Γ is a compact abelian Lie group, and any action of $g \in G$ on G/Γ has the form of a shift $x \mapsto x + g$. Of course, G/Γ is a 1-step nilmanifold. A 2-dimensional parallelepiped in this nilmanifold takes the form $(x + ng, x + (n + h_1)g, x + (n + h_2)g, x + (n + h_1 + h_2)g)$. The first vertex is a function of the other three. In the notation of Proposition 11.5 we can take $\Sigma := (\mathbb{R}/\mathbb{Z})^3$ and $P : \Sigma \to G/\Gamma$ be the map $P(y_{10}, y_{01}, y_{11}) := y_{01} + y_{10} - y_{11}$ and we easily verify that $y_{00} = P(y_{10}, y_{01}, y_{11})$ whenever $(y_{00}, y_{10}, y_{01}, y_{11})$ is a 2-dimensional parallelepiped.

Example 15 (Skew shift). For the sake of illustration, we consider a quotient G/Γ where G is 2-step nilpotent but not connected. The way we have set things up in this paper, then, G/Γ does not qualify as a nilmanifold; however one can modify this example so that it genuinely takes place in a nilmanifold (cf. the proof of Proposition 8.4).

Set $G := \left(\begin{smallmatrix} 1 & \mathbb{Z} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{smallmatrix}\right)$ and $\Gamma := \left(\begin{smallmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{smallmatrix}\right)$. Then G is 2-step nilpotent, and G/Γ may be identified with the torus $(\mathbb{R}/\mathbb{Z})^2$ via the map

$$(x, y) \mapsto \left(\begin{smallmatrix} 1 & 0 & y \\ 0 & 1 & x \\ 0 & 0 & 1 \end{smallmatrix}\right) \Gamma.$$

Taking $g := \left(\begin{smallmatrix} 1 & 1 & 0 \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{smallmatrix}\right)$, it is easy to check the action of g on G/Γ is given by $(x, y) \mapsto (x + \alpha, y + x)$. The 3-dimensional parallelepipeds of this nilflow take the form

$$\big(x + (n + \omega \cdot h)\alpha, y + \tfrac{1}{2}(n + \omega \cdot h)(n + \omega \cdot h + 1)\alpha + (n + \omega \cdot h)x\big)_{\omega \in \{0,1\}^3}.$$

The key point to note here is that the first coordinate is at most linear in n, h, while the second coordinate is at most quadratic. Take the set Σ to be the set of all 7-tuples $((x_\omega, y_\omega))_{\omega \in \{0,1\}^3_*}$ with the linear constraints

$$\begin{aligned} x_{000} + x_{011} &= x_{010} + x_{001} \\ x_{000} + x_{101} &= x_{001} + x_{100} \\ x_{000} + x_{110} &= x_{100} + x_{010}. \end{aligned}$$

The map $P : \Sigma \to (\mathbb{R}/\mathbb{Z})^2$ is given by the alternating sum

$$P(((x_\omega, y_\omega))_{\omega \in \{0,1\}^3_*}) := - \sum_{\omega \in \{0,1\}^3_*} (-1)^{|\omega|}(x_\omega, y_\omega).$$

This is ultimately a reflection of the fact that linear and quadratic functions have vanishing third derivative. Note, in contrast to the previous example, that for the skew

shift a vertex of a 2-dimensional parallelepiped is *not* determined continuously by the other three vertices.

Now we return to the task of proving Proposition 11.2. Let P and Σ be as in Proposition 11.5. The function $x \mapsto F(P(x))$ is continuous on the compact metric space Σ. By the Stone-Weierstrass theorem, we may approximate this function to uniform accuracy $O(\varepsilon)$ by a finite linear combination of tensor products of bounded Lipschitz functions on G/Γ, obtaining the uniform approximation

$$F(P(x)) = \sum_{\alpha \in A} \prod_{\omega \in \{0,1\}_*^{s+1}} H_{\omega,\alpha}(x_\omega) + O(\varepsilon)$$

for some finite index set A and some 1-bounded Lipschitz functions $H_{\omega,\alpha} : G/\Gamma \to [-1,1]$. In particular, since $(g^{n+\omega\cdot h})_{\omega \in \{0,1\}_*^{s+1}}$ lies in Σ and the image of this point under P is $g^n x$, we have

$$F(g^n x) = \sum_{\alpha \in A} \prod_{\omega \in \{0,1\}_*^{s+1}} H_{\omega,\alpha}(g^{n+\omega\cdot h} x) + O(\varepsilon)$$

for all $g \in G$, $x \in G/\Gamma$, $n \in \mathbb{Z}$, and $h \in \mathbb{Z}$.

Now we introduce the parameter $N \geqslant 1$ and average[13] the h parameter over the box $[N]^{s+1}$. In fact it is necessary to perform this averaging somewhat smoothly, to which end we take a smooth function cutoff $\sigma : \mathbb{R} \to [0,1]$ which is supported on $[-1,2]$ and equals 1 on $[0,1]$, and then set

$$F(g^n x) = F_1(n) + F_2(n)$$

where

$$F_1(n) := \sum_{\alpha \in A} \mathbb{E}_{h \in [N]^{s+1}} \sigma(h_1/N) \dots \sigma(h_{s+1}/N) \prod_{\omega \in \{0,1\}_*^{s+1}} H_{\omega,\alpha}(g^{n+\omega\cdot h} x)$$

and $F_2(n) = O(\varepsilon)$. In particular, we have $\|F_1\|_\infty \leqslant 1 + O(\varepsilon)$ since F is bounded by 1. By shrinking the Lipschitz functions $H_{\omega,\alpha}$ by a multiplicative factor of $1 - O(\varepsilon)$, and transferring the error over to F_2, we may in fact ensure that $\|F_1\|_\infty \leqslant 1$.

Now observe that for each fixed α, ω and h the function given by $n \mapsto H_{\omega,\alpha}(g^{n+\omega\cdot h} x) = H_{\omega,\alpha}(g^n (g^{\omega\cdot h} x))$ is a Lipschitz nilsequence on the s-step nilmanifold G/Γ, with Lipschitz constant independent of N, g and x. We remarked, in §8, that the Lipschitz nilsequences form an algebra in a certain sense. From this remark we conclude that F_1 is an averaged Lipschitz s-step nilsequence on the product space $(G/\Gamma)^{\{0,1\}_*^{s+1}}$, again with Lipschitz constant independent of N, g and x. To conclude the proof it suffices to show that F_1 is also bounded in $U^{s+1}[N]^*$ uniformly in N, g and x. By the triangle inequality and the definition of the $U^{s+1}[N]^*$ norm, it thus suffices to show that the absolute value of

$$\mathbb{E}_{n \in [N]; h \in [N]^{s+1}} f(n) \sigma(h_1/N) \dots \sigma(h_{s+1}/N) \prod_{\omega \in \{0,1\}_*^{s+1}} H_{\omega,\alpha}(g^{n+\omega\cdot h} x) \tag{11.5}$$

[13] One could take a limit here as $N \to \infty$, using an ergodic theorem to ensure suitable convergence; this would make the decomposition $F = F_1 + F_2$ independent of N, but at the cost of replacing the finite averaging in the definition of an averaged nilsequence with an infinite one. We omit the details.

is uniformly bounded in N, g, x whenever $f : [N] \to \mathbb{R}$ satisfies $\|f\|_{U^{s+1}[N]} \leqslant 1$. From this point onwards we do not care what the functions $n \mapsto H_{\omega,\alpha}(g^n x)$ actually are: it is merely important that they are 1-bounded. For that reason we write $\mathbf{b}_\omega(n) = H_{\omega,\alpha}(g^n x)$, whereupon the quantity (11.5) that we are to show is uniformly bounded becomes

$$\mathbb{E}_{n\in[N];h\in[N]^{s+1}} f(n)\sigma(h_1/N)\dots\sigma(h_{s+1}/N) \prod_{\omega\in\{0,1\}^{s+1}_*} \mathbf{b}_\omega(n+\omega\cdot h). \tag{11.6}$$

At this point we transfer to a group $\mathbb{Z}_{N'}$ where $N' = 10sN$ (say). Slightly abusing notation, the expression (11.6) is, up to factors of $O_s(1)$, equal to

$$\mathbb{E}_{n\in\mathbb{Z}_{N'};h\in\mathbb{Z}^{s+1}_{N'}} f(n)\sigma(h_1/N)\dots\sigma(h_{s+1}/N) \prod_{\omega\in\{0,1\}^{s+1}_*} \mathbf{b}_\omega(n+\omega\cdot h). \tag{11.7}$$

Here we have extended f from $[N]$ to all of $\mathbb{Z}_{N'}$ by defining it to be zero outside of $[N]$. Now by taking a Fourier expansion on $\mathbb{Z}^{s+1}_{N'}$ we may write

$$\sigma(h_1/N)\dots\sigma(h_{s+1}/N) = \sum_{r_1,\dots,r_{s+1}} c_{r_1,\dots,r_{s+1}} e\big((r_1h_1+\dots+r_{s+1}h_{s+1})/N\big).$$

By choosing the cutoff σ to be sufficiently smooth, we may ensure that

$$\sum_{r_1,\dots,r_{s+1}} |c_{r_1,\dots,r_{s+1}}| = O_s(1).$$

Thus to show that (11.7) is uniformly bounded it suffices to show the same for

$$\mathbb{E}_{n\in\mathbb{Z}_{N'};h\in\mathbb{Z}^{s+1}_{N'}} f(n)e\big((r_1h_1+\dots+r_{s+1}h_{s+1})/N\big) \prod_{\omega\in\{0,1\}^{s+1}_*} \mathbf{b}_\omega(n+\omega\cdot h) \tag{11.8}$$

for all $r_1,\dots,r_{s+1} \in \mathbb{Z}_{N'}$. It is easy to see that the exponential may be split up and incorporated into the $\mathbf{b}_\omega(\ )$ terms, and therefore we have reduced the matter to placing a bound on

$$\mathbb{E}_{n\in\mathbb{Z}_{N'};h\in\mathbb{Z}^{s+1}_{N'}} f(n) \prod_{\omega\in\{0,1\}^{s+1}_*} \mathbf{b}_\omega(n+\omega\cdot h). \tag{11.9}$$

Now we are assuming that $\|f\|_{U^{s+1}[N]} \leqslant 1$. By Lemma B.5 this implies that $\|f\|_{U^{s+1}(\mathbb{Z}_{N'})} = O_s(1)$. The boundedness now follows from the Gowers-Cauchy-Schwarz inequality (B.12). Tracing backwards, we see in turn that (11.9), (11.7), (11.6) and (11.5) are all $O_s(1)$, thereby concluding the proof. □

Although we will not need this fact here, it is interesting to note that Proposition 11.2 allows one to extend Proposition 8.2 from bounded f to integrable f:

Corollary 11.6 (Nilsequences obstruct uniformity, II)**.** *Let $s \geqslant 0$ and $\delta \in (0,1)$. Let $G/\Gamma = (G/\Gamma, d_{G/\Gamma})$ be a nilmanifold with some fixed smooth metric $d_{G/\Gamma}$, and let $(F(g^n x))_{n\in\mathbb{N}}$ be a bounded s-step nilsequence with Lipschitz constant at most M. Let $f : [N] \to \mathbb{R}$ be a function for which*

$$\mathbb{E}_{n\in[N]}|f(n)| \leqslant 1$$

and

$$|\mathbb{E}_{n\in[N]} f(n)F(g^n x)| \geqslant \delta.$$

Then we have

$$\|f\|_{U^{s+1}[N]} \gg_{s,\delta,M,G/\Gamma} 1.$$

Proof. We apply Proposition 11.2 with ε equal to a small multiple of δ, and conclude from the triangle inequality that

$$|\mathbb{E}_{n\in[N]} f(n)F_1(n)| \geqslant \delta/2.$$

Since F_1 has a $U^{s+1}[N]^*$ norm of $O_{s,\delta,M,G/\Gamma}(1)$, the claim follows. □

12. A splitting of the von Mangoldt function

To summarise so far, we have reduced the task of proving that the GI(s) and MN(s) conjectures imply the Main Theorem to the much easier task of establishing Proposition 11.3. This is a correlation estimate involving $\Lambda'_{b,W}$. It is convenient to return at this point to the original von Mangoldt function Λ. The contribution from the prime powers which are introduced when $\Lambda'_{b,W}$ is replaced by $\Lambda_{b,W}$ is easily seen to be negligible, and so it suffices to establish the estimate

$$\mathbb{E}_{n\in[N]}(\Lambda_{b,W}(n) - 1)F_1(n) = o_{M,M',G/\Gamma,s}(1).$$

Recalling the definition (5.1) of $\Lambda_{b,W}(n)$, we are thus trying to establish the bound

$$\mathbb{E}_{n\in[N]}(\frac{\phi(W)}{W}\Lambda(Wn+b) - 1)F_1(n) = o_{M,M',G/\Gamma,s}(1). \tag{12.1}$$

At this point we perform a standard decomposition of Λ into a "smooth" piece $\Lambda^\sharp$ corresponding to small divisors and a "rough" piece $\Lambda^\flat$ corresponding to large divisors. We take a small exponent $\gamma = \gamma_s > 0$, whose exact value will be specified later, and set $R := N^\gamma$. Observe from (8.1) that

$$\Lambda(n) = -\log R \sum_{d|n} \mu(d)\chi(\frac{\log d}{\log R})$$

where $\chi : \mathbb{R}^+ \to \mathbb{R}^+$ is the identity function $\chi(x) := x$. We now perform a smooth splitting $\chi = \chi^\sharp + \chi^\flat$, where $\chi^\sharp(x)$ vanishes for $|x| \geqslant 1$ and $\chi^\flat(x)$ vanishes for $|x| \leqslant 1/2$, the precise form of this splitting being unimportant. This induces a splitting $\Lambda = \Lambda^\sharp + \Lambda^\flat$, where

$$\Lambda^\sharp(n) := -\log R \sum_{d|n} \mu(d)\chi^\sharp(\frac{\log d}{\log R}) \quad \text{and} \quad \Lambda^\flat(n) := -\log R \sum_{d|n} \mu(d)\chi^\flat(\frac{\log d}{\log R}). \tag{12.2}$$

Thus to prove (12.1) it will suffice to show the estimates

$$\mathbb{E}_{n\in[N]}(\frac{\phi(W)}{W}\Lambda^\sharp(Wn+b) - 1)F_1(n) = o_{s,M'}(1) \tag{12.3}$$

and

$$\mathbb{E}_{n\in[N]}\frac{\phi(W)}{W}\Lambda^\flat(Wn+b)F_1(n) = o_{M,G/\Gamma,s}(1). \tag{12.4}$$

We begin by establishing the bound (12.3). It is here that we need the dual norm bound (11.4). Indeed, from that bound we have

$$\begin{aligned}\big|\mathbb{E}_{n\in[N]}(\frac{\phi(W)}{W}\Lambda^\sharp(Wn+b)-1)F_1(n)\big| &\leqslant \big\|\frac{\phi(W)}{W}\Lambda^\sharp(Wn+b)-1\big\|_{U^{s+1}[N]}\|F_1\|_{U^{s+1}[N]*}\\ &\leqslant M'\big\|\frac{\phi(W)}{W}\Lambda^\sharp(Wn+b)-1\big\|_{U^{s+1}[N]}.\end{aligned}$$

It suffices, then to show that

$$\big\|\frac{\phi(W)}{W}\Lambda^\sharp(Wn+b)-1\big\|_{U^{s+1}[N]} = o_s(1). \tag{12.5}$$

This is a multilinear correlation estimate for a truncated divisor sum, and can be treated by standard sieve theory methods related to the correlation estimates of Goldston and Yıldırım [16, 17, 18] provided that the exponent γ is sufficiently small (an appropriate choice would be, for example, $\gamma_s := \frac{1}{10}2^{-s}$). We provide the details of this computation in Appendix D. This establishes (12.3).

It remains to establish the bound (12.4). Recall that F_1 is an averaged nilsequence. From the triangle inequality, it will thus suffice to prove the bound

$$\mathbb{E}_{n\in[N]}\frac{\phi(W)}{W}\Lambda^\flat(Wn+b)F(g^nx) = o_{M,G/\Gamma,s}(1) \tag{12.6}$$

for all 1-bounded s-step nilsequences $F(g^nx)$ of Lipschitz constant M. We emphasise that the o-term is required to depend only on $M, G/\Gamma$ and s, and should be otherwise be independent of F, g and x.

We will eventually apply the MN(s) conjecture, which comes with the safety net of an error term which decays like $\log^{-A} N$ for any A. With this in mind, we begin by removing the W-dependence in (12.6) in a rather crude fashion. Since $\phi(W)/W \leqslant 1$, we ignore this factor completely.

Now by a simple substitution we have

$$\mathbb{E}_{n\in[N]}\Lambda^\flat(Wn+b)F_1(g^nx) = W\mathbb{E}_{b<n\leqslant WN+b}1_{n\equiv b(\mathrm{mod}\ W)}\Lambda^\flat(n)F_1(g^{(n-b)/W}x). \tag{12.7}$$

Now any Lie group G over $\mathbb{R}$ for which the exponential map $\exp : \mathfrak{g} \to G$ from the associated Lie algebra is surjective is *divisible*, meaning that given any $g \in G$ and any positive integer m there is an element $g^{1/m} \in G$ with $(g^{1/m})^m = g$. When G is simply-connected and nilpotent, exp is a homeomorphism (see [7] for details). In our setting, write $g' := g^{1/W}$ and $x' := g^{-b/W}x$. Then for all $n \equiv b(\mathrm{mod}\ W)$ we have

$$F_1(g'^nx') = F_1(g^{(n-b)/W}x). \tag{12.8}$$

Note that the left-hand side here makes perfect sense for *any* n, not just for n such that $n \equiv b(\mathrm{mod}\ W)$.

The constraint $1_{n\equiv b(\mathrm{mod}\ W)}$ may be expanded as a Fourier series

$$1_{n\equiv b(\mathrm{mod}\ W)} = \frac{1}{W}\sum_{r\in\mathbb{Z}_W} e(-rb/W)e(rn/W)$$

on $\mathbb{Z}_W$. We substitute this and (12.8) into (12.7), noting that each function $n \mapsto e(rn/W)$ may be realised as a 1-bounded, $O(1)$-Lipschitz nilsequence on the 1-step nilmanifold $\mathbb{R}/\mathbb{Z}$. Replacing G/Γ with $G/\Gamma \times \mathbb{R}/\mathbb{Z}$, we see that in order to prove (12.6) it suffices to show that

$$W\mathbb{E}_{b<n\leqslant WN+b}\Lambda^\flat(n)F(g^n x) = o_{M,G/\Gamma,s}(1) \tag{12.9}$$

for all M-Lipschitz 1-bounded nilsequences $(F(g^n x))_{n\in N}$ on an s-step nilmanifold G/Γ.

In fact we will establish the stronger estimate

$$|\sum_{n\in[N]} \Lambda^\flat(n)F(g^n x)| \ll_{M,G/\Gamma,s,A} N\log^{-A} N \tag{12.10}$$

for any $A > 0$. Note that w was chosen to be so slowly growing that $W = O(\log N)$, so this estimate really is stronger than (12.9). We expand the left-hand side of (12.10) using (12.2) and reduce to showing that

$$|\sum_{n\in[N]}\sum_{d|n} \mu(d)\chi^\flat(\frac{\log d}{\log R})F(g^n x)| \ll_{M,G/\Gamma,s,A} N\log^{-A} N. \tag{12.11}$$

The left-hand side may be rearranged as

$$|\sum_{m\in[N]}\sum_{d\in[N/m]} \mu(d)\chi^\flat(\frac{\log d}{\log R})F((g^m)^d x)|.$$

Observe that $\chi^\flat$ is supported on $|x| \geqslant 1/2$, and so the summand vanishes unless $d \geqslant R^{1/2}$, in which case $m \leqslant N/R^{1/2}$. We now apply the Möbius and nilsequences conjecture MN(s). Together with a straightforward summation by parts to remove the smooth cutoff $\chi^\flat$ this shows that

$$|\sum_{d\in[N/m]} \mu(d)\chi^\flat(\frac{\log d}{\log R})F((g^m)^d x)| \ll_{M,G/\Gamma,s,A} \frac{N}{m}\log^{-A}\frac{N}{m}.$$

Note that we are making critical use here of the fact that the bounds in the MN(s) conjecture are uniform in the g parameter in order to deal with the fact that we have dilated g to g^m. Since $m \leqslant N/R^{1/2}$, we see that $\log^{-A}(N/m) \ll_A \log^{-A} N$. Summing in m and absorbing the logarithmically divergent sum $\sum_{m\in[N]}\frac{1}{m}$ into the $\log^{-A} N$ factor we obtain (12.11) as desired. This in turn implies (12.10) and hence, by our earlier series of reductions, (12.4). Together with (12.3), which we have already established, this concludes the proof of Proposition 11.3. By our long series of earlier reductions, this (finally!) completes the proof of the Main Theorem. □

13. Variations on the main argument and other remarks

It is conceivable that our methods here extend to certain "finite complexity" multilinear averages involving systems of *polynomials* $\psi_j(n)$ rather than affine-linear forms. Indeed, the machinery of "PET induction" (see e.g. [5]) allows us in principle to use repeated applications of Cauchy-Schwarz to control certain of these averages by Gowers uniformity norms. A model problem would be to count the number of p, n for which the

numbers $p, p+n, p+n^2, \dots, p+n^k$ are all prime. A naïve attempt to do this meets with what seems to be an insurmountable obstacle. Namely, in order to restrict the range of the primes concerned to an interval such as $[N]$, certain other parameters (for example the "shifts" h in the definition of the Gowers norms) have to be restricted to a much smaller range, say of size $O(N^{1/100})$. This makes it impossible to pass back and forth between $[N]$ and $\mathbb{Z}_{N'}$ as we have done above, and the evaluation of exponential sums with μ or Λ on such a range seems to be beyond hope, even assuming the GRH. It may be that the PET induction scheme can be "globalised" to avoid these issues, but we do not know how to address this at present.

For the benefit of readers who are only interested in the unconditional "quadratic" ($s = 2$) applications of this paper such as Corollary 1.7 or Examples 5-7 we outline a shorter path to the Main Theorem in that case. This approach avoids Lie theory completely, and probably represents the best approach to obtaining bounds for error terms. Note, however, that with either approach our error terms are completely ineffective unless the GRH is assumed. The introduction of Lie theory, though strictly speaking unnecessary, seems to make our work easier to understand from the conceptual point of view. This is especially the case when $s \geqslant 3$, where it is not even clear how Lie theory-free analogues of the GI(s) and MN(s) conjecture might be formulated.

In the quadratic case it is possible to replace the concept of a 2-step nilsequence by more concrete objects. In a sense these are more basic than 2-step nilsequences, if only because in [26] we introduce these objects first and then build nilsequences from them. Note, however, that this may be an artifact of our approach.

These more basic objects can then be manipulated by hand without resorting to machinery such as the Host-Kra theory in Appendix E. Let us consider, by way of illustration, the following more concrete version of the inverse Gowers-norm conjecture GI(2) which was proven in [26].

Theorem 13.1 (U^3 inverse theorem with bracket polynomials)**.** *Let $f : [N] \to [-1, 1]$ be such that $\|f\|_{U^3[N]} > \delta$ for some $0 < \delta \leqslant 1$ and $N \geqslant 1$. Then there exists a positive integer $J = O_\delta(1)$ and real numbers $a_j, b_j, \xi_{j,1}, \xi_{j,2}, \xi_{j,3}$ for $j \in [J]$ such that*

$$|\mathbb{E}_{n\in[N]} f(n)e(\phi(n))| \gg_\delta 1 \tag{13.1}$$

where ϕ is the function

$$\phi(n) := -\sum_{j\in[J]} \left(a_j\{\xi_{j,1}n\}\{\xi_{j,2}n\} + b_j\{\xi_{j,3}n\}\right).$$

Remark. As before, $\{x\}$ denotes the fractional part of x, which we take to lie in $(-\frac{1}{2}, \frac{1}{2}]$.

This result follows quickly from [26, Theorem 10.9] using Lemma B.5 to work in a cyclic group of prime order. We refer to the phase $\phi(n)$ (13.1) as a "bracket polynomial". By modifying the arguments in §10, one can transfer this theorem to the case when f is bounded by a pseudorandom measure ν rather than by 1, thereby reducing Theorem

7.2 to the establishment of the exponential sum estimate

$$\mathbb{E}_{n\in[N]}(\Lambda_{b,W}(n)-1)e\big(-\sum_{j\in[J]}\big(a_j\{\xi_{j,1}n\}\{\xi_{j,2}n\}+b_j\{\xi_{j,3}n\}\big)\big)=o_J(1)$$

uniformly over all $b\in[W]$ with $\gcd(b,W)=1$. This could in principle[14] be established directly by Vinogradov's method, following the machinery in [27], though the argument would be rather lengthy. Alternatively one can deduce this result from the corresponding results for the Möbius function established in [27] using a variant of the arguments in this paper.

A key difference is that the Host-Kra machinery and the machinery of averaged nilsequences are no longer required. Instead, the above function $e(\phi(n))$ can be replaced by a smoother variant, constructed for instance using a variant of the dual function machinery in [24], in order to obtain a function which is bounded in $(U^3)^*$. This provides an analogue of Proposition 11.3, and from that point onwards one may proceed similarly.

One could also use a still more "basic" type of obstruction for the U^3-norm, namely phases which are locally quadratic on Bohr sets (cf. [26, §2]). These require even less unpacking than the bracket quadratics above, and indeed it was found to be rather convenient to work with these functions in [27]. It takes a while to even define these functions properly, however, and they suffer from a few technical deficiencies which affect various other steps of the argument. Perhaps the most serious is that if $n\mapsto f(n)$ is such a function then $n\mapsto f(dn)$ need not quite be, a phenomenon which causes trouble in §12. □

14. A brief discussion of bounds

We have shied away from giving any explicit bounds on our $o(1)$ error terms. There are at least two reasons for this. Firstly, it is notationally easier to avoid doing so. Secondly, and much more importantly, unless one assumes the GRH we do not have any explicit bounds!

By way of illustration, let us consider the statement

$$\mathbb{E}_{x,d\leqslant N}\mu(x)\mu(x+d)\mu(x+2d)\mu(x+3d)=o(1), \tag{14.1}$$

which follows from the case $s=2$ of Proposition 9.1. A discussion of correlations involving Λ would go along similar lines, but there is the distraction of the singular product $\beta_\infty\prod_p\beta_p$.

As we remarked, the error term here is completely ineffective without assuming GRH. Indeed to show that the left-hand side in (14.1) is at most δ, we would ultimately (deep inside the paper [27]) need estimates for the sum of the Möbius function over arithmetic

[14]Indeed, this exponential sum is a more complicated variant of the more traditional exponential sum $\sum_{n\in[N]}\Lambda(n)e(\alpha n^2)$, which was considered for instance in [15, 33] .

progressions with common difference $q \sim \log^{A(\delta)} N$. Although such estimates exist, the error terms involve an ineffective constant $C(A(\delta))$ due to the possible presence of Landau-Siegel zeros.

Assuming the GRH one could prove using our methods that

$$|\mathbb{E}_{x,d\leqslant N}\mu(x)\mu(x+d)\mu(x+2d)\mu(x+3d)| \leqslant C\log^{-c} N$$

for some explicit C and some explicit (but small) $c > 0$. To obtain such a result it would be best to avoid the use of Lie theory as outlined in §13, since the many approximation arguments involved in that theory are quite costly from the quantitative point of view.

Improved results in additive combinatorics (particularly a solution to the so-called Polynomial Freiman-Ruzsa conjecture, which could be used as an input in [26]) could lead to a bound of the shape $\exp(-\log^c N)$. However it seems that obtaining a bound N^{-c} is very difficult.

Unconditionally, a bound in (14.1) of the form $O(f(n))$ for some explicit function $f(n)$ tending to zero as $n \to \infty$ and some ineffective implied constant $O()$ would be very interesting.

To set the above discussion in context, we mention the best available results for three-term progressions, which follow from estimates for $\sup_{\alpha\in\mathbb{R}/\mathbb{Z}} |\mathbb{E}_{n\leqslant N}\mu(n)e(\alpha n)|$. These seem to be as follows.

$$\mathbb{E}_{x,d\leqslant N}\mu(x)\mu(x+d)\mu(x+2d) \ll \begin{cases} C_A \log^{-A} N & \text{any } A > 0 \quad \text{Davenport [10]} \\ C_\epsilon N^{-1/4+\epsilon} & \text{on GRH} \quad \text{Baker–Harman [2].} \end{cases}$$

Bounds of a similar type could be obtained for any instance of Proposition 9.1 with $s = 1$.

Appendix A. Elementary convex geometry

In this appendix we recall some profoundly classical facts concerning convex bodies which will allow us to manipulate cutoffs such as 1_K readily, beginning with an ancient observation of Archimedes.

Lemma A.1 (Archimedes comparison principle)**.** *Let $K_1 \subseteq K_2 \subseteq \mathbb{R}^d$ be bounded convex bodies. Then the surface area of K_1 is less than or equal to the surface area of K_2.*

Proof. It is easy to see that the intersection of K_2 with a half-space has lesser or equal surface area than K_2. Since K_1 can be approximated to arbitrary accuracy by the intersection of finitely many half-spaces, the claim follows. □

Corollary A.2 (Boundary region estimate)**.** *Let $K \subseteq [-N,N]^d$ be a convex body. If $\varepsilon \in (0,1)$, then the εN-neighbourhood of the boundary ∂K has volume $O_d(\varepsilon N^d)$.*

Proof. Rescale so that $N = 1$. By differentiating in ε we see that it suffices to show that any convex body in $[-2,2]^d$ has surface area $O_d(1)$. But this follows from the Archimedes comparison principle. One could also derive this fact using the theory of mixed volumes; see [38]. □

At this point we can now readily prove (1.3) using the Gauss volume-packing argument. By intersecting K with the half-spaces $\{x \in \mathbb{R}^d : \psi_j(x) > 0\}$ it suffices to show that

$$|K \cap \mathbb{Z}^d| = \mathrm{vol}_d(K) + O_d(N^{d-1})$$

for all convex bodies $K \subseteq [-N,N]$. However, given that $|K \cap \mathbb{Z}^d|$ is equal to the volume of the set $(K \cap \mathbb{Z}^d) + [-1/2,1/2]^d$, which differs from K only on the $O_d(1)$-neighbourhood of ∂K, the claim then follows from Corollary A.2.

Now we give an analytic consequence of Corollary A.2.

Corollary A.3 (Lipschitz approximation of convex indicators)**.** *Let $K \subseteq [-N,N]^d$ be a convex body and let $\varepsilon \in (0,1)$. Then we can write $1_K = F_\varepsilon + O(G_\varepsilon)$, where $F_\varepsilon, G_\varepsilon$ are non-negative Lipschitz functions on $[-2N,2N]^d$ with Lipschitz constants $O(\frac{1}{\varepsilon N})$ and bounded in magnitude by 1, and where $\int_{\mathbb{R}^d} G_\varepsilon(x)\,dx = O_d(\varepsilon N^d)$.*

Proof. We take

$$F_\varepsilon(x) := \max(1 - \frac{\mathrm{dist}_{\mathbb{R}^d}(x,K)}{\varepsilon N}, 0) \quad \text{and} \quad G_\varepsilon(x) := \max(1 - \frac{\mathrm{dist}_{\mathbb{R}^d}(x,\partial K)}{2\varepsilon N}, 0).$$

The claim follows easily from Corollary A.2. □

In practice, Corollary A.3 allows us to replace a rough cutoff such as 1_K with the smoother operation of Lipschitz cutoffs. This can then be combined with Fourier analysis to replace the Lipschitz cutoffs in turn with modulations by linear phases, which turn out to be utterly harmless in our analysis. This might remind readers of the Pólya-Vinogradov completion-of-sums method, or the Erdős-Turán inequality.

Appendix B. Gowers norm theory

In this appendix we develop the general "elementary" theory of Gowers uniformity norms, which were introduced in [21] and subsequently, in the rather different context of ergodic theory, in [32]. By elementary in this context, we basically mean that we only pursue here those results which can be obtained as an easy consequence of the Cauchy-Schwarz inequality. This is in contrast to the more advanced inverse theory involving nilsequences, Fourier analysis, and suchlike. The theory here is an amalgam of parts of [21, §3], [23], [24, §5], [26, §1], [32], [41, §3], [44, 45], or [46, Ch. 11].

It is convenient to work rather abstractly at first, dealing with complex-valued functions of many variables. This level of abstraction will be useful for us when we prove the generalised von Neumann theorem, Proposition 7.1, in §C. The argument is essentially that of [24, §5], generalised to handle arbitrary systems of linear forms rather than merely k-term APs, but the introduction of extra notation somewhat eases the process of actually carrying this out.

Definition B.1 (Gowers box norms)**.** Let $(X_\alpha)_{\alpha\in A}$ be a finite non-empty collection of finite non-empty sets, and for any $B \subseteq A$ write $X_B := \prod_{\alpha\in B} X_\alpha$ for the Cartesian product. If $f : X_A \to \mathbb{C}$ is a complex-valued function, we define the *Gowers box norm* $\|f\|_{\square(X_A)} \in \mathbb{R}^+$ to be the unique non-negative real number such that

$$\|f\|^{2^{|A|}}_{\square(X_A)} := \mathbb{E}_{x_A^{(0)},x_A^{(1)}\in X_A} \prod_{\omega_A\in\{0,1\}^A} \mathcal{C}^{|\omega_A|} f(x_A^{(\omega_A)}) \tag{B.1}$$

where $\mathcal{C} : z \mapsto \overline{z}$ is complex conjugation, and for any $x_A^{(0)} = (x_\alpha^{(0)})_{\alpha\in A}$ and $x_A^{(1)} = (x_\alpha^{(1)})_{\alpha\in A}$ in X_A and $\omega_A = (\omega_\alpha)_{\alpha\in A}$ in $\{0,1\}^A$, we write $x_A^{(\omega)} := (x_\alpha^{(\omega_\alpha)})_{\alpha\in A}$ and $|\omega_A| := \sum_{\alpha\in A}\omega_\alpha$. We adopt the convention that if A is empty (so that f is a constant), then $\|f\|_{\square(X_A)} := f$.

It is not immediately obvious that the right-hand side of (B.1) is non-negative, or that the term "norm" is appropriate. We will establish both of these facts below.

Examples 2. If $A = \{1\}$, then

$$\|f\|_{\square(X_1)} = \big(\mathbb{E}_{x_1^{(0)},x_1^{(1)}\in X_1} f(x_1^{(0)})f(x_1^{(1)})\big)^{1/2} = |\mathbb{E}_{x_1\in X_1} f(x_1)|$$

while if $A = \{1,2\}$, then $\|f\|_{\square(X_{1,2})} =$

$$\left(\mathbb{E}_{x_1^{(0)},x_1^{(1)}\in X_1;x_2^{(0)},x_2^{(1)}\in X_2} f(x_1^{(0)},x_2^{(0)})\overline{f(x_1^{(0)},x_2^{(1)})f(x_1^{(1)},x_2^{(0)})}f(x_1^{(1)},x_2^{(1)})\right)^{1/4}.$$

In general, the $2^{|A|}$th power of the $\square(X_A)$ norm on f_A is a multilinear average of f_A over $|A|$-dimensional boxes (hence the name).

It is easy to verify the recursive relationship

$$\|f\|^{2^{|A|}}_{\square(X_A)} = \mathbb{E}_{x_\alpha^{(0)},x_\alpha^{(1)}\in X_\alpha} \|f(\cdot,x_\alpha^{(0)})\overline{f(\cdot,x_\alpha^{(1)})}\|^{2^{|A|-1}}_{\square(X_{A\setminus\{\alpha\}})} \tag{B.2}$$

whenever $\alpha \in A$, which can be used as an alternate definition of the box norms. In particular we see that the box norms $\|f\|_{\square(X_A)}$ are non-negative for A non-empty. These norms are also conjugation-invariant, homogeneous, and enjoy the positivity property

$$\|f\|_{\square(X_A)} \leqslant \|\nu\|_{\square(X_A)} \tag{B.3}$$

whenever $f : X_A \to \mathbb{C}$ and $\nu : X_A \to \mathbb{R}^+$ obey the pointwise bound $|f(x_A)| \leqslant \nu(x_A)$ for all $x_A \in X_A$.

The box norms are also invariant under a large class of phase modulations. Indeed one easily verifies from (B.2) and induction that

$$\|fe(\sum_{B\subsetneq A}\phi_B)\|_{\square(X_A)} = \|f\|_{\square(X_A)} \tag{B.4}$$

where $e : \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ is the standard character $e(x) := e^{2\pi i x}$ and for each proper subset $B \subseteq A$, the phase function $\phi_B : X_B \to \mathbb{R}/\mathbb{Z}$ is arbitrary. Thus the $\square(X_A)$ norm is insensitive to "lower order" modulations which involve only a proper subset of the variables in X_A.

A fundamental inequality[15] concerning these norms is

Lemma B.2 (Gowers-Cauchy-Schwarz inequality)**.** *Let $(X_\alpha)_{\alpha\in A}$ be a finite collection of finite non-empty sets. For every $\omega_A \in \{0,1\}^A$ let $f_{\omega_A} : X_A \to \mathbb{C}$ be a function. Then*

$$\big|\mathbb{E}_{x_A^{(0)},x_A^{(1)}\in X_A} \prod_{\omega_A\in\{0,1\}^A} \mathcal{C}^{|\omega_A|} f_{\omega_A}(x_A^{(\omega_A)})\big| \leqslant \prod_{\omega_A\in\{0,1\}^A} \|f_{\omega_A}\|_{\square(X_A)}. \tag{B.5}$$

Proof. We induct on $|A|$. When $|A| = 0$ the claim trivially holds, and in fact there is equality. Now suppose that $|A| \geqslant 1$ and the claim has already been proven for smaller sets A.

Partition A as $A' \cup \{\alpha\}$ for some $\alpha \in A$. We can rewrite the left-hand side of (B.5) as

$$|\mathbb{E}_{x_{A'}^{(0)},x_{A'}^{(1)}\in X_{A'}} \prod_{\omega_\alpha\in\{0,1\}} \mathcal{C}^{\omega_\alpha} F_{\omega_\alpha}(x_{A'}^{(0)}, x_{A'}^{(1)})|$$

where

$$F_{\omega_\alpha}(x_{A'}^{(0)}, x_{A'}^{(1)}) := \mathbb{E}_{x_\alpha^{(\omega_\alpha)}\in X_\alpha} \prod_{\omega_{A'}\in\{0,1\}^{A'}} \mathcal{C}^{|\omega_{A'}|} f_{(\omega_{A'},\omega_\alpha)}(x^{(\omega_{A'})}, x_\alpha).$$

By Cauchy-Schwarz it thus suffices to show that

$$\mathbb{E}_{x_{A'}^{(0)},x_{A'}^{(1)}\in X_{A'}} |F_{\omega_\alpha}(x_{A'}^{(0)}, x_{A'}^{(1)})|^2 \leqslant \prod_{\omega_{A'}\in\{0,1\}^{A'}} \|f_{(\omega_{A'},\omega_\alpha)}\|_{\square(X_A)}^2$$

for each $\omega_\alpha \in \{0,1\}$. We can expand the left-hand side as

$$\mathbb{E}_{x_\alpha^{(0)},x_\alpha^{(1)}\in X_\alpha} \mathbb{E}_{x_{A'}^{(0)},x_{A'}^{(1)}\in X_{A'}} \prod_{\omega_{A'}\in\{0,1\}^{A'}} \mathcal{C}^{|\omega_{A'}|}\Big(f_{(\omega_{A'},\omega_\alpha)}(x_{A'}^{(\omega_{A'})}, x_\alpha^{(0)})\overline{f_{(\omega_{A'},\omega_\alpha)}(x_{A'}^{(\omega_{A'})}, x_\alpha^{(1)})}\Big).$$

Applying the induction hypothesis, we can bound this by

$$\mathbb{E}_{x_\alpha^{(0)},x_\alpha^{(1)}\in X_\alpha} \prod_{\omega_{A'}\in\{0,1\}^{A'}} \|f_{(\omega_{A'},\omega_\alpha)}(\cdot, x_\alpha^{(0)})\overline{f_{(\omega_{A'},\omega_\alpha)}(\cdot, x_\alpha^{(1)})}\|_{\square(X_{A'})}$$

and the claim now follows from Hölder's inequality and (B.2). □

From (B.5) we easily deduce the *Gowers triangle inequality*

$$\|f + g\|_{\square(X_A)} \leqslant \|f\|_{\square(X_A)} + \|g\|_{\square(X_A)}$$

[15]In our treatment here, this inequality plays a more central role than in earlier papers; we are using it as a kind of "universal Cauchy-Schwarz inequality", in the sense that any other inequality that we need, which would in earlier papers be proven by multiple applications of the ordinary Cauchy-Schwarz inequality, is instead proven here by a single application of the Gowers-Cauchy-Schwarz inequality. This seems to fit with the philosophy that the Gowers norms are somehow "universal" or "characteristic" for all averages of a certain complexity.

as can be seen by raising both sides to the power $2^{|A|}$. Let us also observe, setting all but one of the functions in (B.5) to be Kronecker delta functions, that if $\|f\|_{\square(X_A)} = 0$ and $|A| \geqslant 2$ then f vanishes identically. Thus we see that the $\square(X_A)$-norm is indeed a norm for $|A| \geqslant 2$, whilst for $|A| = 1$ it is merely a semi-norm.

As a consequence of the Gowers-Cauchy-Schwarz inequality we obtain

Corollary B.3 (Second Gowers-Cauchy-Schwarz inequality)**.** *Let $(X_\alpha)_{\alpha\in A}$ be a collection of finite non-empty sets. For every $B \subseteq A$ let $f_B : X^B \to \mathbb{C}$ be a function. Then*

$$|\mathbb{E}_{x_A\in X_A} \prod_{B\subseteq A} f_B(x_B)| \leqslant \prod_{B\subseteq A} \|f_B^{\overline{2}^{|A|-|B|}}\|_{\square(X_B)}^{1/2^{|A|-|B|}} \tag{B.6}$$

where $x_B \in X_B$ is the restriction of x_A to the indices B, and for any complex number z we define $z^{\overline{2}^n} := z$ when $n = 0$ and $z^{\overline{2}^n} := |z|^{2^n}$ for $n > 0$.

Proof. For each $\omega_A \in \{0,1\}^A$ we let $f_{\omega_A} : X_A \to \mathbb{C}$ be the function

$$f_{\omega_A}(x_A) := \mathcal{C}^{|\omega_A|} f_B(x_B)$$

where $B := \{\alpha \in A : \omega_\alpha = 1\}$. Then we can rewrite the above left-hand side as

$$|\mathbb{E}_{x_A^{(0)},x_A^{(1)}\in X_A} \prod_{\omega_A\in\{0,1\}^A} \mathcal{C}^{|\omega_A|} f_{\omega_A}(x_A^{(\omega_A)})|$$

which by the Gowers-Cauchy-Schwarz inequality is bounded by

$$\prod_{\omega_A\in\{0,1\}^A} \|f_{\omega_A}\|_{\square(X_A)}.$$

However, direct calculation (using (B.2), for instance) shows that

$$\|f_{\omega_A}\|_{\square(X_A)} = \|f_B^{\overline{2}^{|A|-|B|}}\|_{\square(X_B)}^{1/2^{|A|-|B|}}$$

where $B := \{\alpha \in A : \omega_\alpha = 1\}$, and the claim follows. □

As a special case of Corollary B.3 (together with (B.3)), we see that

$$|\mathbb{E}_{x_A\in X_A} f_A(x_A) \prod_{B\subsetneq A} f_B(x_B)| \leqslant \|f_A\|_{\square(X_A)} \tag{B.7}$$

whenever the functions f_B are bounded in magnitude by 1 for $B \subsetneq A$; compare this with (B.4). The inequality (B.7) asserts that the $\square$ norm is stable with respect to lower order functions and can be viewed as a type of generalised von Neumann theorem.

Remark. If f_A is also bounded by 1, then there is a converse to (B.7), namely that there exist bounded functions f_B for which

$$|\mathbb{E}_{x_A\in X_A} f_A(x_A) \prod_{B\subsetneq A} f_B(x_B)| \geqslant \|f_A\|_{\square(X_A)}^{2^{|A|}}.$$

Indeed this follows easily from raising (B.1) to the power $2^{|A|}$ and using the pigeonhole principle to freeze the $x_A^{(1)}$ variables. Thus we see that the lower order functions $\prod_{B\subsetneq A} f_B(x_B)$ are "characteristic" for the $\square(X_A)$ norm: if $\|f_A\|_{\square(X_A)}$ is large then f_A

correlates with a function of the form $\prod_{B \subsetneq A} f_B(x_B)$. One can pursue this idea to eventually obtain the hypergraph version of the Szemerédi regularity lemma, a task which was carried out fully in [44].

In our applications we will need to generalise (B.7) to the case where the f_B are bounded by some other functions ν_B. Fortunately this is also an easy consequence of Corollary B.3:

Corollary B.4 (Weighted generalised von Neumann theorem)**.** *Let $(X_\alpha)_{\alpha \in A}$ be a finite collection of finite non-empty sets. For every $B \subseteq A$ let $f_B : X_B \to \mathbb{C}$ and $\nu_B : X_B \to \mathbb{R}^+$ be functions such that $|f_B(x_B)| \leqslant \nu_B(x_B)$ for all $x_B \in X_B$. Then*

$$|\mathbb{E}_{x_A \in X_A} \prod_{B \subseteq A} f_B(x_B)| \leqslant \|f_A\|_{\square^A(\nu;X_A)} \prod_{B \subsetneq A} \|\nu_B\|^{1/2^{|A|-|B|}}_{\square^B(\nu;X_B)} \tag{B.8}$$

where for any $B \subseteq A$ and $g_B : X_B \to \mathbb{C}$ we define $\|g_B\|_{\square^B(\nu;X_B)}$ to be the unique nonnegative real number satisfying

$$\|g_B\|^{2^{|B|}}_{\square^B(\nu;X_B)} := \mathbb{E}_{x_B^{(0)}, x_B^{(1)} \in X_B} \Big(\prod_{\omega_B \in \{0,1\}^B} \mathcal{C}^{|\omega_B|} g_B(x_B^{(\omega_B)}) \Big) \prod_{C \subsetneq B} \prod_{\omega_C \in \{0,1\}^C} \nu_C(x_C^{(\omega_C)}).$$

Remark. It follows from (B.10) below that the right-hand side of the last equation is non-negative, and so $\|g_B\|_{\square(\nu;X_B)}$ is well-defined. Note for instance that

$$\|\nu_B\|_{\square^B(\nu;X_B)} := \big(\mathbb{E}_{x_B^{(0)}, x_B^{(1)} \in X_B} \prod_{C \subseteq B} \prod_{\omega_C \in \{0,1\}^C} \nu_C(x_C^{(\omega_C)}) \big)^{1/2^{|B|}}. \tag{B.9}$$

and

$$\|f_B\|_{\square(1;X_B)} = \|f_B\|_{\square(X_B)}.$$

Proof. By a limiting argument we may assume that the ν_B are strictly positive throughout X_B. We refactorise

$$\prod_{B \subseteq A} f_B(x_B) = \prod_{B \subseteq A} \tilde{f}_B(x_B)$$

where

$$\tilde{f}_B(x_B) := \frac{f_B(x_B)}{\nu_B(x_B)} \prod_{C \subseteq B} \nu_C(x_B)^{1/2^{|A|-|C|}}.$$

Applying Corollary B.3 we can thus bound the left-hand side of (B.8) by

$$\|\tilde{f}_A\|_{\square(X_A)} \prod_{B \subsetneq A} \|\tilde{f}_B^{2^{|A|-|B|}}\|^{1/2^{|A|-|B|}}_{\square(X_B)}.$$

However, direct calculation shows that

$$\|\tilde{f}_A\|_{\square(X_A)} = \|f_A\|_{\square(\nu;X_A)}, \tag{B.10}$$

whilst the pointwise bound

$$|\tilde{f}_B(x_B)| \leqslant \prod_{C \subseteq B} \nu_C(x_B)^{1/2^{|A|-|C|}}$$

together with (B.3) gives

$$|\|\tilde{f}_B^{2^{|A|-|B|}}\|_{\square(X_B)}^{1/2^{|A|-|B|}}| \leqslant \|\prod_{C\subseteq B}\nu_C(x_B)^{1/2^{|B|-|C|}}\|_{\square(X_B)}^{1/2^{|A|-|B|}}$$
$$= \|\nu_B\|_{\square(\nu;X_B)}^{1/2^{|A|-|B|}}$$

and the claim follows. □

Remark. In order for this inequality to be useful, one needs to compare the weighted $\square$ norm $\|f\|_{\square(\nu;X_A)}$ with the unweighted norm $\|f\|_{\square(X_A)}$. For any fixed set of weights ν, this is not possible when the ν are unbounded; however, if the ν also depend on an additional parameter y, then we will be able to establish comparability estimates of this type after averaging in y, assuming that ν obeys suitable "linear forms conditions". See Appendix C; similar ideas appear in [24, 45].

Now we pass from this abstract setting to a more "additive" setting. Given any $s \geqslant 0$, any finite additive group Z and any function $f : Z \to \mathbb{C}$, we define the *Gowers uniformity norm* $\|f\|_{U^{s+1}(Z)}$ by the formula

$$\|f\|_{U^{s+1}(Z)} := \|f(x_1 + \ldots + x_{s+1})\|_{\square^{s+1}(Z^{s+1})}.$$

Equivalently, we have

$$\|f\|_{U^{s+1}(Z)}^{2^{s+1}} = \mathbb{E}_{x^{(0)},x^{(1)}\in Z^{s+1}}\prod_{\omega\in\{0,1\}^{s+1}}\mathcal{C}^{|\omega|}f(\sum_{j=1}^{s+1}x_j^{(\omega_j)})$$
$$= \mathbb{E}_{x\in Z;h\in Z^{s+1}}\prod_{\omega\in\{0,1\}^{s+1}}\mathcal{C}^{|\omega|}f(x+\sum_{j=1}^{s+1}\omega_j h_j).$$

Because the $U^{s+1}(Z)$ norm is derived from the box norm of dimension $s+1$, many properties of the latter norm automatically descend to the former norm. For instance, the $U^{s+1}(Z)$ norm is indeed a norm for $s \geqslant 1$, and from (B.4) we have the invariance

$$\|e(\phi)f\|_{U^{s+1}(Z)} = \|f\|_{U^{s+1}(Z)} \tag{B.11}$$

whenever $s \geqslant 1$ and $\phi : Z \to \mathbb{R}/\mathbb{Z}$ is an affine-linear phase or more generally a polynomial phase of degree at most s. In our applications we shall take Z to be a cyclic group $\mathbb{Z}_{N'}$, and our functions f shall usually be real-valued. Also, from Lemma B.2 we have the Gowers-Cauchy-Schwarz inequality for Z, which was first observed in [21] and reads as follows:

$$|\mathbb{E}_{x\in Z;h\in Z^{s+1}}\prod_{\omega\in\{0,1\}^{s+1}}\mathcal{C}^{|\omega|}f_\omega(x+\sum_{j=1}^{s+1}\omega_j h_j)| \leqslant \prod_{\omega\in\{0,1\}^{s+1}}\|f_\omega\|_{U^{s+1}(Z)}. \tag{B.12}$$

For technical reasons we shall need to localise the Gowers norms slightly. Let A be any finite non-empty subset of an additive group Z, which may or may not be finite. Then

for any $f : A \to \mathbb{C}$, we define the Gowers uniformity norm $\|f\|_{U^{s+1}(A)}$ by the formula

$$\begin{aligned}\|f\|_{U^{s+1}(A)}^{2^{s+1}} &= \mathbb{E}_{x^{(0)},x^{(1)}:\sum_{j=1}^{s+1} x_j^{(\omega_j)} \in A \ \forall \omega \in \{0,1\}^{s+1}} \prod_{\omega \in \{0,1\}^{s+1}} \mathcal{C}^{|\omega|} f(\sum_{j=1}^{s+1} x_j^{(\omega_j)}) \\ &= \mathbb{E}_{x,h:x+\sum_{j=1}^{s+1} \omega_j h_j \in A \ \forall \omega \in \{0,1\}^{s+1}} \prod_{\omega \in \{0,1\}^{s+1}} \mathcal{C}^{|\omega|} f(x + \sum_{j=1}^{s+1} \omega_j h_j).\end{aligned} \tag{B.13}$$

In the particular case $A = [N]$, which is used several times in the paper, we shall adopt the abbreviation

$$\|f\|_{U^{s+1}[N]} := \|f\|_{U^{s+1}([N])}.$$

If A is contained in a *finite* additive group Z, then these local Gowers norms are related to their global counterparts by the identity

$$\|f\|_{U^{s+1}(A)} = \|f1_A\|_{U^{s+1}(Z)} / \|1_A\|_{U^{s+1}(Z)} \tag{B.14}$$

for any $f : A \to \mathbb{C}$, where $f1_A : Z \to \mathbb{C}$ is the extension by zero of f from A to Z. The local norm $U^{s+1}(A)$ is also intrinsic in the following sense: if $A \subseteq Z$, $A' \subseteq Z'$, and $\phi : A \to A'$ is a *Freiman isomorphism* in the sense that it is 1-1 onto its image and for any $a_1, a_2, a_3, a_4 \in A$, we have $a_1+a_2 = a_3+a_4$ if and only if $\phi(a_1)+\phi(a_2) = \phi(a_3)+\phi(a_4)$, then we have $\|f \circ \phi\|_{U^{s+1}(A)} = \|f\|_{U^{s+1}(A')}$ for all $f : A' \to \mathbb{C}$. A particular consequence of this is the following lemma.

Lemma B.5 (Comparability of $U^{s+1}(I)$ and $U^{s+1}(\mathbb{Z}_{N'})$)**.** *Let $N' \geqslant 1$ be an integer, let $\alpha > 0$, and let $I = \{a, a+1, \dots, b\}$ be an interval of integers whose length satisfies $\alpha N' \leqslant |I| \leqslant N'/2$. Let $f : I \to \mathbb{C}$ be a function on I, and let $\tilde{f} : \mathbb{Z}_{N'} \to \mathbb{C}$ be the function formed from f by identifying I with a subset of $\mathbb{Z}_{N'}$ and setting $\tilde{f}(x) = 0$ for $x \notin I$. Then we have*

$$\|\tilde{f}\|_{U^{s+1}(\mathbb{Z}_{N'})} = c\|f\|_{U^{s+1}(I)} \tag{B.15}$$

where $c = c_{I,N',s} > 0$ is a constant which is independent of f, and which is bounded above and below by quantities depending only on α and s.

Proof. As $|I| \leqslant N'/2$, the interval $I \subseteq \mathbb{Z}$ is Freiman isomorphic to its counterpart in $\mathbb{Z}_{N'}$. The claim then follows from (B.14) together the easily confirmed observation that $\|1_I\|_{U^{s+1}(\mathbb{Z}_{N'})}$ is bounded above and below by quantities depending only on α and s. □

Remark. We will typically apply this lemma with $I = [N]$ and with N' comparable to a moderately large multiple of N. See, for example, the proof of Proposition 10.1.

Appendix C. Proof of the generalised von Neumann theorem

The purpose of this appendix is to prove Proposition 7.1.

Proposition 7.1 (Generalised von Neumann theorem)**.** *Let s, t, d, L be positive integer parameters as usual. Then there are constants C_1 and D, depending on s, t, d and L, such that the following is true. Let C, $C_1 \leqslant C \leqslant O_{s,t,d,L}(1)$, be arbitrary and suppose*

that $N' \in [CN, 2CN]$ is a prime. Let $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ be a D-pseudorandom measure, and suppose that $f_1, \dots, f_t : [N] \to \mathbb{R}$ are functions with $|f_i(x)| \leqslant \nu(x)$ for all $i \in [t]$ and $x \in [N]$. Suppose that $\Psi = (\psi_1, \dots, \psi_t)$ is a system of affine-linear forms in s-normal form with $\|\Psi\|_N \leqslant L$. Let $K \subseteq [-N, N]^d$ be a convex body such that $\Psi(K) \subseteq [N]^t$. Suppose also that

$$\min_{1 \leqslant j \leqslant t} \|f_j\|_{U^{s+1}[N]} \leqslant \delta$$

for some $\delta > 0$. Then we have

$$\sum_{n \in K} \prod_{i \in [t]} f_i(\psi_i(n)) = o_\delta(N^d) + \kappa(\delta) N^d. \tag{C.1}$$

Recall that this is a variant of [24, Proposition 5.3], which was proven by a long series of applications of the Cauchy-Schwarz inequality. We shall phrase our argument using Corollary B.3, but the argument is essentially that of [24, §5]. It is also necessary to perform some regularisation to deal with the convex body K, a technical feature not present in [24, Proposition 5.3].

Moving to a cyclic group. Let us first make some very minor reductions. We start by moving the whole problem to the group $\mathbb{Z}_{N'}$. We will always assume that $N' = O_{s,t,d,L}(N)$, but one may wish to take N' to be quite a bit larger than N in order that a pseudorandom measure ν can be constructed so as to make Proposition 7.1 applicable. We embed $[N]$ inside $\mathbb{Z}_{N'}$ in the usual manner, and extend the functions $f_1, \dots, f_t$ to all of $\mathbb{Z}_{N'}$ by defining them to be zero outside of $[N]$. From Lemma B.5 we then have

$$\|f_j\|_{U^{s+1}(\mathbb{Z}_{N'})} \ll_C \delta$$

for some $j \in \{1, \dots, t\}$. Similarly, we may identify the set $K \cap \mathbb{Z}^d$ with a subset K' of $\mathbb{Z}^d_{N'}$. We can also view Ψ as a map from $\mathbb{Z}^d_{N'}$ to $\mathbb{Z}^t_{N'}$. Note that Ψ will then map K' to $[N]^d$. To summarise, we have reduced matters to establishing the following.

Proposition 7.1′ (Transfer to $\mathbb{Z}_{N'}$)**.** *Let s, t, d, L be positive integer parameters as usual. Then there is a constant D, depending on s, t, d and L, such that the following is true. Let $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ be a D-pseudorandom measure, and suppose that $f_1, \dots, f_t : \mathbb{Z}_{N'} \to \mathbb{R}$ are functions with $|f_i(x)| \leqslant \nu(x)$ for all $i \in [t]$ and $x \in \mathbb{Z}_{N'}$. Suppose that $\Psi = (\psi_1, \dots, \psi_t)$ is a system of affine-linear forms in s-normal form with $\|\Psi\|_N \leqslant L$. Let $K' \subseteq \mathbb{Z}^d_{N'}$ be identified with $K \cap \mathbb{Z}^d$ for some convex $K \subseteq [-\frac{1}{4}N', \frac{1}{4}N']^d$. Suppose also that*

$$\min_{1 \leqslant j \leqslant t} \|f_j\|_{U^{s+1}(\mathbb{Z}_{N'})} \leqslant \delta$$

for some $\delta > 0$. Then we have

$$\mathbb{E}_{n \in \mathbb{Z}^d_{N'}} 1_{K'}(n) \prod_{i \in [t]} f_i(\psi_i(n)) = o_\delta(1) + \kappa(\delta). \tag{C.2}$$

Remark. Note the disappearance of C. This was an artefact of the relationship between N and N', which has now been forgotten.

From this point onwards we do our linear algebra over $\mathbb{Z}_{N'}$, rather than over $\mathbb{Q}$. Note that the notion of s-normal form coincides in the two settings provided that $N' \geqslant$

$N_0(s,t,d,L)$ is sufficiently large. Furthermore no two of the homogeneous parts $\dot{\psi}_i$ are parallel when considered (mod N'). This fact (which is very easily checked) is a simple instance of a kind of "Lefschetz principle".

Removing the convex cutoff. The next step is to partially eliminate the cutoff $1_{K'}(n)$ by replacing it by a more analytically tractable Lipschitz cutoff. We introduce a metric on $\mathbb{Z}^d_{N'}$ by declaring the distance between $(n_1,\dots,n_d)$ and $(m_1,\dots,m_d)$ to be $(\sum_{j=1}^d \|\frac{n_i-m_i}{N'}\|^2_{\mathbb{R}/\mathbb{Z}})^{1/2}$, where $\|x\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance to the nearest integer. This is the metric induced from the standard embedding of $\mathbb{Z}^d_{N'}$ into the torus $(\mathbb{R}/\mathbb{Z})^d$. To establish Proposition 7.1′, we claim that it suffices to establish the bound

$$\mathbb{E}_{n\in\mathbb{Z}^d_{N'}} F(n)\prod_{i\in[t]} f_i(\psi_i(n)) = o_{\delta,M}(1) + \kappa_M(\delta) \tag{C.3}$$

whenever $M>0$, $F:\mathbb{Z}^d_{N'}\to[-1,1]$ has Lipschitz constant M and the functions f_i are bounded pointwise by ν and satisfy $\min_{1\leqslant i\leqslant t}\|f_i\|_{U^{s+1}(\mathbb{Z}_{N'})}\leqslant\delta$. To see why, let $\varepsilon>0$ be a small quantity to be chosen later. It will suffice to prove that

$$\mathbb{E}_{n\in\mathbb{Z}^d_{N'}} 1_{K'}(n)\prod_{i\in[t]} f_i(\psi_i(n)) = \ o_\varepsilon(1) + \kappa_\varepsilon(\delta) + \kappa(\varepsilon),$$

as the claim then follows by setting ε to be a sufficiently slowly decaying function of δ.

To establish this bound, we apply Corollary A.3 to effect the decomposition

$$1_{K'}(n) = F_\varepsilon(n) + O(G_\varepsilon(n))$$

for all $n\in\mathbb{Z}^d_{N'}$, where $F_\varepsilon, G_\varepsilon : \mathbb{Z}^d_{N'}\to[0,1]$ are Lipschitz in the above metric with constant $O(1/\varepsilon)$. Furthermore, from the Lipschitz and integral bounds in Corollary A.3 we easily obtain the estimate

$$\mathbb{E}_{n\in\mathbb{Z}_{N'}} G_\varepsilon(n) = o_\varepsilon(1) + \kappa(\varepsilon). \tag{C.4}$$

Here we are basically using nothing more than the standard fact that Lipschitz functions are uniformly Riemann integrable. From (C.3) we have

$$\mathbb{E}_{n\in\mathbb{Z}^d_{N'}} F_\varepsilon(n)\prod_{i\in[t]} f_i(\psi_i(n)) = o_\varepsilon(1) + \kappa_\varepsilon(\delta)$$

and so by the triangle inequality and the fact that $|f_i(x)|\leqslant\nu(x)$ it is enough to show that

$$\mathbb{E}_{n\in\mathbb{Z}^d_{N'}} G_\varepsilon(n)\prod_{i\in[t]} \nu(\psi_i(n)) = o_\varepsilon(1) + \kappa(\varepsilon). \tag{C.5}$$

Now a standard application of the linear forms condition (see [24, Lemma 5.2]) gives

$$\|\nu - 1\|_{U^{s+1}(\mathbb{Z}_{N'})} = o(1).$$

Now the function $\frac{1}{2}(\nu-1)$ satisfies $\frac{1}{2}|\nu(x)-1|\leqslant\frac{1}{2}(\nu(x)+1)$, and this latter function is easily seen to be a pseudorandom measure (see [24, Lemma 3.4]). Thus from (C.3) we have

$$\mathbb{E}_{n\in\mathbb{Z}^d_{N'}} G_\varepsilon(n)\prod_{i\in[t]} g_i(\psi_i(n)) = o_\varepsilon(1)$$

whenever all the functions g_i are either 1 or $\nu - 1$, and not all of them are 1. When $g_i = 1$ for all i we have the bound $o_\varepsilon(1)+\kappa(\epsilon)$, from (C.4). The bound (C.5) now follows immediately upon writing $\nu = 1 + (\nu - 1)$ and expanding as a sum of 2^t terms.

It remains to prove (C.3). We now claim that we may dispense with the Lipschitz cutoff F entirely, and reduce to proving the estimate

$$\mathbb{E}_{n\in\mathbb{Z}^d_{N'}} \prod_{i\in[t]} f_i(\psi_i(n)) = \ o_\delta(1) + \kappa(\delta), \tag{C.6}$$

which involves no cutoff function at all. To see this, first observe that (C.6) implies the extension

$$\mathbb{E}_{n\in\mathbb{Z}^d_{N'}} e(m\cdot n/N) \prod_{i\in[t]} f_i(\psi_i(n)) = \ o_\delta(1) + \kappa(\delta). \tag{C.7}$$

for any frequency $m \in \mathbb{Z}^d_{N'}$. Indeed, if m lies in the span of $\dot\psi_1, \dots, \dot\psi_t$ then we may simply factor $e(m \cdot n/N)$ into terms that can be absorbed into the $f_1, \dots, f_t$ factors, noting that we can trivially extend (C.6) to cover the case when $f_1, \dots, f_t$ are complex-valued instead of real-valued. If m does not lie in this span, then it is easy to see that the left-hand side of (C.7) in fact vanishes.

Now we return to (C.3). Let $X > 0$ be arbitrary. By a standard Fourier-analytic argument, given in detail in [27, Lemma A.9], we may decompose

$$F(n) = \sum_{j=1}^{J} c_j e(m_j \cdot n/N) + O_d(M \log X/X)$$

where $J = O_d(X^d)$, $c_j = O(1)$ are coefficients, and $m_j \in \mathbb{Z}^d_{N'}$ are frequencies. Inserting this into (C.3), we have

$$\begin{aligned}&\mathbb{E}_{n\in\mathbb{Z}^d_{N'}} F(n) \prod_{i\in[t]} f_i(\psi_i(n))\\ &= \sum_{j=1}^{J} c_j \mathbb{E}_{n\in\mathbb{Z}^d_{N'}} e(m_j\cdot n/N) \prod_{i\in[t]} f_i(\psi_i(n)) + O_d\big(\frac{M\log X}{X}\big)\mathbb{E}_{n\in\mathbb{Z}^d_{N'}} \prod_{i\in[t]} \nu(\psi_i(n)).\end{aligned}$$

Using (C.7) to control the first term and the linear forms condition to estimate the second, we see that this is bounded by

$$O_d(X^d)(o_\delta(1) + \kappa(\delta)) + O_d\big(\frac{M\log X}{X}\big)(1 + o(1)).$$

Taking X to be a sufficiently slowly growing function of $1/\delta$ we obtain (C.3) as desired.

Main argument. It remains to prove (C.6). By symmetry we may assume that f_1 is the function with minimal U^{s+1} norm, thus

$$\|f_1\|_{U^{s+1}(\mathbb{Z}_{N'})} \leqslant \delta.$$

Recall that the system $\Psi : \mathbb{Z}^d \to \mathbb{Z}^t$ is in s-normal form. By permuting the basis vectors $e_1, \dots, e_d$ if necessary, we may then assume that $\prod_{j=1}^{s+1} \dot\psi_i(e_j)$ vanishes for $i \neq 1$ and is non-zero for $i = 1$.

In summary, we are reduced to proving

Proposition 7.1″ (Reduced generalised von Neumann theorem)**.** *Let s, t, d, L be positive integer parameters as usual. Then there is a constant D, depending on s, t, d and L, such that the following is true. Let $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ be a D-pseudorandom measure, and suppose that $f_1, \dots, f_t : \mathbb{Z}_{N'} \to \mathbb{R}$ are functions with $|f_i(x)| \leqslant \nu(x)$ for all $i \in [t]$ and $x \in \mathbb{Z}_{N'}$. Suppose that $\Psi = (\psi_1, \dots, \psi_t)$ is a system of affine-linear forms such that $\prod_{j=1}^{s+1} \dot{\psi}_i(e_j)$ vanishes for $i \neq 1$ and is non-zero for $i = 1$. Then we have*

$$\left|\mathbb{E}_{n \in \mathbb{Z}^d_{N'}} \prod_{i \in [t]} f_i(\psi_i(n))\right| \leqslant \|f_1\|_{U^{s+1}(\mathbb{Z}_{N'})} + o(1). \tag{C.8}$$

To prove the estimate (C.8), note first that the coefficients $\dot{\psi}_1(e_j)$, $j \in [s+1]$, are non-zero and bounded by $O_{s,t,d,L}(1)$, and hence are invertible in $\mathbb{Z}_{N'}$ provided that $N \geqslant N_0(s, t, d, L)$. Thus we may dilate[16] the first $s+1$ variables and assume that $\dot{\psi}_1(e_j) = 1$ for $j \in [s+1]$, a manoeuvre which affords a little notational simplicity if nothing more. With this normalisation we have, writing $n = (x_1, \dots, x_d)$ and $y = (x_{s+2}, \dots, x_d)$, that

$$\psi_1(x_1, \dots, x_{s+1}, y) = x_1 + \dots + x_{s+1} + \psi_1(0, y).$$

The other forms ψ_i, $i = 2, \dots, t$ do not involve all of the variables $x_1, \dots, x_{s+1}$, since the system Ψ is in normal form. This will be a crucial fact for us and to handle it we look, for each ψ_i, at the set $\Omega(i)$ of indices $j \in [s+1]$ for which $\dot{\psi}_i(e_j) \neq 0$, and then group the forms according to their associated set $\Omega(i)$. Thus $\Omega(1) = [s+1]$ and $\Omega(i) \subsetneq [s+1]$ for $i = 2, \dots, t$. Observe that the indices $j = s+2, \dots, d$ and the associated variable $y = (x_{s+2}, \dots, x_d)$ will be largely irrelevant in the sequel. With this nomenclature we may write the left-hand side of (C.8) as

$$|\mathbb{E}_{y \in \mathbb{Z}^{d-s-1}_{N'}} \mathbb{E}_{x_{[s+1]} \in \mathbb{Z}^{s+1}_{N'}} \prod_{B \subseteq [s+1]} F_{B,y}(x_B)| \tag{C.9}$$

where $x_{[s+1]} = (x_j)_{j \in [s+1]}$, x_B is the restriction of $x_{[s+1]}$ to B, and

$$F_{B,y}(x_B) := \prod_{i \in [t] : \Omega(i) = B} f_i(\psi_i(x_B, y)).$$

We have abused notation ever so slightly by regarding f_i as a function on $\mathbb{Z}^B_{N'} \times \mathbb{Z}^{d-s-1}_{N'}$ rather than on $\mathbb{Z}^{s+1}_{N'} \times \mathbb{Z}^d_{N'}$, supressing mention of the irrelevant variables x_j, $j \in [s+1] \setminus \Omega(i)$. Observe that

$$F_{[s+1],y}(x_{[s+1]}) = f_1(\psi_1(x_{[s+1]}, y)) = f_1(x_1 + \dots + x_{s+1} + \psi_1(0, y)).$$

Now we have the pointwise bounds $|F_{B,y}(x_B)| \leqslant \nu_{B,y}(x_B)$, where

$$\nu_{B,y}(x_B) := \prod_{i \in [t] : \Omega(i) = B} \nu(\psi_i(x_B, y)).$$

[16] This dilation converts the coefficients from bounded integers, to rationals with bounded numerator and denominator. However, when the time comes to apply the linear forms condition, one can clear denominators and reduce back to estimates involving only bounded integers again.

Invoking Corollary B.4, we may bound (C.9) by

$$\mathbb{E}_{y\in\mathbb{Z}_{N'}^{d-s-1}}\|F_{[s+1],y}\|_{\square(\nu_{[s+1],y};\mathbb{Z}_{N'}^{[s+1]})}\prod_{B\subsetneq[s+1]}\|\nu_{B,y}\|_{\square(\nu_{B,y};\mathbb{Z}_{N'}^{B})}^{1/2^{s+1-|B|}}.$$

The reader may wish to recall the definition of the quantities appearing here, which are provided in the statement of Corollary B.4.

Applying Hölder's inequality[17], we see that to show (C.8) it suffices to show that

$$\mathbb{E}_{y\in\mathbb{Z}_{N'}^{d-s-1}}\|F_{[s+1],y}\|_{\square^{s+1}(\nu_{[s+1],y};\mathbb{Z}_{N'}^{[s+1]})}^{2^{s+1}}\leqslant\|f_1\|_{U^{s+1}(\mathbb{Z}_{N'})}+o(1) \tag{C.10}$$

and that

$$\mathbb{E}_{y\in\mathbb{Z}_{N'}^{d-s-1}}\|\nu_{B,y}\|_{\square(\nu_{\subsetneq B,y})}^{2^{|B|}}=1+o(1) \tag{C.11}$$

for all non-empty $B\subsetneq[s+1]$. Note that except for f_1, the unknown functions $f_2,\dots,f_t$ have all been eliminated. This procedure will be familiar to readers who have looked at (for example) [24, Ch. 5].

We begin with (C.11). We expand the left-hand side, obtaining

$$\begin{aligned}\mathbb{E}_{y\in\mathbb{Z}_{N'}^{d-s-1}}\|\nu_{B,y}\|_{\square^{B}(\nu_{\subsetneq B,y})}^{2^{|B|}}&=\mathbb{E}_{x_B^{(0)},x_B^{(1)}\in X_B}\prod_{C\subseteq B}\prod_{\omega_C\in\{0,1\}^C}\nu_{C,y}(x_C^{(\omega_C)})\\&=\mathbb{E}_{x_B^{(0)},x_B^{(1)}\in X_B}\prod_{C\subseteq B}\prod_{\omega_C\in\{0,1\}^C}\prod_{i\in[t]:\Omega(i)=C}\nu(\psi_i(x_C^{(\omega_C)},y)).\end{aligned}$$

Because of the definition of $\Omega(i)$, and the hypothesis that no two of the ψ_i were affine-linear combinations of each other, we see that the affine-linear forms

$$(x_B^{(0)},x_B^{(1)},y)\mapsto\psi_i(x_C^{(\omega_C)},y),$$

as C varies over subsets of B and i varies over those $i\in[t]$ such that $\Omega(i)=C$, also have the property that no two forms are affine-linear combinations of each other. In other words, this system has finite complexity. Thus (C.11) will follow from the linear forms condition (6.2) provided that the degree D of pseudorandomness is sufficiently large.

Now we turn to (C.10). The left-hand side expands as

$$\begin{aligned}\mathbb{E}_{x_{[s+1]}^{(0)},x_{[s+1]}^{(1)}\in\mathbb{Z}_{N'}^{s+1};y\in\mathbb{Z}_{N'}^{d-s-1}}\prod_{\omega\in\{0,1\}^{s+1}}&f_1(\sum_{j=1}^{s+1}x_i^{(\omega_j)}+\psi_1(0,y))\times\\&\times\prod_{C\subsetneq[s+1]}\prod_{\omega_C\in\{0,1\}^C}\prod_{i\in[t]:\Omega(i)=C}\nu(\psi_i(x_C^{(\omega_C)},y)).\end{aligned}$$

[17] This is really an application of the Cauchy-Schwarz inequality several times, since the exponent is a power of two.

Substituting $h := x^{(1)}_{[s+1]} - x^{(0)}_{[s+1]}$ and $z := x^{(0)}_1 + \ldots + x^{(0)}_{s+1} + \psi_1(0,y)$, we may rewrite this as

$$\mathbb{E}_{x^{(0)}_{[s+1]}, h \in \mathbb{Z}^{s+1}_{N'}; y \in \mathbb{Z}^{d-s-1}_{N'}} \prod_{\omega \in \{0,1\}^{s+1}} f_1(z + \sum_{j=1}^{s+1} \omega_j h_j) \prod_{C \subsetneq [s+1]} \prod_{\omega_C \in \{0,1\}^C} \prod_{i \in [t]: \Omega(i) = C} \nu\big(\psi_i(x^{(0)}_C, y) + \\ + \sum_{j \in C} \omega_j \dot{\psi}_i(e_j) h_j\big).$$

Observe that for fixed h, the map $(x^{(0)}_{[s+1]}, y) \mapsto z$ is uniform, in the sense that each z is mapped to by exactly $(N')^{d-1}$ preimages. Thus we may rewrite the preceding expression as

$$\mathbb{E}_{z \in \mathbb{Z}_{N'}; h \in \mathbb{Z}^{s+1}_{N'}} W(z,h) \prod_{\omega \in \{0,1\}^{s+1}} f_1(z + \sum_{j=1}^{s+1} \omega_j h_j)$$

where

$$W(z,h) := \mathbb{E}_{\substack{x^{(0)}_{[s+1]} \in \mathbb{Z}^{s+1}_{N'}; y \in \mathbb{Z}^{d-s-1}_{N'} \\ z = x^{(0)}_1 + \ldots + x^{(0)}_{s+1} + \psi_1(0,y)}} \prod_{C \subsetneq [s+1]} \prod_{\omega_C \in \{0,1\}^C} \prod_{i \in [t]: \Omega(i) = C} \nu(\psi_i(x^{(0)}_C, y) + \sum_{j \in C} \omega_j \dot{\psi}_i(e_j) h_j).$$

Comparing this with (B.13), we see that to prove (C.10) it suffices to show that

$$\mathbb{E}_{z \in \mathbb{Z}_{N'}; h \in \mathbb{Z}^{s+1}_{N'}} (W(z,h) - 1) \prod_{\omega \in \{0,1\}^{s+1}} f_1(z + \sum_{j=1}^{s+1} \omega_j h_j) = o(1).$$

By Cauchy-Schwarz and the hypothesis $|f_1(x)| \leqslant \nu(x)$, it suffices to establish the estimates

$$\mathbb{E}_{z \in \mathbb{Z}_{N'}; h \in \mathbb{Z}^{s+1}_{N'}} |W(z,h) - 1|^n \prod_{\omega \in \{0,1\}^{s+1}} \nu(z + \sum_{j=1}^{s+1} \omega_j h_j) = 0^n + o(1)$$

for $n = 0$ and $n = 2$. Expanding, we reduce to showing that

$$\mathbb{E}_{z \in \mathbb{Z}_{N'}; h \in \mathbb{Z}^{s+1}_{N'}} W(z,h)^n \prod_{\omega \in \{0,1\}^{s+1}} \nu(z + \sum_{j=1}^{s+1} \omega_j h_j) = 1 + o(1)$$

for $n = 0, 1, 2$.

This will follow from the linear forms condition. We shall just verify the case $n = 2$, as the cases $n = 0, 1$ follow from that case (they utilise a subset of the linear forms that are used in the $n = 2$ case). When $n = 2$, we can expand out the left-hand side as

$$\mathbb{E}_* \big(\prod_{\omega \in \{0,1\}^{s+1}} \nu(z + \sum_{j=1}^{s+1} \omega_j h_j) \big) \times \\ \prod_{C \subsetneq [s+1]} \prod_{\omega_C \in \{0,1\}^C} \prod_{i \in [t]: \Omega(i) = C} \nu\big(\psi_i(x^{(0)}_C, y) + \sum_{j \in C} \omega_j \dot{\psi}_i(e_j) h_j\big) \nu\big(\psi_i(\tilde{x}^{(0)}_C, \tilde{y}) + \sum_{j \in C} \omega_j \dot{\psi}_i(e_j) h_j\big) \tag{C.12}$$

where the average $\mathbb{E}_*$ is over all sextuples

$$(z, h, x^{(0)}_{[s+1]}, \tilde{x}^{(0)}_{[s+1]}, y, \tilde{y}) \in \mathbb{Z}_{N'} \times \mathbb{Z}^{s+1}_{N'} \times \mathbb{Z}^{s+1}_{N'} \times \mathbb{Z}^{s+1}_{N'} \times \mathbb{Z}^{d-s-1}_{N'} \times \mathbb{Z}^{d-s-1}_{N'}$$

subject to the affine constraints

$$z = x_1^{(0)} + \ldots + x_{s+1}^{(0)} + \psi_1(0,y) = \tilde{x}_1^{(0)} + \ldots + \tilde{x}_{s+1}^{(0)} + \psi_1(0,\tilde{y}). \tag{C.13}$$

Naturally, we wish to apply the linear forms condition, on the assumption that ν is D-pseudorandom for sufficiently large D. To do this we must first eliminate the constraints (C.13). To do this, we substitute for $x_{s+1}^{(0)}$ and $\tilde{x}_{s+1}^{(0)}$ in terms of the other variables, that is to say we write

$$x_{s+1}^{(0)} = z - x_1^{(0)} - \cdots - x_s^{(0)} - \psi_1(0,y)$$

and

$$\tilde{x}_{s+1}^{(0)} = z - \tilde{x}_1^{(0)} - \cdots - \tilde{x}_s^{(0)} - \psi_1(0,\tilde{y}).$$

In this way we may rewrite (C.12) as an unconstrained average over the $2d+s-1$ variables $h, x_{[s]}^{(0)}, \tilde{x}_{[s]}^{(0)}, y, \tilde{y}$.

When written in terms of this set of variables, it is clear that all the linear forms in (C.12) have integer coefficients which are bounded in terms of s, t, d and L. To apply the linear forms condition, all we must do is satisfy ourselves that no two of these forms are affinely dependent, that is to say no two of them have parallel homogeneous parts.

To see this, first observe that the 2^{s+1} homogeneous forms $z + \sum_{j=1}^{s+1} \omega_j h_j$ are pairwise distinct, and that they are also different from any other form appearing in (C.12) even after performing the above substitutions, because the latter forms all involve at least one of the variables from $x_{[s]}^{(0)}$, $\tilde{x}_{[s]}^{(0)}$ (here we are using the fact that C is a *proper* subset of $[s+1]$).

Now consider an affine form $\psi_i(x_C^{(0)}, y) + \sum_{j\in C} \omega_j \dot{\psi}_i(e_j) h_j$ appearing in (C.12). Recalling that $C = \Omega(i)$, the set of all j for which $\dot{\psi}_i(e_j) \neq 0$, we see that in our new system of variables this form may be written as the slightly alarming expression

$$\sum_{j=1}^{s} \dot{\psi}_i(e_j)(x_j^{(0)} + \omega_j h_j) + \dot{\psi}_i(e_{s+1})(z - x_1^{(0)} - \cdots - x_s^{(0)} - \psi_1(0,y) + \omega_{s+1}h_{s+1}) + \psi_i(0,y). \tag{C.14}$$

There is a similar expression involving tildes. We claim first of all that at least one of the variables $x_1^{(0)}, \ldots, x_s^{(0)}$ must appear with non-zero coefficient. If this were not the case then we would have $\dot{\psi}_i(e_j) = \dot{\psi}_i(e_{s+1})$ for $j \leqslant s$ and hence, since $C \subsetneq [s+1]$, C is empty. Hence so is the product over $\omega_C \in \{0,1\}^C$ in (C.12). Thus no form (C.14) with this property appears in (C.12), thereby confirming the claim.

The claim just proved immediately implies that no form (C.14) has homogeneous part parallel to that of a form with tildes. It remains to prove that the forms in (C.14) have pairwise non-parallel homogeneous parts.

Suppose that we are given a form (C.14) written as

$$q\big(r_1 x_1^{(0)} + \cdots + r_s x_s^{(0)} + r'z + (\text{terms involving } h \text{ and } y)\big)$$

where $q \neq 0$. We claim that the set C from which the form came may be identified knowing only $r_1, \dots, r_s, r'$. Indeed we must have $qr' = \dot{\psi}_i(e_{s+1})$, whence $\dot{\psi}_i(e_j) = \lambda(r_j + r')$ for $j \leqslant s$. The set C may be found simply by looking at which of these quantities do not vanish. It is immediately clear that $\omega_C \in \{0,1\}^C$ may also be recovered.

The only way in which two forms (C.14) could have parallel homogeneous parts, then, is if there is some fixed choice of ω, some $i \neq i'$ and some rational $q, q' \neq 0$ such that

$$q\big(\sum_{j=1}^{s} \dot{\psi}_i(e_j)(x_j^{(0)} + \omega_j h_j) + \dot{\psi}_i(e_{s+1})(z - \sum_{j=1}^{s} x_j^{(0)} - \psi_1(0,y) + \omega_{s+1}h_{s+1}) + \psi_i(0,y)\big)$$
$$= q'\big(\sum_{j=1}^{s} \dot{\psi}_{i'}(e_j)(x_j^{(0)} + \omega_j h_j) + \dot{\psi}_{i'}(e_{s+1})(z - \sum_{j=1}^{s} x_j^{(0)} - \psi_1(0,y) + \omega_{s+1}h_{s+1}) + \psi_{i'}(0,y)\big)$$

for all choices of the variables. After some simple manipulations one confirms that $q\dot{\psi}_i(e_j) = q'\dot{\psi}_{i'}(e_j)$ for $j \leqslant s+1$ and that $q\psi_i(0,y) = q'\psi_{i'}(0,y)$. Thus $\dot{\psi}_i$ is parallel to $\dot{\psi}_{i'}$, contrary to the assumption that the system $\Psi = (\psi_i)_{i=1}^t$ has finite complexity.

We have verified that it was valid to invoke the linear forms condition, provided that D is large enough. This completes the proof of Proposition 7.1' and hence that of Proposition 7.1. □

Appendix D. Goldston-Yildirim correlation estimates

One aim of this section is to construct a pseudorandom measure ν such that a suitable multiple of ν majorises the modified von Mangoldt function $\Lambda'_{b,W}$. Specifically, we will prove Proposition 6.4. This was essentially carried out in [24, Chs. 9, 10], building on work of Goldston and Yıldırım [16, 17, 18, 19], but the argument there only led to a majorant for *one* function $\Lambda'_{b,W}$, whereas in the present work we need to simultaneously majorise $\Lambda'_{b_1,W}, \dots, \Lambda'_{b_t,W}$. A few small modifications to the argument in [24] would, however, achieve this. Another aim of this section is to prove (12.5), a crucial estimate on the Gowers norm of a certain truncated von Mangoldt function $\Lambda^\sharp$. This does not follow immediately from the results in [24], though can be proved using similar ideas. We take the opportunity to give a brief but more-or-less self-contained account of these ideas here, while also providing some simplifications.

The heart of the matter is the establishment of correlation estimates for truncated divisor sums $\Lambda_{\chi,R,a} : \mathbb{Z} \to \mathbb{R}$ of the form

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

In this expression R is a moderately large number, which in practice will be a small power of N, $\chi : \mathbb{R} \to \mathbb{R}$ is a smooth, compactly supported function, and $a \in \mathbb{N}$. In our applications we only ever take $a = 1$ or $a = 2$. We extend $\Lambda_{\chi,R,a}$ to the negative

numbers in the obvious manner. Indeed, the compact support of χ ensures that $\Lambda_{\chi,R,a}$ is periodic.

Remark. Observe that $\Lambda_{\chi,R,a} = \chi(0)^a \log R$ on "almost primes" - numbers coprime to $\prod_{p \leqslant R} p$. For the purposes of gaining intuition about these functions one might think of them heuristically as being weights on the almost primes, though they do also have some weight on other numbers. The reason we need to deal with $\Lambda_{\chi,R,2}(n)$ is to correct for the rather unfortunate fact that $\Lambda_{\chi,R,1}(n)$ can be negative. This trick is of course closely related to the Λ^2 sieve of Selberg.

Associated to these truncated divisor sums are certain numbers which we call *sieve factors*.

Definition D.1 (Sieve factors)**.** Let $\chi : \mathbb{R} \to \mathbb{R}$ be compactly supported and suppose that $a \geqslant 1$. Then we define the *sieve factor* $c_{\chi,a}$ by the formula

$$c_{\chi,a} := \int_{\mathbb{R}} \dots \int_{\mathbb{R}} \prod_{B \subseteq [a]} \big(\sum_{j \in B} (1 + i\xi_j)\big)^{(-1)^{|B|-1}} \prod_{j=1}^{a} \varphi(\xi_j)\ d\xi_j, \tag{D.1}$$

where φ is the modified Fourier transform of χ, defined by the formula

$$e^x \chi(x) = \int_{-\infty}^{\infty} \varphi(\xi) e^{-ix\xi}\ d\xi. \tag{D.2}$$

The sieve factor $c_{\chi,a}$ looks very complicated (though explicitly computable), but in the special cases $a = 1, 2$ it has a particularly simple form:

Lemma D.2. *We have $c_{\chi,1} = -\chi'(0)$ and $c_{\chi,2} = \int_0^\infty |\chi'(x)|^2\ dx$. More generally, $c_{\chi,a}$ is a real number.*

Proof. We deal first with the case $a = 1$. From (D.1) and (D.2) we have

$$c_{\chi,1} = \int_{\mathbb{R}} (1 + i\xi)\varphi(\xi)\ d\xi = \chi(0) - \frac{d}{dx}(e^x \chi(x))|_{x=0}$$

and the claim follows. Now we handle the case $a = 2$. We have

$$c_{\chi,2} = \int_{\mathbb{R}} \int_{\mathbb{R}} \frac{(1 + i\xi)(1 + i\xi')}{2 + i(\xi + \xi')} \varphi(\xi)\varphi(\xi')\ d\xi d\xi'$$

Using the identity

$$\frac{1}{2 + i(\xi + \xi')} = \int_0^\infty e^{-(1+i\xi)x} e^{-(1+i\xi')x}\ dx$$

we can rewrite $c_{\chi,2}$ as

$$\int_0^\infty \left(\int_{\mathbb{R}} \varphi(\xi)(1 + i\xi) e^{-(1+i\xi)x}\ d\xi \right)^2 dx.$$

But from differentiating (D.2) we see that the expression in parentheses is $-\chi'(x)$, and the claim follows.

Finally, for general a, we observe that since χ is real, we have $\varphi(-\xi) = \overline{\varphi(\xi)}$. Taking complex conjugates of (D.1) and substituting $\xi_j \mapsto -\xi_j$ we obtain the claim. □

Roughly speaking, we will be able to show the analogue of the generalised Hardy-Littlewood conjecture for these sums $\Lambda_{\chi,R,a}$ so long as χ is suitably smooth and R is a sufficiently small power of N. More precisely, we prove the following.

Theorem D.3 (Goldston-Yıldırım estimate)**.** *Let t, d, L be positive integers, let N be a large positive integer as usual, and let $\Psi = (\psi_1, \dots, \psi_t)$ be a system of affine-linear forms with $\|\Psi\|_N \leqslant L$. Assume that no two of the forms ψ_i are rational multiples of one another. Let $a = (a_1, \dots, a_t) \in \mathbb{N}^t$ be a t-tuple of integers. Let $K \subseteq [-N, N]^d$ be a convex body, and let $\chi_1, \dots, \chi_t : \mathbb{R} \to \mathbb{R}$ be smooth, compactly supported functions. Let $R = N^\gamma$, where $\gamma > 0$ is sufficiently small depending on t, d, L, χ and a. Call a prime p* exceptional *if there exist two forms ψ_i, ψ_j which are linearly dependent modulo p, and let P_Ψ denote the set of all exceptional primes. Write $X := \sum_{p \in P_\Psi} p^{-1/2}$. Then we have*

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i \in [t]} \Lambda_{\chi_i,R,a_i}(\psi_i(n)) = \prod_{i \in [t]} c_{\chi_i,a_i} \cdot \mathrm{vol}_d(K) \cdot \prod_p \beta_p + O(\frac{N^d}{\log^{1/20} R} e^{O(X)}), \qquad \text{(D.3)}$$

where the local factors β_p for each prime p were defined in (1.4), (1.6)*, and the sieve factors $c_{\chi,a}$ were defined in Definition D.1. The implied constants here can depend on $t, d, L, \chi_1, \dots, \chi_t$ and a.*

Remarks. Note that we are *not* assuming that the system Ψ has finite complexity but, as stated, we do assume that no two of the forms ψ_i are rational multiples of one another. This means that P_Ψ is finite but not necessarily bounded in terms of t, d, L. If, for example, we have $d = 1$, $t = 2$ and $\psi_1(n) = n$, $\psi_2(n) = n + M$, then P_Ψ can be somewhat large if M has many prime factors. If Ψ does have finite complexity then X is bounded in terms of t, d, L and the error term becomes $o(N^d)$. In other situations this term can be more substantial. We have not attempted to find an error term which is best possible, being happy to settle for one that suffices for our application, and in particular for the correlation condition (Definition 6.3).

Theorem D.3 should be compared with Conjecture 1.2. The space $\Psi^{-1}((\mathbb{R}^+)^t)$, which appears in (1.4), is not present here because the truncated divisor sums Λ_{χ_i,R,a_i} extend periodically to the negative numbers, in contrast to the von Mangoldt function Λ.

Remark. In the works of Goldston, Pintz and Yıldırım [16, 17, 18, 19] the choice of cutoff χ was critically important. In our analysis it is not, ultimately because the inverse Gowers-norm conjecture GI(s) applies even for arbitrarily small $\delta > 0$. This allows us to use simpler and smoother enveloping sieves in which the sieve factors are large. We do, of course, require these factors to be independent of N. In taking χ to be very smooth, a number of simplifications are possible. Following notes of the second author [42, 43] (see also [31]), we avoid the use of any deep facts from analytic number theory such as the classical zero-free region for the Riemann zeta function. One may instead make do with the elementary observation that the Riemann zeta function $\zeta(s)$ has the asymptotic $\zeta(s) = \frac{1}{s-1} + O(1)$ for s near 1 and $\Re(s) > 1$. We note that these simplifications could also be applied (retrospectively) to Chapters 9 and 10 of [24].

Remark. Observe that if $R = N^\gamma$, then $0 \leqslant \Lambda'(n) \leqslant \frac{1}{\gamma\chi(0)^2}\Lambda_{\chi,R,2}(n)$ for all n, $R < n \leqslant N$. Thus we can use Theorem D.3 to obtain upper bounds for the expression (1.7) which lose a multiplicative factor of $\left(\frac{c_{\chi,2}}{\gamma\chi(0)^2}\right)^t$, which is independent of N. This observation,

coupled with a good choice of χ and γ, is rather close to the Selberg Λ^2 sieving technique. As is well-known there are significant barriers (the "parity problem") to reducing this multiplicative loss to something approaching 1.

PROOF OF THEOREM D.3. To simplify the notation we allow all implicit constants to depend on $t, d, L, \chi_1, \dots, \chi_t$ and a. We may assume that N (and hence R) are large with respect to these parameters, as the claim is trivial otherwise.

It is convenient to introduce the index set

$$\Omega := \{(i,j) : i \in [t]; j \in [a_i]\} \subseteq \mathbb{N}^2.$$

With this notation, it is a simple matter to expand the left-hand side of (D.3) as

$$\log^t R \sum_{(m_{i,j})_{(i,j)\in\Omega} \in \mathbb{N}^\Omega} \Big(\prod_{(i,j)\in\Omega} \mu(m_{i,j}) \chi_i\big(\frac{\log m_{i,j}}{\log R}\big)\Big) \sum_{n \in K \cap \mathbb{Z}^d} \prod_{(i,j)\in\Omega} 1_{m_{i,j}|\psi_i(n)}.$$

The μ factors allow us to restrict $m_{i,j}$ to $\mathbb{N}_*$, the set of square-free natural numbers. If, for each $i \in [t]$, we set $m_i := \mathrm{lcm}(m_{i,1}, \dots, m_{i,a_i})$, then we can rewrite the above expression as

$$\log^t R \sum_{(m_{i,j})_{(i,j)\in\Omega} \in \mathbb{N}_*^\Omega} \Big(\prod_{(i,j)\in\Omega} \mu(m_{i,j}) \chi_i\big(\frac{\log m_{i,j}}{\log R}\big)\Big) \sum_{n \in K \cap \mathbb{Z}^d} \prod_{i\in[t]} 1_{m_i|\psi_i(n)}.$$

Since χ is compactly supported we may restrict m_i to be at most $R^{O(1)}$ for all i. In particular if we set $m := \prod_{i\in[t]} m_i$ then $m \leqslant R^{O(1)}$ also. From the Chinese remainder theorem we see that as a function of n, the expression $\prod_{i\in[t]} 1_{m_i|\psi_i(n)}$ is periodic with respect to the lattice $m \cdot \mathbb{Z}^d$. By a volume packing argument similar to that used to prove (1.3) in Appendix A, we have

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{i\in[t]} 1_{m_i|\psi_i(n)} = \mathrm{vol}_d(K) \alpha_{m_1,\dots,m_t} + O(mN^{d-1})$$

where $\alpha_{m_1,\dots,m_t}$ is the local factor

$$\alpha_{m_1,\dots,m_t} := \mathbb{E}_{n \in \mathbb{Z}_m^d} \prod_{i\in[t]} 1_{m_i|\psi_i(n)}.$$

The total contribution of the error term $O(mN^{d-1})$ to (D.3) can be estimated crudely by $O(R^{O(1)} N^{d-1} \log^t R)$, which will be $o(N^d)$ if the exponent γ that defines R is sufficiently small. Thus we can discard this term and reduce our task to that of showing that

$$\log^t R \sum_{(m_{i,j})_{(i,j)\in\Omega} \in \mathbb{N}_*^\Omega} \Big(\prod_{(i,j)\in\Omega} \mu(m_{i,j}) \chi_i\big(\frac{\log m_{i,j}}{\log R}\big)\Big) \alpha_{m_1,\dots,m_t} = \prod_{i\in[t]} c_{\chi_i,a_i} \cdot \prod_p \beta_p + O(e^{O(X)} \log^{-1/20} R). \tag{D.4}$$

Note that we have eliminated the convex body K and the scale parameter[18] N. From the Chinese remainder theorem we make the key observation that $\alpha_{m_1,\dots,m_t}$ is multiplicative

[18]Observe that, although the o-notation concerns the situation when $N \to \infty$, this is exactly the same as letting $R \to \infty$.

in $m_1, \dots, m_t$, so that if we decompose $m_i = \prod_p p^{r_{p,i}}$ then

$$\alpha_{m_1,\dots,m_t} = \prod_p \alpha_{p^{r_{p,1}},\dots,p^{r_{p,t}}}. \tag{D.5}$$

Note that as the $m_{i,j}$ are square-free, the $r_{p,i}$ are either 0 or 1.

The next step is to use Fourier expansion to replace the weights χ_i by more multiplicative functions. Indeed, as χ_i is smooth and compactly supported we have the Fourier expansion (D.2) for some smooth φ_i which is rapidly decreasing in the sense that $|\varphi_i(\xi)| \ll_A (1+\xi)^{-A}$ for all $A > 0$. Thus we have

$$\chi_i\big(\frac{\log m_{i,j}}{\log R}\big) = \int_{-\infty}^{\infty} m_{i,j}^{-\frac{1+i\xi}{\log R}} \varphi_i(\xi)\, d\xi.$$

We could insert this Fourier expansion into (D.4) directly, but it will be easier if we first take advantage of the rapid decrease of φ_i to truncate the Fourier integral to the interval $I := \{\xi \in \mathbb{R} : |\xi| \leqslant \log^{1/2} R\}$ (say), thereby obtaining

$$\chi_i\big(\frac{\log m_{i,j}}{\log R}\big) = \int_I m_{i,j}^{-\frac{1+i\xi}{\log R}} \varphi_i(\xi)\, d\xi + O_A(m_{i,j}^{-1/\log R} \log^{-A} R)$$

for any A. Since $\chi_i(\log m_{i,j}/\log R)$ is itself bounded by $O(m_{i,j}^{-1/\log R})$, we conclude that

$$\prod_{(i,j)\in\Omega} \chi_i\big(\frac{\log m_{i,j}}{\log R}\big) = \int_I \dots \int_I \prod_{(i,j)\in\Omega} m_{i,j}^{-z_{i,j}} \varphi_i(\xi_{i,j})\, d\xi_{i,j} + O_A\big(\log^{-A} R \prod_{(i,j)\in\Omega} m_{i,j}^{-1/\log R}\big), \tag{D.6}$$

where we have written $z_{i,j} := (1 + i\xi_{i,j})/\log R$. Let us first deal with the contribution of the error term $O_A(\log^{-A} R \prod_{(i,j)\in\Omega} m_{i,j}^{-1/\log R})$ to (D.4). Taking absolute values everywhere, we can bound these contributions by

$$\ll_A (\log R)^{O(1)-A} \sum_{(m_{i,j})_{(i,j)\in\Omega}\in\mathbb{N}^\Omega} \alpha_{m_1,\dots,m_t} \prod_{(i,j)\in\Omega} m_{i,j}^{-1/\log R}.$$

Using the multiplicativity, we can factorise this expression as an Euler product

$$(\log R)^{O(1)-A} \prod_p \sum_{(r_{i,j})_{(i,j)\in\Omega}\in\mathbb{N}^\Omega} \alpha_{p^{r_1},\dots,p^{r_t}} p^{-(\sum_{(i,j)\in\Omega} r_{i,j})/\log R}$$

where $r_i := \max(r_{i,1}, \dots, r_{i,a_i})$. Crude computations then show that $\alpha_{p^{r_1},\dots,p^{r_t}}$ is equal to 1 when $r_1 = \dots = r_t = 0$ and $O(1/p)$ otherwise (cf. Lemma 1.3) and hence we can bound the above expression by

$$(\log R)^{O(1)-A} \prod_p (1 - \frac{1}{p^{1+1/\log R}})^{-O(1)}.$$

Since the Riemann zeta function $\zeta(s) = \prod_p (1 - \frac{1}{p^s})^{-1}$ has a simple pole at $s = 1$ with residue 1, we see that

$$\prod_p (1 - \frac{1}{p^s})^{-1} = \frac{1}{s-1} + O(1) \tag{D.7}$$

whenever $\Re(s) > 1$ and $s - 1$ is sufficiently close to 1. This allows us to bound the above expression by

$$O_A((\log R)^{O(1)-A})$$

which will be acceptable if A is large enough. Thus we only need to deal with the contribution of the main term of (D.6) to (D.4). After swapping sums and integrals[19], we write this term as

$$\log^t R \int_I \dots \int_I \sum_{(m_{i,j})_{(i,j)\in\Omega}\in\mathbb{N}_*^\Omega} \prod_{(i,j)\in\Omega} \mu(m_{i,j}) m_{i,j}^{-z_{i,j}} \alpha_{m_1,\dots,m_t} \varphi_i(\xi_{i,j})\, d\xi_{i,j}.$$

Using the multiplicativity of α once more, we can write this expression as

$$\log^t R \int_I \dots \int_I \prod_p E_{p,\xi} \cdot \prod_{(i,j)\in\Omega} \varphi_i(\xi_{i,j})\, d\xi_{i,j}, \tag{D.8}$$

where $\xi = (\xi_{i,j})_{(i,j)\in\Omega} \in I^\Omega$ and $E_{p,\xi}$ is the Euler factor

$$E_{p,\xi} := \sum_{(m_{i,j})_{(i,j)\in\Omega}\in\{1,p\}^\Omega} \Big(\prod_{(i,j)\in\Omega} \mu(m_{i,j}) m_{i,j}^{-z_{i,j}} \alpha_{m_1,\dots,m_t} \Big).$$

Our task is to show that (D.8) is $\big(\prod_{i\in[t]} c_{\chi_i,a_i}\big) \prod_p \beta_p + O(e^{O(X)} \log^{-1/20} R)$. To tackle this we must understand the Euler factors $E_{p,\xi}$. We may rewrite this expression as

$$E_{p,\xi} = \sum_{B\subseteq\Omega} (-1)^{|B|} \frac{\alpha(p,B)}{p^{\sum_{(i,j)\in B} z_{i,j}}}. \tag{D.9}$$

In this expression $\alpha(p,B) := \alpha_{p^{r_1},\dots,p^{r_t}}$, where $r_i := 1$ whenever $(i,j) \in B$ for at least one j, and $r_i := 0$ otherwise. Note that $\alpha(p,\emptyset) = 1$.

Call a set $B \subseteq \Omega$ *vertical* if it is non-empty and contained inside a vertical fibre $\{i\} \times [a_i]$ for some $i \in [t]$. If B is vertical then $\alpha(p,B) = \mathbb{E}_{n\in\mathbb{Z}_p^d} 1_{p|\psi_i(n)}$, which is equal to $1/p$ if $p \geqslant p_0(t,d,L)$ is sufficiently large. To say something about $\alpha(p,B)$ when B is neither empty nor vertical, recall that we described a prime p as exceptional, and wrote $p \in P_\Psi$, if there exist i, i' such that ψ_i is a multiple of $\psi_{i'}$ in $\mathbb{Z}_p$. For $p \notin P_\Psi$, we see from Lemma 1.3 that $\alpha(p,B) = O(1/p^2)$ whenever B is not vertical or empty. If $p \in P_\psi$ then the best we can say in general is that $\alpha(p,B) = O(1/p)$.

From the above discussion we have

$$E_{p,\xi} = (1 + O(1/p^2)) E'_{p,\xi} \quad \text{for } p \notin P_\Psi, \tag{D.10}$$

where $E'_{p,\xi}$ is the Euler factor[20]

$$E'_{p,\xi} := \prod_{B\subseteq\Omega, B \text{ vertical}} \left(1 - \frac{1}{p^{1+\sum_{(i,j)\in B} z_{i,j}}}\right)^{(-1)^{|B|-1}}. \tag{D.11}$$

[19]One can justify the exchange of integrals and summations because I is compact, and the summation can be shown to be absolutely convergent, either by using the crude bounds above, or by using bounds such as (D.10) below.

[20]To provide a link to the discussion of [24], we observe that

$$\prod_p E'_{p,\xi} = \prod_{B\subseteq\Omega, B \text{ vertical}} \zeta\big(1 + \sum_{(i,j)\in B} z_{i,j}\big)^{(-1)^{|B|}}.$$

For $p \in P_\Psi$ we must rely instead on the far weaker bound

$$E_{p,\xi} = (1 + O(1/p))E'_{p,\xi}. \tag{D.12}$$

From the estimate (D.7) and the fact that $|z_{i,j}| = O(\log^{-1/2} R)$ when $\xi_{i,j} \in I$ we have

$$\begin{aligned}\prod_p E'_{p,\xi} &= \prod_{B \subseteq \Omega, B \text{ vertical}} \left(\frac{1}{\sum_{(i,j)\in B} z_{i,j}} + O(1) \right)^{(-1)^{|B|}} \\ &= (1 + O(\log^{-1/2} R)) \prod_{B \subseteq \Omega, B \text{ vertical}} \big(\sum_{(i,j)\in B} z_{i,j} \big)^{(-1)^{|B|-1}}. \end{aligned} \tag{D.13}$$

Our aim now is to establish a corresponding estimate for $\prod_p E_{p,\xi}$. Note that we cannot afford the loss of a multiplicative constant which would result from a naïve application of (D.10).

Proposition D.4 (Euler product estimate)**.** *We have*

$$\prod_p E_{p,\xi} = \left(\prod_p \beta_p + O(e^{O(X)} \log^{-1/20} R) \right) \prod_p E'_{p,\xi}$$

for any $\xi \in I^\Omega$.

Proof. From Lemma 1.3 we have $\beta_p = 1 + O(1/p)$ for $p \in P_\Psi$ and $\beta_p = 1 + O(1/p^2)$ otherwise. For starters this implies the very crude bound

$$\prod_p \beta_p \leqslant e^{O(X)}, \tag{D.14}$$

which we will use later on. Our first main task is to dispose of the contribution of the large primes p, when (say) $p > \log^{1/10} R$. Using the estimates for β_p just mentioned, we have

$$\prod_{p \leqslant \log^{1/10} R} \beta_p \leqslant e^{O(X)}. \tag{D.15}$$

We also have

$$\begin{aligned}\prod_{p > \log^{1/10} R} \beta_p &\leqslant \exp\big(O(\sum_{p > \log^{1/10} R : p \in P_\Psi} p^{-1})\big) \\ &\leqslant \exp\big(O(X \log^{-1/20} R)\big) \\ &= 1 + O(e^{O(X)} \log^{-1/20} R),\end{aligned}$$

where the last bound follows from the elementary inequality $e^{\lambda X} \leqslant 1 + \lambda e^X$, valid for $\lambda \leqslant 1$ and $X \in \mathbb{R}_{\geqslant 0}$. Similarly, using the inequality $e^{-\lambda X} \geqslant 1 - \lambda e^X$, we obtain the corresponding lower bound, and thus

$$\prod_{p > \log^{1/10} R} \beta_p = 1 + O(e^{O(X)} \log^{-1/20} R). \tag{D.16}$$

From this and (D.14) we see that it will suffice to show that

$$\prod_p E_{p,\xi} = \Big(\prod_{p \leqslant \log^{1/10} R} \beta_p + O(e^{O(X)} \log^{-1/20} R) \Big) \prod_p E'_{p,\xi}.$$

Now from (D.10), (D.12) we have

$$\prod_{p>\log^{1/10} R} E_{p,\xi} = \exp\Big(\sum_{p>\log^{1/10} R} p^{-2} + \sum_{p\in P_\Psi : p>\log^{1/10} R} p^{-1}\Big) \prod_{p>\log^{1/10} R} E'_{p,\xi}.$$

Since $\sum_{p>\log^{1/10} R} p^{-2} = O(\log^{-1/10} R)$ and $\sum_{p\in P_\Psi : p>\log^{1/10} R} p^{-1} = O(X\log^{-1/20} R)$, we conclude that

$$\begin{aligned}\prod_{p>\log^{1/10} R} E_{p,\xi} &= \exp\big(O(1+X)\log^{-1/20} R\big) \prod_{p>\log^{1/10} R} E'_{p,\xi}\\ &= \big(1+O(e^{O(X)}\log^{-1/20} R)\big) \prod_{p>\log^{1/10} R} E'_{p,\xi},\end{aligned}$$

the last step following as in the proof of (D.16). From this and (D.14) we see that it suffices to show that

$$\prod_{p\leqslant\log^{1/10} R} E_{p,\xi} = \Big(\prod_{p\leqslant\log^{1/10} R} \beta_p + O(e^{O(X)}\log^{-1/20} R)\Big) \prod_{p\leqslant\log^{1/10} R} E'_{p,\xi}. \tag{D.17}$$

To do this, we will prove the following lemma.

Lemma D.5. *We have*

$$E_{p,\xi} = \big(\beta_p + O(\frac{\log p}{\log^{1/2} R})\big)E'_{p,\xi}.$$

for all $p \leqslant \log^{1/10} R$.

Proof that Lemma D.5 implies (D.17). Suppose first that there is $p_0 \leqslant \log^{1/10} R$ such that $\beta_{p_0} = 0$. Then, using the fact that $\beta_p = 1 + O(1/p)$, we have

$$\prod_{p\leqslant\log^{1/10} R} \big(\beta_p + O(\frac{\log p}{\log^{1/2} R})\big) = O(\frac{\log p}{\log^{1/2} R})e^{O(X)},$$

which is acceptable. If no β_p is vanishes then, since $\beta_p = 1+O(1/p)$ and β_p is a rational with denominator dividing p^d, we have a bound $\beta_p \gg 1$ with the implied constant depending only on the global parameters t, d, L. Thus, using (D.15), we have

$$\begin{aligned}\prod_{p\leqslant\log^{1/10} R} \big(\beta_p + O(\frac{\log p}{\log^{1/2} R})\big) &= \prod_{p\leqslant\log^{1/10} R} \beta_p \cdot \prod_{p\leqslant\log^{1/10} R} \big(1+O(\frac{\log p}{\log^{1/2} R})\big)\\ &= \big(\prod_{p\leqslant\log^{1/10} R} \beta_p\big)\cdot\big(1+O(\log^{-1/3} R)\big)\\ &= \prod_{p\leqslant\log^{1/10} R} \beta_p + O(e^{O(X)}\log^{-1/3} R)).\end{aligned}$$

Thus (D.17) holds in this case also. □

Proof of Lemma D.5. Observe that since $\xi \in I^\Omega$, we have

$$p^{\sum_{(i,j)\in B} z_{i,j}} = 1 + O(\log p/\log^{1/2} R)$$

for all B and all $p \leqslant \log^{1/10} R$. Dividing (D.9) by (D.11) (noting that the latter has magnitude comparable to 1) and performing Taylor expansion in $w = p^{\sum z_{i,j}}$ about $w = 1$ it is not hard to check that

$$\frac{E_{p,\xi}}{E'_{p,\xi}} = \frac{\widetilde{E}_p}{\widetilde{E}'_p} + O(\frac{\log p}{\log^{1/2} R}),$$

where $\widetilde{E}_p, \widetilde{E}'_p$ are defined setting all the $z_{i,j}$ equal to zero in (D.9) and (D.11) respectively. Thus

$$\widetilde{E}_p := \sum_{B \subseteq \Omega} (-1)^{|B|} \alpha(p, B) \tag{D.18}$$

and

$$\widetilde{E}'_p := \sum_{B \subseteq \Omega, B \text{ vertical}} \big(1 - \frac{1}{p}\big)^{(-1)^{|B|-1}}. \tag{D.19}$$

To prove the lemma, then, it suffices to prove the identity

$$\beta_p = \frac{\widetilde{E}_p}{\widetilde{E}'_p}. \tag{D.20}$$

Recalling (D.18) and (D.19), it will suffice to show that

$$\sum_{B \subseteq \Omega} (-1)^{|B|} \alpha(p, B) = \beta_p \prod_{B \subseteq \Omega, B \text{ vertical}} (1 - \frac{1}{p})^{(-1)^{|B|-1}}. \tag{D.21}$$

Using the binomial theorem, the right-hand side of (D.21) simplifies to $\beta_p(1-\frac{1}{p})^t$, which by (1.6) is equal to

$$\mathbb{E}_{n \in \mathbb{Z}_p^d} 1_{p \nmid \psi_1(n)} \dots 1_{p \nmid \psi_t(n)}.$$

By the inclusion-exclusion principle this can be written as

$$\sum_{r_1, \dots, r_t \in \{0,1\}} (-1)^{r_1 + \dots + r_t} \mathbb{E}_{n \in \mathbb{Z}_p^d} \prod_{r_i = 0} 1_{p | \psi_i(n)},$$

which in turn is just

$$\sum_{r_1, \dots, r_t \in \{0,1\}} (-1)^{r_1 + \dots + r_t} \alpha_{p^{r_1}, \dots, p^{r_t}}.$$

We are to show that this is equal to the left-hand side of (D.21), namely

$$\sum_{B \subseteq \Omega} (-1)^{|B|} \alpha(p, B).$$

To do this, we compare coefficients of $\alpha_{p^{r_1}, \dots, p^{r_t}}$ on both sides. To evaluate the coefficient on the left-hand side, let I be the set of indices for which $r_i \neq 0$. Then this coefficient is easily seen to be

$$\prod_{i \in I} \sum_{B_i \subseteq [a_i]} (-1)^{|B_i|}$$

which, by the binomial theorem, is simply $(-1)^{|I|}$. This gives (D.20), and the claim follows. □

We return to the proof of (D.4). Recall that we had reduced this to the task of finding an approproate asymptotic for (D.8). Substituting the result of Proposition D.4 into

(D.8) and applying (D.14), it is easy to reduce this in turn to showing the following two facts. Firstly, that

$$\log^t R \int_I \dots \int_I \Big(\prod_p E'_{p,\xi}\Big) \prod_{(i,j)\in\Omega} \varphi_i(\xi_{i,j})\, d\xi_{i,j} = \prod_{i\in[t]} c_{\chi_i,a_i} + O(\log^{-1/20} R); \tag{D.22}$$

and secondly that

$$\log^t R \int_I \dots \int_I \prod_p |E'_{p,\xi}| \prod_{(i,j)\in\Omega} |\varphi_i(\xi_{i,j})|\, d\xi_{i,j} = O(1). \tag{D.23}$$

Let us begin with the second task, that of proving (D.23). We simply substitute $z_{i,j} = (1+i\xi_{i,j})/\log R$ into (D.13). The contribution from the terms $\log R$ is precisely $\log^{-t} R$, by a simple application of the binomial theorem $\sum_{B\subseteq C: B\neq\emptyset}(-1)^{|B|} = 0^{|C|} - 1$. For the terms involving the $\xi_{i,j}$ we have the crude estimate

$$\prod_{(i,j)\in\Omega} (1+|\xi_{i,j}|)^{O(1)},$$

and so

$$\prod_p |E'_{p,\xi}| \ll \frac{1}{\log^t R} \prod_{(i,j)\in\Omega} (1+|\xi_{i,j}|)^{O(1)}.$$

However since χ is smooth its modified Fourier transform satisfies $|\varphi_i(\xi_{i,j})| \ll_A (1+|\xi_{i,j}|)^{-A}$ for any $A > 0$, as we have already remarked. The claim then follows by taking A large enough.

Now we prove (D.22). Using the rapid decay of the functions φ once more together with (D.13) we see that it suffices to show that

$$\log^t R \int_I \dots \int_I \prod_{B\subseteq\Omega, B \text{ vertical}} \Big(\sum_{(i,j)\in B} z_{i,j}\Big)^{(-1)^{|B|-1}} \prod_{(i,j)\in\Omega} \varphi_i(\xi_{i,j})\, d\xi_{i,j} = \prod_{i\in[t]} c_{\chi_i,a_i} + O(\log^{-1/20} R).$$

The first move is to reinstate the integrals over all of $\mathbb{R}$, rather than just over I. Doing this introduces an error which is $\ll_A \log^{-A} R$ for any $A > 0$, on account of the rapid decrease of φ. Once this is done the multiple integral is easily seen to factor, there being one integral for each index i. After scaling out the factors of $\log R$, the claim follows from the definition (D.1) of the sieve weights $c_{\chi,a}$. The result follows, and we have concluded the proof of Theorem D.3. □

Construction of the enveloping sieve. Now we are ready to prove Proposition 6.4, the statement of which was as follows.

Proposition 6.4 (Domination by a pseudorandom measure)**.** *Let $D > 1$ be arbitrary. Then there is a constant $C_0 := C_0(D)$ such that the following is true. Let $C \geqslant C_0$, and suppose that $N' \in [CN, 2CN]$. Let $b_1, \dots, b_t \in \{0, 1, \dots, W-1\}$ be coprime to*

$W := \prod_{p \leqslant w} p$. Then there exists a D-pseudorandom measure $\nu : \mathbb{Z}_{N'} \to \mathbb{R}^+$ which obeys the pointwise bounds

$$1 + \Lambda'_{b_1,W}(n) + \ldots + \Lambda'_{b_t,W}(n) \ll_{D,C} \nu(n)$$

for all $n \in [N^{3/5}, N]$, where we identify n with an element of $\mathbb{Z}_{N'}$ in the obvious manner.

The definition of D-pseudorandom was given, and discussed, in §6. See in particular Definitions 6.2 and 6.3 and the paragraphs following the latter. Let $\gamma = \gamma(C, D) > 0$ be a parameter to be chosen later and set $R := N^\gamma$. Fix an arbitrary smooth even function $\chi : \mathbb{R} \to \mathbb{R}$ which is supported on $[-1, 1]$ and satisfies $\chi(0) = 1$ and $\int_0^1 |\chi'(x)|^2 \, dx = 1$. For such a function we have $c_{\chi,2} = 1$, thanks to Lemma D.2.

We define the preliminary weight $\tilde{\nu} : [N] \to \mathbb{R}^+$ by setting

$$\tilde{\nu}(n) := \mathbb{E}_{i \in [t]} \frac{\phi(W)}{W} \Lambda_{\chi,R,2}(Wn + b_i)$$

and then transfer this to $\mathbb{Z}_{N'}$ by setting $\nu(n) := \frac{1}{2} + \frac{1}{2}\tilde{\nu}(n)$ when $n \in [N]$ and $\nu(n) := 1$ otherwise.

By construction, $\tilde{\nu}$ is certainly non-negative. To verify the pointwise bounds, it suffices to show that

$$\Lambda'_{b_i,W}(n) \ll_{C,D} \frac{\phi(W)}{W} \Lambda_{\chi,R,2}(Wn + b_i)$$

for all $i \in [t]$ and $n \in [N^{3/5}, N]$. The left-hand side is only non-zero when $Wn + b_i$ is a prime which is greater than $N^{3/5}$. Supposing that $\gamma < 3/5$, we see that in this case the left-hand side is equal to $\frac{\phi(W)}{W} \log N$, while the right-hand side is $\frac{\phi(W)}{W} \log R$. Since $R = N^\gamma$ and γ depends only on C, D, the claim follows.

It remains to show that ν is a D-pseudorandom measure. Our argument here shall follow that in [24] rather closely, but will use Theorem D.3 as a substitute for [24, Propositions 9.5,9.6]. For that reason we shall skip some of the details which are more or less exact repetition of those in [24].

Let us first verify the (D, D, D)-linear forms condition. By decomposing ν up into its various components as in [24], it certainly suffices to establish the somewhat general bound

$$\sum_{n \in K \cap \mathbb{Z}^d} \prod_{j \in [m]} \tilde{\nu}(\psi_j(n)) = \mathrm{vol}_d(K) + o(N^d)$$

where $\Psi = (\psi_1, \ldots, \psi_m)$ is a system of affine-linear forms, no two of which are affinely related, $m, d, \|\Psi\|_N$ are all $O_D(1)$, and $K \subseteq [-N, N]^d$ is a convex body with $\Psi(K) \subseteq [-N, N]^m$. Splitting $\tilde{\nu}$ up further, we thus reduce to showing that

$$\left(\frac{\phi(W)}{W}\right)^m \sum_{n \in K \cap \mathbb{Z}^d} \prod_{j \in [m]} \Lambda_{\chi,R,2}(\psi_j(Wn + b_{i_j})) = \mathrm{vol}_d(K) + o(N^d) \tag{D.24}$$

for all $i_1, \ldots, i_m \in [t]$.

Now we apply Theorem D.3. As we are assuming that no two of the forms $\psi_j(n)$ are affinely related, the same is true for the forms $\psi_j(Wn + b_{i_j})$. In particular we see that the exceptional primes, if they exist, are bounded in size by $O(w) = O(\log\log N)$. In particular we have $X = O(\log\log^{1/2} N)$ and so $e^{O(X)}\log^{-1/20} R = o(1)$. We can thus write the left-hand side of (D.24) as

$$\left(\frac{\phi(W)}{W}\right)^m c_{\chi,2}^m \operatorname{vol}_d(K) \prod_p \beta_p + o(N^d)$$

where we suppress the dependence of constants on t, m, d, L, D. Because all the b_{i_j} are coprime to W, we see that $\beta_p = (\frac{p}{p-1})^t$ for all $p \leqslant w$, and in particular $\prod_{p\leqslant w}\beta_p = \left(\frac{W}{\phi(W)}\right)^t$. Also, for $p > w$ we see from Lemma 1.3 that $\beta_p = 1 + O(1/p^2)$, and so $\prod_{p>w}\beta_p = 1 + o(1)$. Since $c_{\chi,2} = 1$, the claim follows.

Now we verify the D-correlation condition for ν. As before we can pass from ν to $\tilde\nu$, and reduce to showing that

$$\sum_{n\in I}\prod_{j\in[m]}\tilde\nu(n+h_j) \ll N \sum_{1\leqslant j<j'\leqslant m}\tau(h_j - h_{j'})$$

for all $m = O_D(1)$, all $h_1, \dots, h_m \in [N]$, and all intervals $I \subseteq [N]$, and where $\tau : [-N, N] \to \mathbb{R}^+$ obeys the moment bounds $\mathbb{E}_{n\in[-N,N]}\tau(n)^q \ll_q 1$ for all $q > 0$. We may assume that no two of the h_i are equal as in this case one can use crude divisor estimates, setting $\tau(0)$ to be moderately large (see [24] for details). Again, we split up $\tilde\nu$ and reduce to showing that

$$\left(\frac{\phi(W)}{W}\right)^m \left(\sum_{n\in I}\prod_{j\in[m]}\Lambda_{\chi,R,2}(W(n+h_j)+b_{i_j})\right) \ll N \sum_{1\leqslant j<j'\leqslant m}\tau(h_j - h_{j'})$$

whenever $i_1, \dots, i_m \in [t]$. We can apply Theorem D.3 with the system of forms $\Psi = (W(n + h_j) + b_{i_j})_{j=1}^m$ and write the left-hand side as

$$\left(\frac{\phi(W)}{W}\right)^m \left(c_{\chi,2}^m |I| \prod_p \beta_p + O(Ne^{O(X)}\log^{-1/20} R)\right).$$

As before we can discard the sieve factor $c_{\chi,2} = 1$, and we have $\prod_{p\leqslant w}\beta_p = \left(\frac{W}{\phi(W)}\right)^m$.

It thus suffices to show that

$$\prod_{p>w}\beta_p + e^{O(X)}\log^{-1/20} R \ll \sum_{1\leqslant j<j'\leqslant m}\tau(h_j - h_{j'}).$$

From Lemma 1.3 we see that for $p > w$ we have $\beta_p = 1 + O(1/p)$, with the improvement $\beta_p = 1 + O(1/p^2)$ as long as $p \notin P_\Psi$, that is as long as p does not divide $W(h_j - h_{j'}) + b_{i_j} - b_{i_{j'}}$ for any $1 \leqslant j < j' \leqslant m$. Thus

$$\prod_{p>w}\beta_p \ll \prod_{\substack{p>w\\ p\in P_\Psi}}\left(1 + O(\frac{1}{p})\right) \ll \exp\left(O\Big(\sum_{\substack{p>w\\ p\in P_\Psi}}\frac{1}{p}\Big)\right). \tag{D.25}$$

On the other hand, since $w = O(\log\log N)$ is so small we have

$$e^{O(X)} \log^{-1/20} R \leqslant \exp(O(\sum_{p\in P_\Psi} \frac{1}{p^{1/2}})) \log^{-1/20} R$$
$$\ll \exp\left(O(\sum_{\substack{p>w\\ p\in P_\Psi}} \frac{1}{p^{1/2}})\right).$$

It follows from this analysis that if we set

$$\tau(n) := \sum_{1\leqslant j<j'\leqslant m} \exp\left(O(\sum_{\substack{p>w\\ p|Wn+b_{i_j}-b_{i_{j'}}}} \frac{1}{p^{1/2}})\right)$$

then we obtain the desired correlation estimate. To show the moment bounds on τ it suffices to show that

$$\mathbb{E}_{n\in[N]} \exp(q \sum_{\substack{p>w\\ p|Wn+h}} \frac{1}{p^{1/2}}) \ll_q 1$$

for all $h = O(W)$. By repeating the proof of [24, Lemma 9.9] we can deduce this bound from

$$\sum_{n\in[N]} \prod_{\substack{p>w\\ p|Wn+h}} (1 + p^{-1/4}) \ll_q N.$$

Using the bound

$$\prod_{\substack{p>w\\ p|Wn+h}} (1 + p^{-1/4}) \leqslant \sum_{\substack{(d,W)=1\\ d|Wn+h}} d^{-1/4}$$

we reduce to showing that

$$\sum_{\substack{(d,W)=1\\ d=O(WN)}} d^{-1/4} \sum_{\substack{n\in[N]\\ d|Wn+h}} 1 \ll_q N.$$

But we have

$$\sum_{\substack{n\in[N]\\ d|Wn+h}} 1 = O(1 + N/d)$$

by the Chinese remainder theorem, and the claim then follows easily. This concludes the proof of Proposition 6.4. □

The correlation estimate for $\Lambda^\sharp$. The final task of this appendix is to establish the correlation estimate (12.5), which was the crucial fact that $\Lambda^\sharp_{b,W} - 1$ has small Gowers norm. We allow all constants to depend on s. Expanding out the $U^{s+1}[N]$ norm, it suffices to show the slightly more general bound

$$\sum_{(n,h)\in K} \prod_{\omega\in\{0,1\}^{s+1}} (\frac{\phi(W)}{W} \Lambda^\sharp(W(n + \omega\cdot h) + b) - 1) = o(N^{s+2})$$

whenever K is a convex body in $[-N,N]^{s+2}$. Expanding out the product, it suffices to show that

$$\sum_{(n,h)\in K}\prod_{\omega\in B}\frac{\phi(W)}{W}\Lambda^\sharp(W(n+\omega\cdot h)+b) = \mathrm{vol}_{s+2}(K)+o(N^{s+2})$$

for all $B\subseteq\{0,1\}^{s+1}$. Now observe that $\Lambda^\sharp = -\Lambda_{\chi^\sharp,R,1}$, and so we may invoke Theorem D.3 with the system of forms $\Psi = (W(n+\omega\cdot h)+b)_{\omega\in B}$ to write the left-hand side as

$$\big(-\frac{\phi(W)}{W}\big)^{|B|}c_{\chi^\sharp,1}^{|B|}\,\mathrm{vol}_{s+2}(K)\prod_p\beta_p + O(N^{s+2}e^{O(X)}/\log^{1/20}R).$$

As in the preceding section, we compute that $\prod_{p\leqslant w}\beta_p = \left(\frac{W}{\phi(W)}\right)^{|B|}$, while $\beta_p = 1+O(1/p^2)$ for $p>w$. Furthermore all exceptional primes p have $p\leqslant w$, and thus since w is so small

$$e^{O(X)}/\log^{1/20}R = O(\log^{-1/20}R)\exp\big(O(\sum_{p\leqslant w}\frac{1}{p^{1/2}})\big) = o(1).$$

Finally, from Lemma D.2, we have $c_{\chi^\sharp,1} = -1$. The claim follows. □

Appendix E. Nilmanifold constraints; Host-Kra cube groups

Our aim in this appendix is prove Proposition 11.5, which asserts a constraint concerning parallelepiped in nilmanifolds. It turns out to be convenient to generalise the notion of a parallelepiped to a more general object, namely a *Host-Kra cube*. Thus much of this appendix will be devoted to the algebraic theory of these cubes. We will first introduce such parallelepipeds in the Lie group G, establish the constraint there, and then descend to the quotient space G/Γ and show that the constraint persists down to the quotient. In preparing the material that follows we benefitted much from conversations with Sasha Leibman, and also from remarks made by one of the anonymous referees.

Host-Kra cube groups in G.

Let G be a connected Lie group with identity id_G, with the associated lower central series $G_\bullet$ given by

$$G = G_0 = G_1 \supseteq G_2 \supseteq \dots,$$

where $G_0 = G_1 = G$ and $G_{i+1} = [G,G_i]$. We recall the standard facts that $[G_i,G_j]\subseteq G_{i+j}$, and that each G_i is a closed connected normal Lie subgroup of G; see for instance [7, Ch. 3, §9, Corollary to Prop. 4]. In particular the quotient groups $G_i\backslash G$ are also Lie groups.

To define the Host-Kra cube group $\mathrm{HK}^{s+1}(G_\bullet)$ we first need some combinatorial notation.

Definition E.1 (Simple combinatorics of $\{0,1\}^{s+1}$)**.** We refer to $\{0,1\}^{s+1}$ as *the* cube. Its elements ω may be partially ordered by decreeing that $\omega\leqslant\omega'$ if $\omega_j\leqslant\omega'_j$ for $j=1,\dots,s+1$. A *hyperplane* is any set of the form $H_{j,a} := \{\omega : \omega_j = a\}$. If $0\leqslant d\leqslant s+1$ then we say that a *face of codimension d* is any non-empty intersection F

of d distinct hyperplanes, and we write $d = \mathrm{codim}(F)$. Thus any vertex in $\{0,1\}^{s+1}$ is a face of codimension $s+1$, whilst the whole cube $\{0,1\}^{s+1}$ is a face of codimension 0. We say that two faces are *parallel* if they have the same fixed coordinates, and hence the same codimension. Every face F has a minimal element $\min(F)$ and a maximal element $\max(F)$. We say that a face[21] is *lower* if $\min(F) = 0^{s+1}$. Note that every face is parallel to exactly one lower face, and that lower faces F are in one-to-one correspondence with their maximal elements $\max(F)$, which can be arbitrary. Finally, we say that two parallel faces are *adjacent* if their union is a face of one lower codimension.

Definition E.2 (Face groups)**.** Let $F \subseteq \{0,1\}^{s+1}$ be an face of codimension d. For any element $g \in G$, we write g^F for the element of $G^{\{0,1\}^{s+1}}$ such that $(g^F)_\omega = g$ when $\omega \in F$, and $(g^F)_\omega = \mathrm{id}_G$ otherwise. The *face group* Γ_F is the group generated by all elements g^F with $g \in G_{\mathrm{codim}(F)}$, thus $\Gamma_F \cong G_{\mathrm{codim}(F)}$.

Definition E.3 (Host-Kra cube group)**.** The Host-Kra cube group $\mathrm{HK}^{s+1}(G_\bullet)$ is the subgroup of $G^{\{0,1\}^{s+1}}$ generated by all the face groups Γ_F, as F ranges over faces of $\{0,1\}^{s+1}$.

The Host-Kra cube group could be defined with a more general *filtration* in place of the lower central series $G_\bullet$, that is to say a sequence of subgroups in which the condition that $G_{i+1} = [G, G_i]$ is relaxed to an inclusion $[G_i, G_j] \subseteq G_{i+j}$. We will not need this here.

The significance of the group $\mathrm{HK}^{s+1}(G_\bullet)$ for us is that it contains the parallelepipeds:

Lemma E.4 (Parallelepipeds are Host-Kra cubes)**.** *Given any $g, x \in G$ and $n, h_1, \dots, h_{s+1}$ in $\mathbb{Z}$, the parallelepiped $\mathbf{g} := (g^{n+\omega\cdot h}x)_{\omega\in\{0,1\}^{s+1}}$ lies in $\mathrm{HK}^{s+1}(G_\bullet)$.*

Proof. We may write, in $G^{\{0,1\}^{s+1}}$,

$$\mathbf{g} = (g^{h_{s+1}})^{F_{s+1}}(g^{h_s})^{F_s}\dots(g^{h_1})^{F_1}(g^n x)^{F_0},$$

where $F_0 := \{0,1\}^{s+1}$ and F_i is the hyperplane $F_i := \{\omega : \omega_i = 1\}$ for $i = 1, \dots, s+1$. Thus $\mathbf{g}$ is the product of $s+2$ of the generators of $\mathrm{HK}^{s+1}(G_\bullet)$. □

The face groups G_F are related to each other in a pleasant way:

Lemma E.5 (Face relations)**.** *Let F, F' be faces in $\{0,1\}^{s+1}$.*

(i) *If F, F' are disjoint, then the elements in Γ_F and $\Gamma_{F'}$ commute with one another.*
(ii) *If F and F' intersect then $[\Gamma_F, \Gamma_{F'}] \subseteq \Gamma_{F\cap F'}$.*
(iii) *If F and F' are adjacent and parallel, then $\Gamma_F \subseteq \Gamma_{F'}\Gamma_{F\cup F'}$ and $\Gamma_{F'} \subseteq \Gamma_F\Gamma_{F\cup F'}$.*

Proof. (i) is immediate. To prove (ii), note that any element of $[\Gamma_F, \Gamma'_F]$ has the form $x^{F\cap F'}$ for some $x \in [G_d, G_{d'}]$, where $d := \mathrm{codim}(F)$ and $d' := \mathrm{codim}(F')$. The result

[21] With respect to the partial ordering $\leqslant$, a lower face is exactly the same concept as a principal filter.

follows on noting that $\mathrm{codim}(F \cap F') \leqslant \mathrm{codim}(F) + \mathrm{codim}(F')$, and recalling from group theory that $[G_d, G_{d'}] \subseteq G_{d+d'}$. (iii) is immediate; in this setting we have $x^F x^{F'} = x^{F \cup F'}$. □

From Lemma E.5 (iii) and an easy induction on the codimension we see that every face group Γ_F lies in the group generated by the *lower* face groups. In particular this implies that the entire group $\mathrm{HK}^{s+1}(G_\bullet)$ is generated by the lower face groups. The same result holds for the *upper faces*, but we will not have any further use of this here.

Now we seek a more explicit description of $\mathrm{HK}^{s+1}(G_\bullet)$ by the lower face groups. To achieve this, we need

Definition E.6 (Decreasing ordering of faces)**.** Let $F_1 > \cdots > F_{2^{s+1}}$ be any ordering of the 2^{s+1} lower faces of $\{0,1\}^{s+1}$. We say that this ordering is *decreasing* if whenever $F_i \supseteq F_j$ we have $i \geqslant j$. Thus $F_1 = \{0,1\}^{s+1}$ and $F_{2^{s+1}} = 0^{s+1}$.

Clearly, decreasing orders of faces exist; let us fix such an ordering. Now, observe from Lemma E.5 (i),(ii) that if $i < j$, then we either have $\Gamma_{F_j} \cdot \Gamma_{F_i} \subseteq \Gamma_{F_i} \cdot \Gamma_{F_j}$ or $\Gamma_{F_j} \cdot \Gamma_{F_i} \subseteq \Gamma_{F_i} \cdot \Gamma_{F_j} \cdot \Gamma_{F_k}$ for some $k > j$. From these inclusions we see that any product of elements from the lower face groups Γ_{F_i} can eventually be contained in $\Gamma_{F_1} \cdot \Gamma_{F_2} \cdot \ldots \cdot \Gamma_{F_{2^{s+1}}}$, as one can use the above inclusions to move all occurrences of Γ_{F_1} to the far left, use the closure property $\Gamma_{F_1} \cdot \Gamma_{F_1} = \Gamma_{F_1}$ to concatenate, then move all occurences of Γ_{F_2} to be adjacent to Γ_{F_1}, and so forth. Since the lower face groups generate $\mathrm{HK}^{s+1}(G_\bullet)$, we have thus obtained the factorisation

$$\mathrm{HK}^{s+1}(G_\bullet) = \Gamma_{F_1} \cdot \Gamma_{F_2} \cdot \ldots \cdot \Gamma_{F_{2^{s+1}}}.$$

Thus there exist functions $\tau_i : \mathrm{HK}^{s+1}(G_\bullet) \to \Gamma_{F_i}$ such that

$$\mathbf{g} = \tau_1(\mathbf{g}) \ldots \tau_{2^{s+1}}(\mathbf{g}) \tag{E.1}$$

for all $\mathbf{g} \in \mathrm{HK}^{s+1}(G_\bullet)$.

Remark. Since $\Gamma_{F_i} \cong G_{\mathrm{codim}(F_i)}$ is a closed connected Lie subgroup of $G^{\{0,1\}^{s+1}}$, we can conclude from the above factorisation that $\mathrm{HK}^{s+1}(G_\bullet)$ is also a closed connected group Lie subgroup of $G^{\{0,1\}^{s+1}}$. Furthermore, since the hyperplane face groups consist entirely of parallelepipeds, and the lower dimensional face groups can be expressed as commutators of the hyperplane face groups, we see that $\mathrm{HK}^{s+1}(G_\bullet)$ is in fact the subgroup generated by the parallelepipeds. Thus this is an extremely natural group for studying parallelepipeds.

In the factorisation (E.1), the τ_i are unique: an inspection of the $\max(F_1)$ coefficients of both sides shows that $\tau_1(\mathbf{g})$ is determined uniquely by $\mathbf{g}$, and then after factoring $\tau_1(\mathbf{g})$ out, an inspection of the $\max(F_2)$ coefficients of both sides shows that $\tau_2(\mathbf{g})$ is determined uniquely by $\mathbf{g}$, and so forth. Indeed, this algorithm shows that if $\mathbf{g} = (g_\omega)_{\omega \in \{0,1\}^{s+1}}$, then $\tau_i(\mathbf{g}) \in \Gamma_{F_i}$ is a continuous function of the coordinates $g_{\max(F_1)}, \ldots, g_{\max(F_i)}$ only; indeed, equating Γ_{F_i} with $G_{\mathrm{codim}(F_i)}$, the group element $\tau_i(\mathbf{g})$ is an explicit word in these coordinates. Conversely, $g_{\max(F_i)}$ is a word in $\tau_1(\mathbf{g}), \ldots, \tau_i(\mathbf{g})$ only.

Recall that we are aiming to prove Proposition 11.5, which establishes a constraint amongst the 2^{s+1} vertices of a parallelepiped in G/Γ, an s-step nilmanifold. Henceforth we assume that we are in this setting (the discussion up to now has been valid quite generally). The preceding observations allow one to prove a related fact, namely that if G is s-step nilpotent and if $(g_\omega)_{\omega\in\{0,1\}^{s+1}} \in \mathrm{HK}^{s+1}(G_\bullet)$ then $g_{0^{s+1}}$ is a word in the g_ω, $\omega \in \{0,1\}^{s+1}_*$. Indeed the nilpotence of G implies that the final face group $\Gamma_{F_{2^{s+1}}}$ is trivial, and hence $\tau_{2^{s+1}}(\mathbf{g}) = \mathrm{id}$ for all $\mathbf{g}$. Thus $g_{0^{s+1}} = g_{\max(F_{2^{s+1}})}$ is a actually a word in $\tau_1(\mathbf{g}), \dots, \tau_{2^{s+1}-1}(\mathbf{g})$, and hence in the g_ω, $\omega \in \{0,1\}^{s+1}_*$.

To prove Proposition 11.5, we must show how this constraint "descends" to G/Γ. A step in this direction is the following lemma, which follows immediately from the fact that $g_{0^{s+1}}$ is a word in the g_ω, $\omega \in \{0,1\}^{s+1}_*$.

Lemma E.7. *Suppose that $g = (g_\omega)_{\omega\in\{0,1\}^{s+1}} \in \mathrm{HK}^{s+1}(G_\bullet)$ and that $g_\omega \in \Gamma$ for all $\omega \in \{0,1\}^{s+1}_*$. Then the remaining point $g_{0^{s+1}}$ lies in Γ as well.*

We have defined the Host-Kra cube *group*; now we define the Host-Kra nilmanifold.

Definition E.8 (Host-Kra cube nilmanifold)**.** We define the Host-Kra nilmanifold $\mathrm{HK}^{s+1}(G_\bullet/\Gamma)$ to be the s-step nilmanifold $\mathrm{HK}^{s+1}(G_\bullet)/(\Gamma^{\{0,1\}^{s+1}} \cap \mathrm{HK}^{s+1}(G_\bullet))$.

A priori, this definition does not make sense. The Lie group $\mathrm{HK}^{s+1}(G_\bullet)$ is connected, simply-connected and s-step nilpotent Lie group (the nilpotence follows from the fact that it is a subgroup of $G^{\{0,1\}^{s+1}}$ and the simple-connectedness from the factorisation (E.1) together with the simple-connectedness of the face groups $\Gamma_{F_i} \cong G_{\mathrm{codim}(F_i)}$). We have not, however, shown that $\Gamma^{\{0,1\}^{s+1}} \cap \mathrm{HK}^{s+1}(G_\bullet)$ is cocompact inside it. This is the business of Lemma E.10 below. To prove it, we will need a basic topological property of nilmanifolds, first established in the foundational paper of Mal'cev [37].

Lemma E.9. [37] *Let G be a connected, simply-connected nilpotent Lie group, and let Γ be a discrete cocompact subgroup. Then for any $j \geqslant 1$ the group $\Gamma \cap G_j$ is discrete and cocompact in G_j.*

Remark. To obtain results such as the Main Theorem in the case $s = 2$, we need only consider nilmanifolds which are products of Heisenberg examples. This was observed in Proposition 8.4. In this case, Lemma E.9 can easily be verified by hand using calculations along the lines of those in [27, Appendix B].

Lemma E.10. $\Gamma^{\{0,1\}^{s+1}} \cap \mathrm{HK}^{s+1}(G_\bullet)$ *is a discrete and cocompact subgroup of* $\mathrm{HK}^{s+1}(G_\bullet)$.

Proof. The discreteness is obvious, since $\Gamma^{\{0,1\}^{s+1}}$ is discrete in $G^{\{0,1\}^{s+1}}$. Now by Lemma E.9 there is a a compact set $K_j \subseteq G_j$ such that $G_j = K_j \cap (\Gamma \cap G_j)$. For each i, consider the subgroup $H_i \leqslant \mathrm{HK}^{s+1}(G_\bullet)$ consisting of those $\mathbf{g}$ such that $\tau_1(\mathbf{g}) = \cdots = \tau_i(\mathbf{g}) = \mathrm{id}$. By our earlier observations this is the same as the subgroup $\{\mathbf{g} : g_{\max(F_1)} = \cdots = \mathbf{g}_{\max(F_i)} = \mathrm{id}\}$, and hence in particular H_i is normal in $\mathrm{HK}^{s+1}(G_\bullet)$.

Suppose that $1 \leqslant i \leqslant 2^{s+1}$ and that $\mathbf{g} \in H_{i-1}$. Then $\mathbf{g} = \tau_i(\mathbf{g})\mathbf{h}$, where $\mathbf{h} \in H_i$. Since $\tau_i(\mathbf{g})$ lies in the fact group Γ_{F_i}, we may write it as $(k_i\gamma_i)^{F_i}$ where $k_i \in K_{\mathrm{codim}(F_i)}$ and

$\gamma_i \in \Gamma \cap G_{\mathrm{codim}(F_i)}$. Since H_i is normal, we may hence write

$$\mathbf{g} = (k_i)^{F_i} \cdot \mathbf{h}' \cdot (\gamma_i)^{F_i},$$

where $\mathbf{h}'$ is another element of H_i.

Continuing inductively until $i = 2^{s+1}$, we eventually express an arbitrary element of $\mathrm{HK}^{s+1}(G_\bullet)$ as a product of $(k_1)^{F_1} \dots (k_{2^{s+1}})^{F_{2^{s+1}}}$ times an element of $\Gamma^{\{0,1\}^{s+1}} \cap \mathrm{HK}^{s+1}(G_\bullet)$. Since the set

$$\{(k_1)^{F_1} \dots (k_{2^{s+1}})^{F_{2^{s+1}}} : k_1 \in K_1, \dots, k_{2^{s+1}} \in K_{2^{s+1}}\}$$

is compact, this proves the lemma. □

Proof of Proposition 11.5. The projection $G^{\{0,1\}^{s+1}} \to (G/\Gamma)^{\{0,1\}^{s+1}}$ induces a 1-1 continuous map from the compact set $\mathrm{HK}^{s+1}(G_\bullet/\Gamma)$ to $(G/\Gamma)^{\{0,1\}^{s+1}}$. Henceforth, we consider the former set as a a compact subset of the latter. Let p be the restriction to $\mathrm{HK}^{s+1}(G_\bullet/\Gamma)$ of the obvious projection from $(G/\Gamma)^{\{0,1\}^{s+1}}$ to $(G/\Gamma)^{\{0,1\}^{s+1}_*}$, and let Σ be the range of this map. It follows from Lemma E.7 that this map is 1-1, and hence there is a unique map $P : \Sigma \to G/\Gamma$ such that $(P(x), x) \in \mathrm{HK}^{s+1}(G_\bullet/\Gamma)$ for every $x = (x_\omega)_{\omega \in \{0,1\}^{s+1}_*} \in \Sigma$. The map P is automatically continuous since its graph $\mathrm{HK}^{s+1}(G_\bullet/\Gamma)$ is compact and all spaces involved are Hausdorff.

Proposition 11.5 follows immediately from Lemma E.4. □

References

[1] L. Auslander, L. Green and F. Hahn, *Flows on homogeneous spaces,* Annals of Math. Studies **53**, Princeton 1963.

[2] R. C. Baker and G. Harman, *Exponential sums formed with the Möbius function,* J. London Math. Soc. (2) **43** (1991), no. 2, 193–198.

[3] A. Balog, *The Hardy-Littlewood k-tuple conjecture on average*, Analytic Number Theory (eds. B. Brendt, H.G. Diamond, H. Halberstam and A. Hildebrand), Birkhäuser, 1990, 47–75.

[4] ———, *Linear equations in primes,* Mathematika **39** (1992) 367–378.

[5] V. Bergelson, *Weakly mixing PET*, Ergod. Th. and Dynam. Sys. **7** (1987), 337–349.

[6] V. Bergelson, B. Host and B. Kra, *Multiple recurrence and nilsequences*, with an appendix by I.Z. Ruzsa, Invent. Math. **160** (2005), no. 2, 261–303.

[7] N. Bourbaki, *Lie groups and Lie algebras,* Chapters 1–3. Translated from the French. Reprint of the 1989 English translation. Elements of Mathematics (Berlin). Springer-Verlag, Berlin, 1998. xviii+450 pp.

[8] S. Chowla, *The Riemann hypothesis and Hilbert's tenth problem,* Gordon and Breach, New York-London-Paris 1965.

[9] L. J. Corwin and F. P. Greenleaf, *Representations of nilpotent Lie groups and their applications. Part I. Basic theory and examples,* Cambridge Studies in Advanced Mathematics, **18**, Cambridge University Press, Cambridge, 1990. viii+269 pp.

[10] H. Davenport, *On some infinite series involving arithmetical functions. II*, Quart. J. Math. Oxf. **8** (1937), 313–320.

[11] ———, *Multiplicative number theory,* Third edition. Revised and with a preface by Hugh L. Montgomery. Graduate Texts in Mathematics, **74**, Springer-Verlag, New York, 2000. xiv+177 pp.

[12] L. E. Dickson, *A new extension of Dirichlet's theorem on prime numbers*, Messenger of Math. **33** (1904), 155–161.
[13] H. Furstenberg, *Nonconventional ergodic averages,* The legacy of John von Neumann (Hempstead, NY, 1988), 43–56, Proc. Sympos. Pure Math., **50**, Amer. Math. Soc., Providence, RI, 1990.
[14] ———, *From the Erdős-Turán conjecture to ergodic theory—the contribution of combinatorial number theory to dynamics,* Paul Erdős and his mathematics, I (Budapest, 1999), 261–277, Bolyai Soc. Math. Stud., 11, János Bolyai Math. Soc., Budapest, 2002.
[15] A. Ghosh, *The distribution of αp^2 modulo* 1*,* Proc. London Math. Soc. (3) **42** (1981), no. 2, 252–269.
[16] D. Goldston and C. Y. Yıldırım, *Higher correlations of divisor sums related to primes, I: Triple correlations,* Integers 3 (2003), 66pp.
[17] ———, *Higher correlations of divisor sums related to primes, III: Small gaps between primes,* Proc. London Math. Soc. **95** (2007), 653686.
[18] ———, *Small gaps between primes, I,* preprint available at `http://front.math.ucdavis.edu/`.
[19] D. A. Goldston, J. Pintz, and C.Y. Yıldırım, *Small gaps between primes II*, preprint.
[20] W. T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, GAFA **8** (1998), 529–551.
[21] ———, *A new proof of Szemerédi's theorem*, GAFA **11** (2001), 465–588.
[22] B. J. Green, *Roth's theorem in the primes,* Annals of Math. **161** (2005), no. 3, 1609–1636.
[23] ———, *Generalising the Hardy-Littlewood method for primes,* International Congress of Mathematicians. Vol. II, 373-399, Eur. Math. Soc., Zurich, 2006.
[24] B. J. Green and T. C. Tao, *The primes contain arbitrarily long arithmetic progressions*, Annals of Math. **167** (2008), 481–547.
[25] ———*Restriction theory of the Selberg sieve, with applications,* J. Th. Nombres Bordeaux **18** (2006), 147–182.
[26] ———, *An inverse theorem for the Gowers $U^3(G)$ norm*, Proc. Edinburgh Math. Soc. **51**, no. 1, 73-153.
[27] ———, *Quadratic uniformity of the Möbius function*, to appear in Annales de l'Institut Fourier (Grenoble).
[28] G.H. Hardy and J.E. Littlewood *Some problems of "partitio numerorum"; III: On the expression of a number as a sum of primes,* Acta Math. **44** (1923), 1–70.
[29] H. A. Helfgott, *The parity problem for irreducible polynomials,* preprint.
[30] ———, *The parity problem for reducible polynomials,* J. London Math. Soc. (2) **73** (2006), 415-435.
[31] B. Host, *Progressions arithmétiques dans les nombres premiers* (*d'aprés B. Green and T. Tao*), Séminaire Bourbaki, Mars 2005, 57ieme année, 2004-2005, no. 944.
[32] B. Host and B. Kra, *Nonconventional ergodic averages and nilmanifolds,* Ann. of Math. (2) **161** (2005), no. 1, 397–488.
[33] L. K. Hua, *Additive Theory of Prime Numbers,* American Mathematical Society, Translations of Mathematical Monographs **13**, Providence, Rhode Island, 1966.
[34] H. Iwaniec and E. Kowalski, *Analytic number theory,* American Mathematical Society Colloquium Publications, **53**.
[35] B. Kra, *From ergodic theory to combinatorics and back again,* International Congress of Mathematicians. Vol. III, 57-76, Eur. Math. Soc., Zurich, 2006.
[36] A. Leibman, *Personal communication.*
[37] A. Mal'cev, *On a class of homogeneous spaces*, Izvestiya Akad. Nauk SSSR, Ser Mat. **13** (1949), 9–32.
[38] J. Steiner, *Über paralelle Flächen,* Jbr. Preuss. Akad. Wiss., (1840) 114–118 (Ges. Werke Vol. II, Reiner, Berlin (1882), pp173–176.
[39] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299–345.
[40] T. C. Tao, *A quantitative ergodic theory proof of Szemerédi's theorem*, Electron. J. Combin. **13** (2006). 1 No. 99, 1-49.

[41] ———, *Arithmetic progressions in the primes*, Collectanea Mathematica (2006), Vol. Extra, 37-88. [Proceedings, 7th International Conference on Harmonic Analysis and Partial Differential Equations.]
[42] ———, *Obstructions to uniformity, and arithmetic patterns in the primes*, Quarterly J. Pure Appl. Math. **2** (2006), 199–217.
[43] ———, *A remark on Goldston-Yıldırım correlation estimates,* available at `http://www.math.ucla.edu/~tao/preprints/Expository/gy-corr.dvi`
[44] ———, *A variant of the hypergraph removal lemma*, J. Combin. Thy. A **113** (2006), 1257–1280.
[45] ———, *The Gaussian primes contain arbitrarily shaped constellations*, J. dAnalyse Mathematique **99** (2006), 109-176.
[46] T. Tao and V. Vu, *Additive combinatorics*, Cambridge University Press, 2006.
[47] J.G. van der Corput, *Über Summen von Primzahlen und Primzahlquadraten,* Math. Ann. **116** (1939), 1–50.
[48] I. M. Vinogradov, *Some theorems concerning the primes,* Mat. Sbornik. N.S. **2** (1937), 179–195.
[49] A. Zygmund, *Trigonometric series,* Vol. I, II. Third edition. Cambridge Mathematical Library. Cambridge University Press, Cambridge, 2002. xii; Vol. I: xiv+383 pp.; Vol. II: viii+364 pp.

Centre for Mathematical Sciences, Wilberforce Road, Cambridge CB3 0WA, England

E-mail address: `b.j.green@dpmms.cam.ac.uk`

Department of Mathematics, UCLA, Los Angeles CA 90095-1555, USA.

E-mail address: `tao@math.ucla.edu`