POLYNOMIAL PATTERNS IN THE PRIMES

TERENCE TAO[1] and TAMAR ZIEGLER[2]

[1] UCLA Department of Mathematics, Los Angeles, CA 90095-1596, USA;
email: tao@math.ucla.edu
[2] Einstein Institute of Mathematics, The Hebrew University of Jerusalem,
Jerusalem, 91904, Israel;
email: tamarz@math.huji.ac.il

Abstract

Let $P_1, \ldots, P_k : \mathbb{Z} \to \mathbb{Z}$ be polynomials of degree at most d for some $d \geqslant 1$, with the degree d coefficients all distinct, and admissible in the sense that for every prime p, there exists integers n, m such that $n + P_1(m), \ldots, n + P_k(m)$ are all not divisible by p. We show that there exist infinitely many natural numbers n, m such that $n + P_1(m), \ldots, n + P_k(m)$ are simultaneously prime, generalizing a previous result of the authors, which was restricted to the special case $P_1(0) = \cdots = P_k(0) = 0$ (though it allowed for the top degree coefficients to coincide). Furthermore, we obtain an asymptotic for the number of such prime pairs n, m with $n \leqslant N$ and $m \leqslant M$ with M slightly less than $N^{1/d}$. This asymptotic is already new in general in the homogeneous case $P_1(0) = \cdots = P_k(0) = 0$. Our arguments rely on four ingredients. The first is a (slightly modified) generalized von Neumann theorem of the authors, reducing matters to controlling certain averaged local Gowers norms of (suitable normalizations of) the von Mangoldt function. The second is a more recent concatenation theorem of the authors, controlling these averaged local Gowers norms by global Gowers norms. The third ingredient is the work of Green and the authors on linear equations in primes, allowing one to compute these global Gowers norms for the normalized von Mangoldt functions. Finally, we use the Conlon–Fox–Zhao densification approach to the transference principle to combine the preceding three ingredients together. In the special case $P_1(0) = \cdots = P_k(0) = 0$, our methods also give infinitely many n, m with $n + P_1(m)$, $\ldots, n + P_k(m)$ in a specified set primes of positive relative density δ, with m bounded by $\log^L n$ for some L independent of the density δ. This improves slightly on a result from our previous paper, in which L was allowed to depend on δ.

2010 Mathematics Subject Classification: 11B30, 11N32 (primary)

1. Introduction

In [**12**] Green and the first author established that the primes contain arbitrarily long arithmetic progressions. Equivalently, one has

THEOREM 1. *Let k be a natural number (that is an element of* $\mathbb{N} = \{1, 2, 3, \dots\}$*). Then there exist infinitely many natural numbers* n, m *such that* $n, n + m, \dots, n + (k - 1)m$ *are all prime.*

Among the ingredients of the proof of [**12**] was the deployment of the (global) Gowers uniformity norms introduced in [**8**, **9**], the development of a *generalized von Neumann theorem* controlling certain multiple averages by these norms, and Szemerédi's theorem [**26**] on arithmetic progressions.

Theorem 1 was generalized by the authors [**27**] (see also [**21**] for a subsequent refinement):

THEOREM 2. *Let* $P_1, \dots, P_k \in \mathbb{Z}[m]$ *be polynomials in one indeterminate* m *such that* $P_1(0) = \cdots = P_k(0) = 0$*. Then there exist infinitely many natural numbers* n, m *such that* $n + P_1(m), n + P_2(m), \dots, n + P_k(m)$ *are all prime.*

The proof of Theorem 2 broadly followed the arguments used to prove Theorem 1; for instance, a generalized von Neumann theorem continued to play a crucial role. However, there are some key differences between the two arguments. Most notably, the global Gowers uniformity norms used in [**12**] were replaced by more complicated *averaged local Gowers uniformity norms*, and the Szemerédi theorem was replaced with the multidimensional polynomial recurrence theorem of Bergelson and Leibman [**2**]. It was necessary to deploy the multidimensional version of this theorem (despite the one-dimensional nature of Theorem 2) in order to obtain some uniformity in the recurrence theorem with respect to a certain technical parameter W that arose in the proof. In [**27**], the natural numbers n, m could be chosen so that $m = O(n^{o(1)})$ as $n \to \infty$; in [**28**] this was improved to $m = O(\log^{O(1)} n)$ (with the implied constants depending on $P_1, \dots, P_k$). In [**21**] it was also shown that m could be taken to be one less than a prime.

In a series of papers [**13**, **14**, **16**], a different generalization of Theorem 1 was obtained, namely that the arithmetic progression $n, n + r, \dots, n + (k - 1)r$ was replaced by a more general pattern, and a more quantitative count of the prime patterns was established. We give a special case of the main results of that paper as follows. If A is a finite nonempty set, we use $|A|$ to denote its cardinality, and for any function $f : A \to \mathbb{C}$, we write $\mathbb{E}_{a \in A} f(a)$ for $(1/|A|) \sum_{a \in A} f(a)$. For any N, we let $[N]$ denote the discrete interval

$$[N] := \{n \in \mathbb{N} : n \leqslant N\},$$

and let $\Lambda\colon \mathbb{Z} \to \mathbb{R}$ denote the von Mangoldt function (thus $\Lambda(p^j) := \log p$ for all primes p and natural numbers j, with $\Lambda(n) = 0$ otherwise). All sums and products over p are understood to be restricted to the primes unless otherwise specified. We are using $O()$ and $o()$ asymptotic notation; we review our conventions for this in Section 1.1 below.

THEOREM 3. *Let $P_1, \ldots, P_k \in \mathbb{Z}[\mathrm{m}]$ be linear polynomials (thus $P_i(\mathrm{m}) = a_i\mathrm{m} + b_i$ for some integers a_i, b_i, for each $i = 1, \ldots, k$). Assume that the leading coefficients $a_1, \ldots, a_k$ of the polynomials $P_1, \ldots, P_k$ are distinct. Let N be an asymptotic parameter going to infinity, and let $M = M(N)$ be a quantity such that $M = o_{N \to \infty}(N)$ and $M \gg N \log^{-A} N$ for some fixed constant A. Then*

$$\mathbb{E}_{n \in [N]} \mathbb{E}_{m \in [M]} \Lambda(n + P_1(m)) \ldots \Lambda(n + P_k(m)) = \prod_p \beta_p + o_{N \to \infty}(1)$$

where for each prime p, β_p is the local factor

$$\beta_p := \mathbb{E}_{n \in \mathbb{Z}/p\mathbb{Z}} \mathbb{E}_{m \in \mathbb{Z}/p\mathbb{Z}} \Lambda_p(n + P_1(m)) \ldots \Lambda_p(n + P_k(m))$$

and $\Lambda_p\colon \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}$ is the function $\Lambda_p(n) := (p/(p-1))1_{n \neq 0 \bmod p}$.

We remark that it is easy to establish the absolute convergence of the product $\prod_p \beta_p$; see [13].

Proof. If N/M is a sufficiently slowly growing function of N, this follows immediately from the main theorem of [13] (specialized to the finite complexity tuple $\mathrm{n} + P_1(\mathrm{m}), \ldots, \mathrm{n} + P_k(\mathrm{m})$ of linear forms), together with the results of [14, 16] proving the two conjectures assumed in [13]. The extension to the case where N/M is allowed to grow as fast as a power of $\log N$ follows from the same arguments, as discussed in [6, Appendix A]. □

Specializing this theorem to the case $P_i(\mathrm{m}) := (k-1)\mathrm{m}$, one can compute that $\prod_p \beta_p$ is nonzero, and then it is an easy matter to see that Theorem 3 implies Theorem 1. The Hardy–Littlewood prime tuples conjecture [17] predicts that the condition that the leading coefficients of the $P_1, \ldots, P_k$ are distinct can be relaxed to the condition that the $P_1, \ldots, P_k$ themselves are distinct, which would imply (among other things) the twin prime conjecture, but this is unfortunately well beyond the known techniques used to prove this theorem. The arguments in [13] also treat the case when M/N is comparable to one, as long as an Archimedean local factor β_∞ is inserted on the right-hand side, but to simplify the arguments slightly we work in the local setting $M = o_{N \to \infty}(N)$ to avoid having to consider the Archimedean factor. The prime tuples conjecture

also predicts that the condition $M \gg N \log^{-A} N$ can be replaced with the much weaker condition that M goes to infinity, but again this is beyond the reach of current methods. Even the special case $\mathbb{E}_{n \in [N]} \mathbb{E}_{m \in [M]} \Lambda(n) \Lambda(n + m) = 1 + o_{N \to \infty}(1)$, which essentially measures the error term for the prime number theorem in short intervals on the average, is only known for $M \geqslant N^{1/6+\varepsilon}$ (using the zero-density estimates of Huxley [18]), or for $M \geqslant N^\varepsilon$ assuming the Riemann hypothesis, while to the best of our knowledge the expected asymptotic for $\mathbb{E}_{n \in [N]} \mathbb{E}_{m \in [M]} \Lambda(n) \Lambda(n + m) \Lambda(n + 2m)$ is only known for $M \geqslant N^{5/8+\varepsilon}$ (using the exponential sum estimates of Zhan [30]). (If one replaces the von Mangoldt function with the Möbius function, then there is recent work [22, 23] obtaining such asymptotics for H growing arbitrarily slowly with N. However, the techniques used rely heavily on small prime divisors, and so do not seem to be directly applicable to problems involving the von Mangoldt function.)

The proof of Theorem 3 used the same global Gowers uniformity norms that appeared in the proof of Theorem 1, as well as a very similar generalized von Neumann theorem. However, Szemerédi's theorem was no longer used, as this result does not hold for arbitrary linear polynomials $P_1, \ldots, P_k$ and in any event only provides lower bounds for multiple averages, as opposed to asymptotics. Instead, by using the results of [14, 16] together with some transference arguments, it was shown that a suitable normalization $\Lambda'_{b,W} - 1$ of the von Mangoldt function Λ was small with respect to the global Gowers uniformity norm, which is sufficient to establish the stated result thanks to the generalized von Neumann theorem.

The first main result of this paper is a higher degree generalization of Theorem 3, which (except for a technical additional condition regarding the lower bound on M) is to Theorem 2 as Theorem 3 is to Theorem 1. More precisely, we show

THEOREM 4 (Main theorem). *Let d, r be natural numbers, and let $P_1, \ldots, P_k \in \mathbb{Z}[m_1, \ldots, m_r]$ be polynomials of integer coefficients of degree at most d. Furthermore, assume that the degree d components of $P_1, \ldots, P_k$ are all distinct (or equivalently, that $P_i - P_j$ has degree exactly d for all $1 \leqslant i < j \leqslant k$). Let N be an asymptotic parameter going to infinity, and let $M = M(N)$ be a quantity such that $M/N^{1/d}$ goes to zero sufficiently slowly as $N \to \infty$ (that is to say, there is a quantity $\omega(N)$ going to zero as $N \to \infty$ depending only on d, $r, k, P_1, \ldots, P_k$, and we assume that $M = o_{N \to \infty}(N^{1/d})$ and $M \geqslant \omega(N) N^{1/d}$). Then*

$$\mathbb{E}_{n \in [N]} \mathbb{E}_{\vec{m} \in [M]^r} \Lambda(n + P_1(\vec{m})) \ldots \Lambda(n + P_k(\vec{m})) = \prod_p \beta_p + o_{N \to \infty}(1) \qquad (1)$$

where β_p is the local factor

$$\beta_p := \mathbb{E}_{n \in \mathbb{Z}/p\mathbb{Z}} \mathbb{E}_{\vec{m} \in (\mathbb{Z}/p\mathbb{Z})^r} \Lambda_p(n + P_1(\vec{m})) \ldots \Lambda_p(n + P_k(\vec{m})) \qquad (2)$$

and $\Lambda_p \colon \mathbb{Z}/p\mathbb{Z} \to \mathbb{R}$ is the function $\Lambda_p(n) := (p/(p-1)) 1_{n \neq 0 \bmod p}$.

Note from [27, Lemma 9.5] gives the asymptotic $\beta_p = 1 + O_{P_1,\ldots,P_k}(1/p^2)$ for all sufficiently large p (noting that all such p are 'good' in the sense of [27, Definition 9.4]), so the product $\prod_p \beta_p$ is convergent.

It is likely that the lower bound $M \geqslant \omega(N)N^{1/d}$ can be relaxed to $M \geqslant N^{1/d} \log^{-A} N$ as in Theorem 3, and the upper bound relaxed to $M = O(N)$ at the cost of inserting an Archimedean factor β_∞ on the right-hand side as in [13], but we do not attempt to establish these extensions here to simplify the exposition. As will be clear from the method of proof, one can allow for much smaller values of M—in principle, as small as $\log^L N$ for some large L—as soon as one is able to establish some local Gowers uniformity for (a 'W-tricked' modification of) the von Mangoldt function at scale M^d.

As with Theorem 3, standard conjectures such as the Bateman–Horn conjecture [1] predict that Theorem 4 continues to hold without the requirement that the degree d components of P_i are distinct (so long as the P_i themselves remain distinct), and with M growing arbitrarily slowly with N. Such strengthenings of Theorem 4 remain beyond the methods here. We also remark that a result similar to Theorem 4, involving more general polynomial patterns, was established in [5] under an assumption of sufficiently large 'Birch rank' on the polynomial pattern, which is a rather different regime to the one considered here in as it tends to require a large number of variables compared to the number and degree of polynomials. In the recent paper [3], some special cases of (4) were established, in particular the case when $r = 2$ and $P_i(m_1, m_2) = i(m_1^2 + m_2^2)$ for $i = 1, \ldots, k$.

As in [13], we have a qualitative corollary of Theorem 4:

COROLLARY 1 (Qualitative main theorem). *Let d, r be natural numbers, $P_1, \ldots, P_k \in \mathbb{Z}[m_1, \ldots, m_r]$ be polynomials of degree at most d. Assume that the degree d components of $P_1, \ldots, P_k$ are all distinct, and suppose that for each prime p there exist $n \in \mathbb{Z}$ and $\vec{m} \in \mathbb{Z}^r$ such that $n + P_1(\vec{m}), \ldots, n + P_k(\vec{m})$ are all not divisible by p. Then there exist infinitely many natural numbers $n, m_1, \ldots, m_r$ such that $n + P_1(m_1, \ldots, m_r), \ldots, n + P_k(m_1, \ldots, m_r)$ are simultaneously prime.*

Proof. (Sketch) From hypothesis, the local factors β_p are all nonzero, and one can establish the asymptotic $\beta_p = 1 + O_{k,d,r}(1/p^2)$ for all sufficiently large

primes p. We conclude that the Euler product $\prod_p \beta_p$ is nonzero, and the claim then follows from Theorem 4 since the contribution when one of the $n + P_i(m_1, \ldots, m_r)$ is a prime power, rather than a prime, can easily be shown to be negligible. $\qquad\square$

The case $r = 1, k = 2$ of this corollary had been previously established in [20]. As a typical example of Corollary 1, we see that if $P \in \mathbb{Z}[m]$ is any intersective polynomial (that is to say, a polynomial such as $P(m) = (m^2-2)(m^2-3)(m^2-6)$ that has a root in every modulus), then there exist arbitrarily long arithmetic progressions of primes whose spacing is of the form $P(m)$.

Corollary 1 is a partial generalization of Theorem 2, in that it implies the special case of that theorem when $r = 1$ and the m^d coefficients of the polynomials $P_1(m), \ldots, P_k(m)$ are all distinct. As mentioned above, the arguments here should eventually extend to allow the m^d coefficients of P_i to be equal, so long as the $P_i - P_j$ are nonconstant, once one can establish local Gowers norm control on the von Mangoldt function. In fact, Schinzel's hypothesis H [25] predicts that Corollary 1 should hold even if some of the $P_i - P_j$ are constant, but this claim (which includes the twin prime conjecture as a special case) is well beyond the methods of this paper.

We now briefly summarize the method of proof of Theorem 4. If one directly applies the methods used to prove Theorem 3, replacing (a variant of) the generalized von Neumann theorem from [12] with (a variant of) the more complicated generalized von Neumann theorem from [27], one ends up wishing to control various normalized versions $\Lambda'_{b,W} - 1$ of the von Mangoldt function in certain averaged local Gowers uniformity norms; furthermore, the transference machinery in [12, 27] allows one to replace $\Lambda_{b,W} - 1$ by a bounded function. At this point, one would like to apply an inverse theorem for the averaged local Gowers uniformity norms, but a direct application of the known inverse theorems does not lead to a particularly tractable condition to verify on $\Lambda_{b,W} - 1$. To overcome this issue, we use the concatenation theorems recently developed by us in [29] to control the averaged local Gowers uniform norms arising from the generalized von Neumann theorem by a global Gowers uniform norm. It is at this juncture that the hypothesis that the m^d coefficients of the $P_i(m)$ are all distinct, together with the choice of M as being close to $N^{1/d}$, becomes crucial.

In some cases, we are able to use our methods to partially remove the requirement in Theorem 4 that the degree d components of $P_1, \ldots, P_k$ are distinct, although this requires one to understand the distribution (or Gowers uniformity) of primes in short intervals, for which the known unconditional results still fall well short of what is conjecturally true. As an example of this, we

give the conjectural asymptotic for prime triplets of the form $n, n+m, n+P(m)$, with $n \in [N]$ and m slightly smaller than $N^{1/d}$, when P has degree exactly d:

THEOREM 5. *Theorem 4 is true in the case $r = 1$, $k = 3$, $P_1 = 0$, $P_2 = \mathrm{m}$, and when P_3 has degree exactly d for some $2 \leqslant d \leqslant 5$. If one assumes the generalized Riemann hypothesis (GRH), the condition $d \leqslant 5$ may be removed.*

This theorem will be proven at the end of Section 5. The main idea is to use the machinery of proof of Theorem 4 to essentially eliminate the $\Lambda(n + P_3(m))$ factor, leaving only the task of controlling averages roughly of the form $\mathbb{E}_{n \in [N]} \mathbb{E}_{m \in [M]} \Lambda(n) \Lambda(n + m)$, which can be handled by existing results on primes in short intervals, both with and without GRH. In Remark 2, we briefly discuss some other cases that can in principle be handled by this method, such as $r = 1$, $k = 4$, $P_1 = 0$, $P_2 = \mathrm{m}^k$, $P_2 = 2\mathrm{m}^k$, $P_3 = \mathrm{m}^d$ for certain choices of k, d.

The above results have (somewhat simpler) analogues when the von Mangoldt function Λ is replaced by the Möbius function μ (or the closely related Liouville function λ). In these analogues, the local factors $\prod_p \beta_p$ should be deleted, thus for instance we have

$$\mathbb{E}_{n \in [N]} \mathbb{E}_{m \in [M]^r} \mu(n + P_1(m)) \dots \mu(n + P_k(m)) = o_{N \to \infty}(1)$$

under the hypotheses of Theorem 4. The proof of these variant results is in fact significantly simpler, as all the pseudorandom measures ν that appear in the arguments below can be simply replaced by 1; also, the 'W-trick' is not needed in this case. We leave the modifications of the arguments below needed to obtain these variants to the interested reader.

The methods used to establish Corollary 1 also give a variant involving sets of primes of positive upper density, improving slightly on our previous results in [28].

THEOREM 6 (Narrow polynomial patterns in subsets of the primes). *Let d be a natural number, and let $P_1, \dots, P_k \in \mathbb{Z}[\mathrm{m}]$ be polynomials of integer coefficients of degree at most d, such that $P_1(0) = \dots = P_k(0) = 0$. Let L be a sufficiently large quantity depending on d. Let $\delta > 0$, and let $\mathcal{A}$ be a subset of the primes $\mathcal{P}$ such that*

$$\limsup_{N \to \infty} \frac{|\mathcal{A} \cap [1, N]|}{|\mathcal{P} \cap [1, N]|} \geqslant \delta$$

for some $\delta > 0$. Then one can find infinitely many natural numbers n, m with $n + P_1(m), \dots, n + P_k(m)$ in $\mathcal{A}$ and with $m \leqslant \log^L n$, where L depends only on d, k.

Comparing this result with Corollary 1, we see that the set of primes $\mathcal{P}$ has been replaced by a positive density subset $\mathcal{A}$, and that the condition that the $P_i - P_j$ have degree exactly d has been dropped, replaced instead by the hypothesis $P_1(0) = \cdots = P_k(0)$; we have also set $r = 1$ (as the $r > 1$ case follows easily from the $r = 1$ case, for example by restricting to the diagonal $m_1 = \cdots = m_r$). Another key point is the smallness condition on m. The arguments in [27] essentially established this theorem with the bound $m = n^{o(1)}$, and the subsequent argument in [28] improved this to $m \leqslant \log^{L(\delta)} n$ where the exponent $L(\delta)$ is permitted to depend on δ in addition to d and k. Thus the new contribution of Theorem 6 is the removal of the dependence of L on δ. This follows from the arguments in [28], after replacing [28, Theorem 9] with Theorem 11 below; see Remark 1.

1.1. Notation. If $x \in \mathbb{R}/\mathbb{Z}$, we use $\|x\|_{\mathbb{R}/\mathbb{Z}}$ to denote the distance from x to the nearest integer, and $e(x) := e^{2\pi i x}$.

Given two real numbers A, B with $A \leqslant B$, we use $[A, B]$ to denote the discrete interval $\{n \in \mathbb{Z} : A \leqslant n \leqslant B\}$. We often need to identify this interval with a subset of the cyclic group $\mathbb{Z}/N\mathbb{Z}$ for some modulus N (larger than $B - A$), which can of course be done by applying the reduction map $n \mapsto n \ (N)$ from $\mathbb{Z}$ to $\mathbb{Z}/N\mathbb{Z}$. Similarly for the interval $[M] := \{n \in \mathbb{N} : n \leqslant M\}$.

Given any finite collection $\mathrm{m}_1, \ldots, \mathrm{m}_r$ of indeterminates, we write $\mathbb{Z}[\mathrm{m}_1, \ldots, \mathrm{m}_r]$ for the ring of formal polynomials P in these variables with integer coefficients, thus there is a natural number d for which one has

$$P = \sum_{i_1, \ldots, i_r \geqslant 0 : i_1 + \cdots + i_r \leqslant d} \alpha_{i_1, \ldots, i_r} \mathrm{m}_1^{i_1} \ldots \mathrm{m}_r^{i_r}$$

for some integers $\alpha_{i_1, \ldots, i_r}$. The least such d for which one has such a representation is the *degree* of P. Of course, one can evaluate $P(m_1, \ldots, m_r)$ for any elements $m_1, \ldots, m_r$ of a commutative ring (such as $\mathbb{Z}$ or $\mathbb{Z}/q\mathbb{Z}$) simply by substituting the indeterminates m_i with their evaluations m_i. We write indeterminate variables such as m, n in Roman font, to distinguish them from elements m, n of a specific ring such as $\mathbb{Z}$ or $\mathbb{Z}/q\mathbb{Z}$.

It will be convenient to work with the notion of a *finite multiset*—an unordered collection $\{a_1, \ldots, a_n\}$ of a finite number of objects $a_1, \ldots, a_n$, in which repetitions are allowed. This clearly generalizes the notion of a finite set, in which every element occurs with multiplicity one. If $A = \{a_1, \ldots, a_n\}$ is a nonempty finite multiset, and $f : A \to \mathbb{C}$ is a function on the elements of A, we write

$$\mathbb{E}_{a \in A} f(a) := \frac{1}{n} \sum_{i=1}^{n} f(a_i),$$

which of course generalizes the notion of an average $\mathbb{E}_{a \in A} f(a)$ over a finite nonempty set A; thus for instance $\mathbb{E}_{a \in \{1,2,3,3\}} a = \frac{1+2+3+3}{4}$. If $A = \{a_1, \ldots, a_n\}$ and $B = \{b_1, \ldots, b_m\}$ are finite multisets taking values in an additive group $G = (G, +)$, we write

$$A + B := \{a_i + b_j : i = 1, \ldots, n; j = 1, \ldots, m\}$$

and

$$A - B := \{a_i - b_j : i = 1, \ldots, n; j = 1, \ldots, m\}$$

for the sum and difference multisets. For any integer k, we also write

$$kA := \{ka : a \in A\}$$

(in particular, kA does *not* denote the k-fold iterated sumset of A).

Given a finite nonempty multiset A in a domain X, we define the *density function* $p_A \colon X \to [0, 1]$ to be the quantity

$$p_A(x) := \frac{|\{a \in A : a = x\}|}{|A|}$$

where the numerator is the multiplicity of x in A; thus $\sum_{x \in X} p_A(x) = 1$, and $|A|$ the cardinality of A counting multiplicity; more generally, we define

$$\mathbb{E}_{a \in A} f(a) := \sum_{x \in X} f(x) p_A(x)$$

for all $f \colon X \to \mathbb{R}$. We then define the *total variation distance* $d_{\mathrm{TV}}(A, B)$ between two finite nonempty multisets A, B in X to be

$$d_{TV}(A, B) := \sum_{x \in X} |p_A(x) - p_B(x)|.$$

Thus we have

$$|\mathbb{E}_{a \in A} f(a) - \mathbb{E}_{b \in B} f(b)| \leqslant d_{TV}(A, B) \tag{3}$$

whenever $f \colon X \to \mathbb{R}$ is bounded in magnitude by 1. Informally, if $d_{\mathrm{TV}}(A, A+B)$ is small, one can think of A as having an approximate translation invariance with respect to shifts by B. We observe from the triangle inequality and translation invariance that one has the contraction property

$$d_{\mathrm{TV}}(A + C, B + C) \leqslant \mathbb{E}_{c \in C} d_{\mathrm{TV}}(A + c, B + c) = d_{\mathrm{TV}}(A, B) \tag{4}$$

for any finite nonempty multisets A, B, C. In particular, for finite nonempty multisets A, B, C one has

$$
\begin{aligned}
d_{\mathrm{TV}}(A, A + C) &\leqslant d_{\mathrm{TV}}(A, A + B) + d_{\mathrm{TV}}(A + B, A + B + C) \\
&\quad + d_{\mathrm{TV}}(A + B + C, A + C) \\
&\leqslant d_{\mathrm{TV}}(A, A + B) + d_{\mathrm{TV}}(B, B + C) + d_{\mathrm{TV}}(A + B, A)
\end{aligned}
$$

and thus

$$
d_{\mathrm{TV}}(A, A + C) \leqslant 2 d_{\mathrm{TV}}(A, A + B) + d_{\mathrm{TV}}(B, B + C). \tag{5}
$$

Informally: if A is approximately B-invariant, and B is approximately C-invariant, then A is approximately C-invariant.

Let $G = (G, +)$ be a finite abelian group, and let $Q_1, \ldots, Q_d$ be nonempty finite multisets in G with $d \geqslant 1$. Given a function $f \colon G \to \mathbb{R}$, we define the *Gowers box norm* $\|f\|_{\square_{Q_1,\ldots,Q_d}^d(G)} = \|f\|_{\square_{Q_1,\ldots,Q_d}^d}$ to be the nonnegative real defined by the formula

$$
\|f\|_{\square_{Q_1,\ldots,Q_d}^d}^{2^d} := \mathbb{E}_{x \in G} \mathbb{E}_{h_1 \in Q_1 - Q_1, \ldots, h_d \in Q_d - Q_d} \prod_{\omega \in \{0,1\}^d} f\left(x + \sum_{i=1}^{d} \omega_i h_i\right);
$$

it is easy to see that the right-hand side is nonnegative, so that the $\square_{Q_1,\ldots,Q_d}^d$ norm is well defined. More generally, given functions $f_\omega \colon G \to \mathbb{R}$ for $\omega \in \{0,1\}^d$, we define the *Gowers inner product*

$$
\langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{\square_{Q_1,\ldots,Q_d}^d} := \mathbb{E}_{x \in G} \mathbb{E}_{h_1 \in Q_1 - Q_1, \ldots, h_d \in Q_d - Q_d} \prod_{\omega \in \{0,1\}^d} f_{\omega_1,\ldots,\omega_d}\left(x + \sum_{i=1}^{d} \omega_i h_i\right).
$$

The Gowers norms can also be defined for complex-valued functions by appropriate insertion of complex conjugation operations, but we will not need to do so here. We recall the *Cauchy–Schwarz–Gowers inequality*

$$
\left| \langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{\square_{Q_1,\ldots,Q_d}^d} \right| \leqslant \prod_{\omega \in \{0,1\}^d} \|f_\omega\|_{\square_{Q_1,\ldots,Q_d}^d} \tag{6}
$$

(see for example [13, Lemma B.2]). Among other things, this implies the monotonicity property

$$
\|f\|_{\square_{Q_1,\ldots,Q_{d-1}}^{d-1}} \leqslant \|f\|_{\square_{Q_1,\ldots,Q_d}^d} \tag{7}
$$

for $d > 1$, by setting the f_ω in (6) to be f or 1 in an appropriate fashion. It also implies the triangle inequality for the $\square_{Q_1,\ldots,Q_d}^d$ norm.

When $Q_1 = \cdots = Q_d = Q$, we refer to the Gowers box norm $\||\cdot\||_{\square^d_{Q,\ldots,Q}(G)}$ as the *Gowers uniformity norm* and abbreviate it as $\||\cdot\||_{U^d_Q(G)}$ or simply $\||\cdot\||_{U^d_Q}$. Similarly for the Gowers inner product.

We observe the identity

$$\|f\|^{2^d}_{\square^d_{Q_1,\ldots,Q_d}} = \mathbb{E}_{x \in G} f(x) \mathcal{D}^d_{Q_1,\ldots,Q_d}(f)(x)$$

where the *dual function* $\mathcal{D}^d_{Q_1,\ldots,Q_d}(f) \colon G \to \mathbb{R}$ is defined by

$$\mathcal{D}^d_{Q_1,\ldots,Q_d}(f)(x) := \mathbb{E}_{h_1 \in Q_1 - Q_1,\ldots,h_d \in Q_d - Q_d} \prod_{(\omega_1,\ldots,\omega_d) \in \{0,1\}^d \setminus \{0\}^d} f\left(x + \sum_{i=1}^d \omega_i h_i\right).$$

Again, when $Q_1 = \cdots = Q_d = Q$, we abbreviate $\mathcal{D}^d_{Q,\ldots,Q}$ as $\mathcal{D}^d_Q$. Finally, we recall the L^p norms

$$\|f\|_{L^p(G)} = \|f\|_{L^p} := (\mathbb{E}_{x \in G}|f(x)|^p)^{1/p}$$

for $1 \leqslant p < \infty$, with the usual convention

$$\|f\|_{L^\infty(G)} = \|f\|_{L^\infty} := \sup_{x \in G} |f(x)|.$$

We now set out the asymptotic notation we use. We write $X = O(Y)$, $X \ll Y$, or $Y \gg X$ to denote the estimate $|X| \leqslant CY$ for a constant C. Often, we need the implied C to depend on some parameters, which we indicate with subscripts, thus for instance $X = O_{d,r}(Y)$ denotes the estimate $|X| \leqslant C_{d,r}Y$ for some quantity $C_{d,r}$ depending only on d, r. We will often also be working with an asymptotic parameter such as N going to infinity. In that setting, some quantities will be held fixed (that is, they will be independent of N), while others will be allowed to depend on N. We write $X = o_{N \to \infty}(Y)$, or $X = o(Y)$ for short, if we have $|X| \leqslant c(N)Y$ for some quantity $c(N)$ depending on N and possibly on other fixed quantities that goes to zero as $N \to \infty$ (holding all fixed quantities constant).

2. Controlling averaged Gowers norms by global Gowers norms

A key step in our arguments is establishing that averaged Gowers norms, with shifts depending polynomially on the averaging parameter in a nondegenerate fashion, can be controlled by global Gowers norms. We begin with a polynomial equidistribution lemma, which, roughly speaking, asserts that exponential sums with polynomial phases can only be large if the polynomial is 'major arc'.

LEMMA 1 (Polynomial equidistribution). *Let $d, r \geqslant 1$ be natural numbers, and let $P \in \mathbb{R}[n_1, \ldots, n_r]$ be a polynomial of degree at most d, thus*

$$P(n_1, \ldots, n_r) = \sum_{i_1, \ldots, i_r \geqslant 0 : i_1 + \cdots + i_r \leqslant d} \alpha_{i_1, \ldots, i_r} n_1^{i_1} \ldots n_r^{i_r}$$

for some coefficients $\alpha_{i_1, \ldots, i_r} \in \mathbb{R}$. Let $N_1, \ldots, N_r \geqslant 1$ and $0 < \varepsilon \leqslant 1$ be such that

$$|\mathbb{E}_{n_1 \in [N_1], \ldots, n_r \in [N_r]} e(P(n_1, \ldots, n_r))| \geqslant \varepsilon. \tag{8}$$

Then either one has $N_j \ll_{d,r} \varepsilon^{-O_{d,r}(1)}$ for some $1 \leqslant j \leqslant r$, or else there exists a natural number $q \ll_{d,r} \varepsilon^{-O_{d,r}(1)}$ such that

$$\|q\alpha_{i_1, \ldots, i_r}\|_{\mathbb{R}/\mathbb{Z}} \ll_{d,r} \frac{\varepsilon^{-O_{d,r}(1)}}{N_1^{i_1} \ldots N_r^{i_r}}$$

for all $i_1, \ldots, i_r \geqslant 0$ with $i_1 + \cdots + i_r \leqslant d$.

Proof. In the one-dimensional case $r = 1$, this follows from standard Weyl sum estimates (see for example [**15**, Proposition 4.3]). Now suppose inductively that $r > 1$, and that the claim has already been proven for $r - 1$. For brevity, we allow all implied constants in the asymptotic notation here to depend on r, d.

From (8) we have

$$|\mathbb{E}_{n_1 \in [N_1], \ldots, n_{r-1} \in [N_{r-1}]} e(P(n_1, \ldots, n_r))| \gg \varepsilon$$

for at least $\gg \varepsilon N_r$ values of $n_r \in [N_r]$. On the other hand, we can write

$$P(n_1, \ldots, n_r) = \sum_{i_1, \ldots, i_{r-1} \geqslant 0 : i_1 + \cdots + i_{r-1} \leqslant d} \alpha_{i_1, \ldots, i_{r-1}}(n_r) n_1^{i_1} \ldots n_{r-1}^{i_{r-1}}$$

where for each $i_1, \ldots, i_{r-1}$, $\alpha_{i_1, \ldots, i_{r-1}} \in \mathbb{Z}[n_r]$ is the polynomial

$$\alpha_{i_1, \ldots, i_{r-1}}(n_r) := \sum_{i_r \geqslant 0 : i_1 + \cdots + i_r \leqslant d} \alpha_{i_1, \ldots, i_r} n_r^{i_r}.$$

Applying the induction hypothesis, we see that for $\gg \varepsilon N_r$ values of n_r, one can find $q \ll \varepsilon^{-O(1)}$ such that

$$\|q\alpha_{i_1, \ldots, i_{r-1}}(n_r)\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{\varepsilon^{-O(1)}}{N_1^{i_1} \ldots N_{r-1}^{i_{r-1}}} \tag{9}$$

for all $i_1, \ldots, i_{r-1}$. At present, q can depend on n_r, but by the pigeonhole principle we can pass to a set of $\gg \varepsilon^{O(1)} N_r$ values of n_r and make q independent

of n_r. Then, for each $i_1, \ldots, i_{r-1}$, we have (9) for $\gg \varepsilon^{O(1)} N_r$ values of n_r. Applying [**15**, Lemma 4.5], this implies that either $N_r \ll \varepsilon^{-O(1)}$, or else there exists a natural number $k \ll \varepsilon^{-O(1)}$ (possibly depending on $i_1, \ldots, i_{r-1}$) such that

$$\|kq\alpha_{i_1,\ldots,i_{r-1},i_r}\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{\varepsilon^{-O(1)}}{N_1^{i_1} \ldots N_{r-1}^{i_{r-1}} N_r^{i_r}}$$

for all i_r. At present k can depend on $i_1, \ldots, i_{r-1}$, but by multiplying together all the k associated to the $O(1)$ possible choices of $(i_1, \ldots, i_{r-1})$, we may take k independent of $i_1, \ldots, i_{r-1}$. Replacing q by kq, we obtain the claim. □

Our next result asserts (roughly speaking) that on the average, a multidimensional progression

$$Q(\vec{h}) := P_1(\vec{h}_1)[-M, M] + \cdots + P_k(\vec{h}_k)[-M, M]$$

(with $\vec{h}$ of the order of M, and $P_1, \ldots, P_k$ polynomials of degree exactly $d-1$) will have an approximate global translation symmetry, in the sense that $Q(\vec{h})$ is close on total variation to $Q(\vec{h}) + q[-A^{-2k}M^d, A^{-2k}M^d]$ for some reasonably small q and A, if k is large enough depending on d, r. The equidistribution result from Lemma 1 will play a key role in the proof of this statement.

THEOREM 7 (Approximate global symmetry). *Let $d, r \geqslant 1$ be natural numbers, and suppose that $k \geqslant 1$ is a natural number that is sufficiently large depending on d, r. Let A be a quantity that is sufficiently large depending on d, r, k, and for each $j = 1, \ldots, k$, let $P_j \in \mathbb{Z}[h_1, \ldots, h_r]$ be a polynomial of degree exactly $d-1$ with coefficients that are integers of magnitude at most A. Let M be a quantity that is sufficiently large depending on d, r, k, A. For any $\vec{h} = (\vec{h}_1, \ldots, \vec{h}_k)$ in $(\mathbb{Z}^r)^k$, let $Q(\vec{h})$ be the generalized arithmetic progression*

$$Q(\vec{h}) := P_1(\vec{h}_1)[-M, M] + \cdots + P_k(\vec{h}_k)[-M, M]$$

(viewed as a multiset in $\mathbb{Z}$) and let Q_0 be the arithmetic progression

$$Q_0 := [-A^{-2k}M^d, A^{-2k}M^d]$$

(also viewed as a multiset in $\mathbb{Z}$). Then one has

$$\mathbb{E}_{\vec{h} \in ([M]^r)^k} \inf_{1 \leqslant q \leqslant A^k} d_{\mathrm{TV}}(Q(\vec{h}), Q(\vec{h}) + qQ_0) \ll_{d,r,k} A^{-1}. \quad (10)$$

For future reference, we observe that the claim also holds when the arithmetic progressions are viewed as subsets of a cyclic group $\mathbb{Z}/N\mathbb{Z}$ rather than $\mathbb{Z}$, since applying reduction modulo N can only decrease the total variation norm.

Proof. Let $N := A^{2d} M^d$, then for any $\vec{h} \in ([M]^r)^k$ and $1 \leqslant q \leqslant A^k$, the projection of $\mathbb{Z}$ to $\mathbb{Z}/N\mathbb{Z}$ is injective on $Q(\vec{h})$ and $Q(\vec{h}) + qQ_0$. Thus, we may interpret these progressions instead as lying in the cyclic group $\mathbb{Z}/N\mathbb{Z}$. By Fourier expansion in $\mathbb{Z}/N\mathbb{Z}$, the density function $p_{Q(\vec{h})}$ of the multiset $Q(\vec{h})$ can be written as

$$p_{Q(\vec{h})}(x) = \frac{1}{N} \sum_{\xi \in \mathbb{Z}/N\mathbb{Z}} e(\xi x/N) \mathbb{E}_{a \in Q(\vec{h})} e(-a\xi/N)$$

which factorizes as

$$p_{Q(\vec{h})}(x) = \frac{1}{N} \sum_{\xi \in \mathbb{Z}/N\mathbb{Z}} e(\xi x/N) \prod_{j=1}^{k} D(P_j(\vec{h}_j)\xi)$$

where $D \colon \mathbb{Z}/N\mathbb{Z} \to [-1, 1]$ is the Dirichlet kernel

$$D(\xi) := \mathbb{E}_{m \in [-M,M]} e(-m\xi/N).$$

Similarly we have

$$p_{Q(\vec{h})+qQ_0}(x) = \frac{1}{N} \sum_{\xi \in \mathbb{Z}/N\mathbb{Z}} e(\xi x/N) D'(q\xi) \prod_{j=1}^{k} D(P_j(\vec{h}_j)\xi)$$

where D' is another Dirichlet kernel, defined by the formula

$$D'(\xi) := \mathbb{E}_{m \in [-A^{-2k}M^d, A^{-2k}M^d]} e(m\xi/N).$$

Subtracting and using the triangle inequality, we conclude that

$$d_{\mathrm{TV}}(Q(\vec{h}), Q(\vec{h}) + qQ_0) \leqslant \sum_{\xi \in \mathbb{Z}/N\mathbb{Z}} |D'(q\xi) - 1| \prod_{j=1}^{k} |D(P_j(\vec{h}_j)\xi)|$$

and so it suffices to show that

$$\mathbb{E}_{\vec{h} \in ([M]^r)^k} \inf_{1 \leqslant q \leqslant A^k} \sum_{\xi \in \mathbb{Z}/N\mathbb{Z}} |D'(q\xi) - 1| \prod_{j=1}^{k} |D(P_j(\vec{h}_j)\xi)| \ll_{d,r,k} A^{-1}. \qquad (11)$$

Call $\xi \in \mathbb{Z}/N\mathbb{Z}$ *major arc* if one has

$$\left\| \frac{q\xi}{N} \right\|_{\mathbb{R}/\mathbb{Z}} \leqslant A^{\sqrt{k}}/N$$

for some $1 \leqslant q \leqslant A^{\sqrt{k}}$, and *minor arc* otherwise. We first consider the contribution to (11) of the minor arcs. Here we bound $|D'(q\xi) - 1|$ by 2, and estimate this contribution by

$$2 \sum_{\xi}^{*} \prod_{j=1}^{k} \mathbb{E}_{\vec{h}_j \in [M]^r} |D(P_j(\vec{h}_j)\xi)|$$

where $\sum_{\xi}^{*}$ denotes a summation over minor arc ξ. By several applications of Hölder's inequality, we may bound this by

$$2 \prod_{j=1}^{k} \left(\sum_{\xi} (\mathbb{E}_{\vec{h}_j \in [M]^r} |D(P_j(\vec{h}_j)\xi)|^2)^{k/2} \right)^{1/k}.$$

We have the following distributional control on the expression $\mathbb{E}_{\vec{h}_j \in [M]^r} |D(P_j(\vec{h}_j)\xi)|^2$ appearing above:

PROPOSITION 1 (Distributional bound). *Let the notation and hypotheses be as above, and let n be a natural number. Let $1 \leqslant j \leqslant k$. Then there are at most $O_{d,r}((A2^n)^{O_{d,r}(1)})$ values of minor arc $\xi \in \mathbb{Z}/N\mathbb{Z}$ for which*

$$\mathbb{E}_{\vec{h}_j \in [M]^r} |D(P_j(\vec{h}_j)\xi)|^2 \geqslant 2^{-n}.$$

Furthermore, there are no such minor arc ξ unless

$$2^{-n} \leqslant A^{-k^{1/4}}. \tag{12}$$

Proof. For brevity we allow all implied constants in the asymptotic notation to depend on d and r. If we have $N \ll (2^n)^{O(1)}$ then (12) is trivial (since N is assumed large depending on d, r, k, A), and the first claim is also trivial since there are clearly at most N possible choices for ξ. Thus we may assume that we do not have an estimate of the form $N \ll (2^n)^{O(1)}$.

Assume that $\mathbb{E}_{\vec{h}_j \in [M]^r} |D(P_j(\vec{h}_j)\xi)|^2 \geqslant 2^{-n}$ for some natural number n.

We can expand $\mathbb{E}_{\vec{h}_j \in [M]^r} |D(P_j(\vec{h}_j)\xi)|^2$ as

$$\mathbb{E}_{\vec{h}_j \in [M]^r} |D(P_j(\vec{h}_j)\xi)|^2 = \mathbb{E}_{\vec{h}_j \in [M]^r} \mathbb{E}_{m,m' \in [-M,M]} e\left(\frac{m' P_j(\vec{h}_j)\xi}{N} - \frac{m P_j(\vec{h}_j)\xi}{N} \right).$$

Applying Lemma 1, we conclude that there is a natural number $q_j \ll (A2^n)^{O(1)}$ such that

$$\left\| \frac{q_j \xi}{N} \alpha_{j,i_1,\dots,i_r} \right\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{(A2^n)^{O(1)}}{M^{i_1 + \dots + i_r + 1}}$$

for all the coefficients $\alpha_{j,i_1,\ldots,i_r}$ of P_j. In particular, since P_j has at least one nonzero coefficient $\alpha_{j,i_1,\ldots,i_r}$ with $i_1 + \cdots + i_r = d - 1$, we have

$$\left\| \frac{q_j \xi}{N} \alpha_{j,i_1,\ldots,i_r} \right\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{(A2^n)^{O(1)}}{N} \tag{13}$$

for that coefficient. This is inconsistent with the minor arc hypothesis unless (12) holds.

For fixed q_j, the constraint (13) restricts ξ to at most $O((A2^n)^{O(1)})$ possible values; summing over q_j, we conclude the proposition. $\qquad\square$

Applying the above proposition for each n and summing, we conclude that

$$\sum_{\xi}^{*} (\mathbb{E}_{\vec{h}_j \in [M]^r} |D(P_j(\vec{h}_j)\xi)|^2)^{k/2} \ll_k \sum_{n: 2^{-n} \leqslant A^{-k^{1/4}}} (A2^n)^{O(1)} 2^{-nk/2}$$
$$\ll_k A^{-1}$$

if k is sufficiently large depending on d, r. Thus the contribution of the minor arcs to (11) is acceptable.

It remains to control the contribution to (11) of the major arc ξ. Note that the number of major arc ξ is $O(A^{O(\sqrt{k})})$. As a consequence, any choice of $\vec{h}$ and ξ for which

$$|D(P_j(\vec{h}_j)\xi)| \leqslant A^{-k}$$

for some $1 \leqslant j \leqslant k$ gives a negligible contribution to (11). Thus we may restrict attention to those $\vec{h}$ and ξ for which ξ is major arc and

$$|D(P_j(\vec{h}_j)\xi)| > A^{-k}$$

for all $j = 1, \ldots, k$. Computing the Dirichlet kernel $D(P_j(\vec{h}_j)\xi)$, this implies that

$$\left\| \frac{P_j(\vec{h}_j)\xi}{N} \right\|_{\mathbb{R}/\mathbb{Z}} \ll \frac{A^k}{N}$$

for all $j = 1, \ldots, k$. On the other hand, as ξ is major arc, we have

$$\left\| \frac{\xi}{N} - \frac{a_\xi}{q_\xi} \right\|_{\mathbb{R}/\mathbb{Z}} \leqslant A^{\sqrt{k}}/N \tag{14}$$

for some $1 \leqslant q_\xi \leqslant A^{\sqrt{k}}$ and a_ξ coprime to q_ξ. We also have $P_j(\vec{h}_j) \ll A^{O(1)} M^{d-1}$. As M is large, these estimates are only compatible with each other if q_ξ divides $P_j(\vec{h}_j)$. In particular, q_ξ divides the greatest common divisor of $P_1(\vec{h}_1), \ldots, P_k(\vec{h}_k)$.

Suppose that $\vec{h}$ is such that the greatest common divisor $(P_1(\vec{h}_1), \ldots, P_k(\vec{h}_k))$ of $P_1(\vec{h}_1), \ldots, P_k(\vec{h}_k)$ is at most A^k. Denoting this greatest common divisor by q, we see from (14) that

$$|D'(q\xi) - 1| \ll A^{-k+O(\sqrt{k})}$$

for all ξ that have not already been previously eliminated from consideration. Thus the contribution of these $\vec{h}$ to (11) is acceptable. Thus we only need to restrict attention to those $\vec{h}$ for which $(P_1(\vec{h}_1), \ldots, P_k(\vec{h}_k)) > A^k$. For this case, we bound $|D'(q\xi) - 1|$ by 2 and $D(P_j(\vec{h}_j)\xi)$ by 1, and use the fact that there are only $O(A^{O(\sqrt{k})})$ major arc ξ, and bound this contribution to (11) by

$$\ll A^{O(\sqrt{k})} \mathbb{E}_{\vec{h} \in ([M]^r)^k} 1_{(P_1(\vec{h}_1), \ldots, P_k(\vec{h}_k)) > A^k}.$$

Recall that the $P_j(\vec{h}_j)$ are of size $O(A^{O(1)} M^{d-1})$, and that M is larger than A^{2k}. This implies that if $(P_1(\vec{h}_1), \ldots, P_k(\vec{h}_k)) > A^k$, then either there is an integer q with $A^k < q \leqslant M$ that divides all of the $P_i(\vec{h}_i)$, or else there is a prime p with $M < p \ll A^{O(1)} M^{d-1}$ that divides all of the $P_i(\vec{h}_i)$. Thus we may bound (11) by

$$\ll A^{O(\sqrt{k})} \left(\sum_{A^k \leqslant q \leqslant M} c_q + \sum_{M < p \ll A^{O(1)} M^{d-1}} c_p \right) \tag{15}$$

where

$$c_q := \mathbb{E}_{\vec{h} \in ([M]^r)^k} 1_{q | P_1(\vec{h}_1), \ldots, P_k(\vec{h}_k)}.$$

We may factor

$$c_q := \prod_{i=1}^k \mathbb{E}_{\vec{h}_i \in [M]^r} 1_{q | P_i(\vec{h}_i)}. \tag{16}$$

To estimate the inner average we need the following Schwartz–Zippel type estimate.

LEMMA 2 (Schwartz–Zippel type estimate). *Let $P \in \mathbb{Z}[\mathrm{h}_1, \ldots, \mathrm{h}_r]$ be a polynomial of degree at most d for some $d \geqslant 1$, let p be a prime, and let $m \geqslant 0$ and $j \geqslant 0$ be integers. If P is not divisible by p^{m+1}, then the number of solutions $(h_1, \ldots, h_r) \in (\mathbb{Z}/p^j\mathbb{Z})^r$ to the equation $P(h_1, \ldots, h_r) = 0 \bmod p^j$ is at most $r(j+1)^{d-1} p^{jr-(j-m)/rd}$.*

Proof. We first handle the $r = d = 1$ case. The claim here is that if a linear polynomial $P(\mathrm{h}) = a\mathrm{h} + b$ is not divisible by p^{m+1}, then it has at most p^m roots modulo p^j. By shifting one of the roots (if they exist at all) to the origin, we may assume without loss of generality that $b = 0$, and the claim then follows since the set $\{h \in \mathbb{Z}/p^j\mathbb{Z} : ah = 0\}$ has cardinality at most p^m if a is not divisible by p^{m+1}.

Now suppose that $r = 1$, $d > 1$, and the claim has already been proven for $r = 1$ and smaller values of d. Now we have a polynomial $P(\mathrm{h})$ of degree at most d and not divisible by p^{m+1}, and we wish to show that there are at most $(j+1)^{d-1} p^{j-(j-m)/d}$ roots in $\mathbb{Z}/p^j\mathbb{Z}$. Clearly we may assume there is at least one such root, and by translating to the origin we may factor $P(\mathrm{h}) = \mathrm{h} Q(\mathrm{h}) \bmod p^j$ for some polynomial Q of degree at most $d - 1$ and also not divisible by p^{m+1}. By factoring $h = p^a h'$ for some $0 \leqslant a \leqslant j$ and h' coprime to p, we see that the number of roots of P in $\mathbb{Z}/p^j\mathbb{Z}$ is bounded by $\sum_{a=0}^{j} N_a$, where N_a is the number of roots of the equation $Q(p^a h') = 0 \bmod p^{j-a}$ for $h' \in \mathbb{Z}/p^{j-a}\mathbb{Z}$. Observe that Q is not divisible by $p^{m+1+(d-1)a}$, so by induction hypothesis we have

$$N_a \leqslant (j+1)^{d-2} p^{j-a-(j-a-m-(d-1)a)/(d-1)} = (j+1)^{d-2} p^{j-(j-m)/(d-1)+a/(d-1)}.$$

We also have the trivial bound $N_a \leqslant p^{j-a}$. Using the former bound when $a \leqslant (j-m)/d$ and the latter when $a > (j-m)/d$, we conclude that

$$N_a \leqslant (j+1)^{d-2} p^{j-(j-m)/d}$$

and the claim then follows by summing in a.

Now suppose that $r > 1$, and that the claim has already been proven for smaller values of r. We can write $P = \sum_{i=0}^{d} P_i(\mathrm{h}_1, \ldots, \mathrm{h}_{r-1}) \mathrm{h}_r^i$ for some polynomials $P_i \in \mathbb{Z}[\mathrm{h}_1, \ldots, \mathrm{h}_{r-1}]$ of degree at most d. At least one of these polynomials, say P_{i_0}, is not divisible by p^{m+1}. Let m' denote the integer part of $j - (j-m)/r$, thus

$$0 \leqslant m' \leqslant j - \frac{j-m}{r} \leqslant m' + 1 \leqslant j$$

and in particular

$$\frac{j-m}{r} \leqslant j - m', \qquad \frac{m'+1-m}{r-1}. \tag{17}$$

By induction hypothesis, for all but at most $(r-1)(j+1)^{d-1}$ $p^{(m'+1)(r-1)-(m'+1-m)/(r-1)d}$ choices of $(h_1, \ldots, h_{r-1}) \in (\mathbb{Z}/p^{m'+1}\mathbb{Z})^{r-1}$, $P_{i_0}(h_1, \ldots, h_{r-1})$ is not divisible by $p^{m'+1}$. This implies that for all but at most $(r-1)(j+1)^{d-1} p^{j(r-1)-(m'+1-m)/(r-1)d}$ choices of $(h_1, \ldots, h_{r-1}) \in (\mathbb{Z}/p^j\mathbb{Z})^{r-1}$, $P_{i_0}(h_1, \ldots, h_{r-1})$ is not divisible by $p^{m'+1}$. But if $P_{i_0}(h_1, \ldots, h_{r-1})$ is not divisible by $p^{m'+1}$, we see from the $r = 1$ case that for all but at most $(j+1)^{d-1} p^{j-(j-m')/d}$ choices of $h_r \in \mathbb{Z}/p^j\mathbb{Z}$, $P(h_1, \ldots, h_r)$ is not divisible by p^j. Putting this all together, we conclude that $P(h_1, \ldots, h_r)$ is not divisible

by p^j for at most

$$(r-1)(j+1)^{d-1}p^{j(r-1)-(m'+1-m)/(r-1)d} \times p^j + p^{j(r-1)} \times (j+1)^{d-1}p^{j-(j-m')/d}$$

tuples $(h_1, \ldots, h_r) \in (\mathbb{Z}/p^j\mathbb{Z})^r$. By (17), this quantity is at most $r(j+1)^{d-1}p^{j-(j-m)/rd}$, giving the claim. $\square$

COROLLARY 2. *Let $P \in \mathbb{Z}[h_1, \ldots, h_r]$ be a nonzero polynomial of degree at most d for some $d \geqslant 1$, and with all coefficients of magnitude at most A for some $A \geqslant 1$. Let $q \geqslant 1$ be an integer, and let $\varepsilon > 0$. Then the number of solutions $(h_1, \ldots, h_r) \in (\mathbb{Z}/q\mathbb{Z})^r$ to the equation $P(h_1, \ldots, h_r) = 0 \mod q$ is $O_\varepsilon(A^{1/rd}q^{r-1/rd+\varepsilon})$, where we allow implied constants to depend on d, r.*

Proof. From the fundamental theorem of arithmetic, we may factor

$$q = \prod_{l=1}^{L} p_l^{j_l} \tag{18}$$

for some distinct primes $p_1, \ldots, p_L$ and some positive integers $j_1, \ldots, j_L$. For each $l = 1, \ldots, L$, let m_l be largest integer such that $p_l^{m_l}$ divides P. By the previous lemma, the number of solutions to $P(h_1, \ldots, h_r) = 0 \mod p_l^{j_l}$ is at most $r(j_l+1)^{d-1}p_l^{j_l r-(j_l-m_l)/rd}$, so by the Chinese remainder theorem the number of solutions to $P(h_1, \ldots, h_r) = 0 \mod q$ is at most

$$\prod_{l=1}^{L} r(j_l+1)^{d-1}p_l^{j_l r-(j_l-m_l)/rd}.$$

Using the divisor bound, we have

$$\prod_{l=1}^{L} r(j_l+1)^{d-1} \ll_\varepsilon q^\varepsilon,$$

and from (18) we have

$$\prod_{l=1}^{L} p_l^{j_l r-j_l/rd} = q^{r-1/rd}.$$

Finally, as the greatest common divisor of the coefficients of P is at most A, we have

$$\prod_{l=1}^{L} p_l^{m_l} \leqslant A$$

and the claim follows. $\square$

From the above corollary we see that for $q \leqslant M$, we have

$$\mathbb{E}_{\vec{h}_i \in [M]^r} 1_{q | P_i(\vec{h}_i)} \ll A^{1/rd} q^{-1/2rd}$$

and hence by (16) we have

$$c_q \leqslant q^{-k/4rd}$$

for $A^k \leqslant q \leqslant M$, if k is large. Meanwhile, from the classical Schwartz–Zippel lemma we have

$$\mathbb{E}_{\vec{h}_i \in [M]^r} 1_{p | P_i(\vec{h}_i)} \leqslant \frac{d}{M}$$

for any prime $p > M$, and thus

$$c_q \leqslant \left(\frac{d}{M} \right)^k$$

for $M < p \ll A^{O(1)} M^{d-1}$. The claim (15) then follows if k is chosen sufficiently large depending on d, r, and M sufficiently large depending on d, r, k, A. $\qquad \square$

Next, we recall a key consequence of the concatenation theory developed in [**29**].

THEOREM 8 (Qualitative Bessel inequality for box norms). *Let d be a positive integer. For each $1 \leqslant j \leqslant d$, let $(Q_{i,j})_{i \in I}$ be a finite family of progressions*

$$Q_{i,j} = a_{i,j,1}[-M_{i,j,1}, M_{i,j,1}] + \cdots + a_{i,j,r_{i,j}}[-M_{i,j,r_{i,j}}, M_{i,j,r_{i,j}}]$$

with ranks $r_{i,j}$ at most r, in a cyclic group $\mathbb{Z}/N\mathbb{Z}$. Let f lie in the unit ball of $L^\infty(\mathbb{Z}/N\mathbb{Z})$, and suppose that

$$\mathbb{E}_{i,j \in I} \| f \|_{\square^{d^2}_{(\varepsilon Q_{i,k} + \varepsilon Q_{j,l})_{1 \leqslant k \leqslant d, 1 \leqslant l \leqslant d}} (\mathbb{Z}/N\mathbb{Z})} \leqslant \varepsilon$$

for some $\varepsilon > 0$. Then

$$\mathbb{E}_{i \in I} \| f \|_{\square^d_{Q_{i,1}, \ldots, Q_{i,d}} (\mathbb{Z}/N\mathbb{Z})} \leqslant c(\varepsilon)$$

where $c \colon (0, +\infty) \to (0, +\infty)$ is a function such that $c(\varepsilon) \to 0$ as $\varepsilon \to 0$. Furthermore, c depends only on r and d.

Proof. See [**29**, Theorem 1.23]. $\qquad \square$

We combine this theorem with Theorem 7 to obtain our first result controlling an averaged Gowers norm by global Gowers norms.

THEOREM 9 (Global Gowers norms control averaged Gowers norms). *Let d, r, $D \geqslant 1$ be natural numbers, let d_0 be an integer with $1 \leqslant d_0 \leqslant d$, and suppose that $D_* \geqslant 1$ is a natural number that is sufficiently large depending on d, r, D. Let $\varepsilon > 0$, and let $\delta > 0$ be a quantity that is sufficiently small depending on d, r, D, ε. Let A be sufficiently large depending on $d, r, D, D_*, \varepsilon, \delta$. For each $j = 1, \ldots, D$, let $P_j \in \mathbb{Z}[\mathrm{h}_1, \ldots, \mathrm{h}_r]$ be a polynomial of degree between $d_0 - 1$ and $d - 1$ inclusive with coefficients that are integers of magnitude at most A. Let M be sufficiently large depending on $d, r, D, D_*, \varepsilon, \delta, A$. Let N be a quantity larger than $A^{-1} M^d$. Let $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ be a function bounded in magnitude by 1 such that*

$$\| f \|_{U^{D_*}_{q[A^{-2D_*} M^{d_0}]}} \leqslant \delta \tag{19}$$

for all $1 \leqslant q \leqslant A^{D_*}$. Then

$$\mathbb{E}_{\vec{h} \in [M]^r} \| f \|_{\square^D_{(P_i(\vec{h})[-M,M])_{i \in [D]}}} \leqslant \varepsilon. \tag{20}$$

Here the arithmetic progressions are viewed as multisets in $\mathbb{Z}/N\mathbb{Z}$.

A key point here is that δ does not depend on A.

Proof. Let k be a power of two that is sufficiently large depending on d, r, D; we assume D_* sufficiently large depending on k. Let $\sigma > 0$ be a quantity that is sufficiently small depending on d, r, k, D, ε; we assume δ sufficiently small depending on d, r, k, D, σ. We show that

$$\mathbb{E}_{\vec{h}_1, \ldots, \vec{h}_k \in [M]^r} \| f \|_{\square^{D^k}_{(\sum_{j=1}^k P_{i_j}(\vec{h}_j)[-\sigma M, \sigma M])_{i_1, \ldots, i_k \in [D]}}} \ll \sigma; \tag{21}$$

the claim will then follow from $\log_2 k$ applications of Theorem 8.

We abbreviate $\vec{i} := (i_1, \ldots, i_k)$, $\vec{h} := (\vec{h}_1, \ldots, \vec{h}_k)$, and

$$Q_{\vec{i}}(\vec{h}) := \sum_{j=1}^k P_{i_j}(\vec{h}_j)[-\sigma M, \sigma M]. \tag{22}$$

The left-hand side of (21) can then be written as

$$\mathbb{E}_{\vec{h} \in [M]^{kr}} \| f \|_{\square^{D^k}_{(Q_{\vec{i}}(\vec{h}))_{\vec{i} \in [D]^k}}}.$$

By Hölder's inequality, we may upper bound this by

$$\left(\mathbb{E}_{\vec{h} \in [M]^{kr}} \| f \|^{2^{D^k}}_{\square^{D^k}_{(Q_{\vec{i}}(\vec{h}))_{\vec{i} \in [D]^k}}} \right)^{1/2^{D^k}}.$$

It will thus suffice to show that

$$\|f\|^{2^{D^k}}_{\Box^{D^k}_{(Q_{\vec{i}}(\vec{h}))_{\vec{i}\in[D]^k}}} \ll \sigma^{2^{D^k}}$$

for all but $O(\sigma^{2^{D^k}} M^{kr})$ of the $\vec{h}$ in $[M]^{kr}$.

For each $\vec{i} \in [D]^k$, the multiset $Q_{\vec{i}}(\vec{h})$ is constructed using the polynomials $P_{i_1}, \ldots, P_{i_k}$. By the pigeonhole principle, there exists $d_0 \leqslant d_1 \leqslant d$ (depending on $\vec{i}$) such that at least k/d of the polynomials $P_{i_1}, \ldots, P_{i_k}$ have degree exactly $d_1 - 1$. Let $P_{i_{j_1}}, \ldots, P_{i_{j_{k'}}}$ denote these polynomials, then we can write

$$Q_{\vec{i}}(\vec{h}) = Q'_{\vec{i}}(\vec{h}) + Q''_{\vec{i}}(\vec{h})$$

where

$$Q'_{\vec{i}}(\vec{h}) := \sum_{l=1}^{k'} P_{i_{j_l}}(\vec{h}_{j_l})[-\sigma M, \sigma M]$$

and

$$Q''_{\vec{i}}(\vec{h}) := \sum_{j\in\{1,\ldots,k\}\setminus\{j_1,\ldots,j_{k'}\}} P_{i_j}(\vec{h}_j)[-\sigma M, \sigma M].$$

Applying Theorem 7 and Markov's inequality with these polynomials $P_{i_{j_1}}, \ldots, P_{i_{j_k}}$ (and d replaced by d_1), we conclude that for all but $O(\sigma^{2^{D^k}} M^{kr})$ of the $\vec{h}$ in $[M]^{kr}$, there exists $1 \leqslant q_{\vec{i}} \leqslant A^k$ (depending on $\vec{h}$) for each $\vec{i} \in [D]^k$ such that

$$d_{\mathrm{TV}}(Q'_{\vec{i}}(\vec{h}), Q'_{\vec{i}}(\vec{h}) + q_{\vec{i}}[-A^{-2k}M^{d_1}, A^{-2k}M^{d_1}]) \ll_{d,r,k,D,\sigma} A^{-1}.$$

Applying (4), we conclude that

$$d_{\mathrm{TV}}(Q_{\vec{i}}(\vec{h}), Q_{\vec{i}}(\vec{h}) + q_{\vec{i}}[-A^{-2k}M^{d_1}, A^{-2k}M^{d_1}]) \ll_{d,r,k,D,\sigma} A^{-1}.$$

If q is the product of all the $q_{\vec{i}}$, then from direct computation

$$d_{\mathrm{TV}}(q_{\vec{i}}[-A^{-2k}M^{d_1}, A^{-2k}M^{d_1}], q_{\vec{i}}[-A^{-2k}M^{d_1}, A^{-2k}M^{d_1}] + q[A^{-2D_*}M^{d_0}]) \leqslant A^{-1}$$

and hence by (5)

$$d_{\mathrm{TV}}(Q_{\vec{i}}(\vec{h}), Q_{\vec{i}}(\vec{h}) + q[A^{-2D_*}M^{d_0}]) \ll_{d,r,k,D,\sigma} A^{-1}.$$

From many applications of (3), we thus have

$$\|f\|^{2^{D^k}}_{\Box^{D^k}_{(Q_{\vec{i}}(\vec{h}))_{\vec{i}\in[D]^k}}} = \|f\|^{2^{D^k}}_{\Box^{D^k}_{(Q_{\vec{i}}(\vec{h})+q[A^{-2D_*}M^{d_0}])_{\vec{i}\in[D]^k}}} + O_{d,r,k,D,\sigma}(A^{-1})$$

so it will suffice to show that

$$\|f\|^{2^{D^k}}_{\square^{D^k}_{(Q_{\vec{i}}(\vec{h})+q[A^{-2D_*}M^{d_0}])_{\vec{i}\in[D]^k}}} \ll \sigma^{2^{D^k}}.$$

The left-hand side expands as

$$\mathbb{E}_{a_{0,\vec{i}},a_{1,\vec{i}}\in Q_{\vec{i}}(\vec{h})\forall\vec{i}\in[D]^k}\left\langle T^{\sum_{\vec{i}\in[D]^k}a_{\omega_{\vec{i}},\vec{i}}}f\right\rangle_{U^{D^k}(q[A^{-D_*}M^{d_0}])}$$

where $T^h f(x) := f(x + h)$ is the shift of f by h. By the Cauchy–Schwarz–Gowers inequality (6) and the translation invariance of the Gowers norms, we can bound this by

$$\|f\|_{U^{D^k}(q[A^{-2D_*}M^{d_0}])}$$

and the desired bound now follows from (19) and the monotonicity of the Gowers norms. $\qquad\square$

Theorem 9 is not directly applicable to our applications involving primes, because of the requirement that the function f is bounded in magnitude by 1. In principle, the 'transference principle' introduced in [12] should be able to relax this requirement to allow for unbounded f (so long as f is still bounded pointwise by a suitably 'pseudorandom' majorant), but this turns out to require a fair amount of additional argument. To begin this task, we present a 'dual' form of Theorem 9, which roughly speaking asserts that dual functions associated to averaged Gowers norms can be approximated by polynomial combinations of dual functions associated with global Gowers norms.

THEOREM 10 (Dual function approximation). *Let* $d, r, D \geqslant 1$ *be natural numbers; let* d_0 *be an integer with* $1 \leqslant d_0 \leqslant d$. *Suppose that* $D_* \geqslant 1$ *is a natural number that is sufficiently large depending on* d, r, D. *Let* $\varepsilon > 0$, *and let* $K > 0$ *be a quantity that is sufficiently large depending on* $d, r, D, D_*, \varepsilon$. *Let* A *be sufficiently large depending on* $d, r, D, D_*, \varepsilon, K$. *For each* $j = 1, \dots, D$, *let* $P_j \in \mathbb{Z}[\mathrm{h}_1, \dots, \mathrm{h}_r]$ *be a polynomial of degree between* $d_0 - 1$ *and* $d - 1$ *inclusive with coefficients that are integers of magnitude at most* A. *Let* M *be sufficiently large depending on* $d, r, D, D_*, \varepsilon, K, A$, *and let* M_0, N *be such that* $A^{-1}M \leqslant M_0 \leqslant AM$ *and* $N \geqslant A^{-1}M^d$. *Let* $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ *be a function bounded in magnitude by* 1, *and define the 'averaged dual function'*

$$F := \mathbb{E}_{\vec{h}\in[M_0]^r}\mathcal{D}^D_{P_1(\vec{h})[-M,M],\dots,P_D(\vec{h})[-M,M]}(f). \tag{23}$$

Then one can find a function $\tilde{F} : \mathbb{Z}/N\mathbb{Z} \to [-2, 2]$ *with*

$$\|F - \tilde{F}\|_{L^2(\mathbb{Z}/N\mathbb{Z})} \leqslant \varepsilon \tag{24}$$

such that $\tilde{F}$ is a linear combination

$$\tilde{F} = \sum_{i=1}^{k} c_i F_i \tag{25}$$

with $k \leqslant K$, and for each $i = 1, \ldots, k$, c_i is a scalar with $|c_i| \leqslant 1$, and $F_i \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ is a function of the form

$$F_i = \prod_{j=1}^{k_i} \mathcal{D}_{q_{i,j}[A^{-2D_*}M^{d_0}]}^{D_*}(f_{i,j})$$

with $k_i \leqslant K$, and for each $j = 1, \ldots, k$, $q_{i,j}$ is a natural number with $q_{i,j} \leqslant A^{D_*}$, and $f_{i,j} \colon \mathbb{Z}/N\mathbb{Z} \to [-1, 1]$ is a function bounded in magnitude by 1. As before, the progressions are viewed as multisets in $\mathbb{Z}/N\mathbb{Z}$.

Proof. Let $d, r, d_0, D, D_*, \varepsilon, K, A, P_j, N, M_0, M, f, F$ be as above. For any given K', let $\mathcal{F}_{K'}$ denote the collection of all the functions $\tilde{F}$ that have a decomposition (25) with the indicated properties, but with K replaced by K' throughout. Our task is then to show that there exists $\tilde{F} \in \mathcal{F}_K$ such that $\|F - \tilde{F}\|_{L^2(\mathbb{Z}/N\mathbb{Z})} \leqslant \varepsilon$.

Let $\delta > 0$ be a sufficiently small quantity depending on d, r, D, ε, and let $\sigma > 0$ be sufficiently small depending on $d, r, D, \varepsilon, \delta$. We prove the following *energy decrement* claim: if $K' \geqslant 0$ is a natural number, and $\tilde{F} \in \mathcal{F}_{K'}$ is such that either

$$|\langle \tilde{F}, F - \tilde{F} \rangle_{L^2(\mathbb{Z}/N\mathbb{Z})}| > \delta \tag{26}$$

or

$$\|F - \tilde{F}\|_{U_{q[A^{-2D_*}M^{d_0}]}^{D_*}} > \delta \tag{27}$$

for some $q \leqslant A^{D_*}$, then whenever K'' is sufficiently large depending on δ, σ, K', and A sufficiently large depending on $D_*, \delta, \sigma, K', K''$, there exists $\tilde{F}' \in \mathcal{F}_{K''}$ such that

$$\|F - \tilde{F}'\|_{L^2(\mathbb{Z}/N\mathbb{Z})}^2 \leqslant \|F - \tilde{F}\|_{L^2(\mathbb{Z}/N\mathbb{Z})}^2 - \sigma.$$

Applying this energy decrement claim iteratively starting from $K' = 0$ and $\tilde{F} = 0$ (which implies in particular that $\|F - \tilde{F}\|_{L^2(\mathbb{Z}/N\mathbb{Z})}^2 \leqslant 1$), we obtain after at most $1/\sigma$ iterations that for K large enough, there exists $\tilde{F} \in \mathcal{F}_K$ such that

$$|\langle \tilde{F}, F - \tilde{F} \rangle_{L^2(\mathbb{Z}/N\mathbb{Z})}| \leqslant \delta \tag{28}$$

and

$$\|F - \tilde{F}\|_{U_{q[A^{-2D_*}M^{d_0}]}^{D_*}} \leqslant \delta$$

for all $q \leqslant A^{D_*}$. Applying Theorem 9 (with f set equal to $\frac{1}{3}(F - \tilde{F})$, and adjusting ε and δ slightly), we conclude that

$$\mathbb{E}_{\vec{h} \in [M_0]^r} \|F - \tilde{F}\|_{\square^D_{P_1(\vec{h})[-M,M],\ldots,P_D(\vec{h})[-M,M]}} \leqslant \varepsilon^2/2. \tag{29}$$

We can then use (23), the Cauchy–Schwarz–Gowers inequality (6), and (29) to bound

$$\langle F, F - \tilde{F} \rangle_{L^2(\mathbb{Z}/N\mathbb{Z})} = \mathbb{E}_{\vec{h} \in [M_0]^r} \langle \mathcal{D}^D_{P_1(\vec{h})[-M,M],\ldots,P_D(\vec{h})[-M,M]}(f), F - \tilde{F} \rangle_{L^2(\mathbb{Z}/N\mathbb{Z})}$$

$$\leqslant \mathbb{E}_{\vec{h} \in [M_0]^r} \|F - \tilde{F}\|_{\square^D_{P_1(\vec{h})[-M,M],\ldots,P_D(\vec{h})[-M,M]}}$$

$$\leqslant \varepsilon^2/2$$

and hence by (28) we conclude (24).

It remains to prove the energy decrement claim. Let $K' \geqslant 0$, and let $\tilde{F} \in \mathcal{F}_{K'}$ obey either (26) or (27). First suppose that (26) holds. From the Cauchy–Schwarz inequality this implies that

$$\|\tilde{F}\|_{L^2(\mathbb{Z}/N\mathbb{Z})} \geqslant \delta/3 \tag{30}$$

(since we can bound the L^∞, and hence the L^2, norm of $F - \tilde{F}$ by 3). If we let $c\tilde{F}$ be the orthogonal projection of F to the one-dimensional space spanned by $\tilde{F}$, thus

$$c := \frac{\langle F, \tilde{F} \rangle_{L^2(\mathbb{Z}/N\mathbb{Z})}}{\|\tilde{F}\|^2_{L^2(\mathbb{Z}/N\mathbb{Z})}}.$$

From (30) we see in particular that

$$|c| \leqslant \frac{100}{\delta}. \tag{31}$$

Also, since

$$\langle \tilde{F}, F - \tilde{F} \rangle_{L^2(\mathbb{Z}/N\mathbb{Z})} = \langle \tilde{F}, c\tilde{F} - \tilde{F} \rangle_{L^2(\mathbb{Z}/N\mathbb{Z})}$$

we see that

$$|c - 1| \geqslant \frac{\delta}{10}.$$

By Pythagoras' theorem, we conclude that

$$\|F - c\tilde{F}\|^2_{L^2(\mathbb{Z}/N\mathbb{Z})} \leqslant \|F - \tilde{F}\|^2_{L^2(\mathbb{Z}/N\mathbb{Z})} - \frac{\delta^3}{100}.$$

We would like to take $c\tilde{F}$ to be the function $\tilde{F}'$, but there is the technical difficulty that $c\tilde{F}$ need not take values in $[-2, 2]$. However, from (31) we see that $c\tilde{F}$ takes

values in $[-200/\delta, 200/\delta]$. We now use the Weierstrass approximation theorem to find a polynomial P of degree $O_{\delta,\sigma}(1)$ and coefficients $O_{\delta,\sigma}(1)$ such that

$$|P(t) - \min(\max(t, -1), 1)| \leqslant \sigma^2$$

for all $t \in [-200/\delta, 200/\delta]$. Then $P(c\tilde{F})$ takes values in $[-2, 2]$ and lies in $\mathcal{F}_{K''}$ for K'' large enough depending on δ, σ, K', and

$$\|P(c\tilde{F}) - \min(\max(c\tilde{F}, -1), 1)\|_{L^\infty(\mathbb{Z}/N\mathbb{Z})} \leqslant \sigma^2$$

and thus by the triangle inequality

$$\|F - P(c\tilde{F})\|^2_{L^2(\mathbb{Z}/N\mathbb{Z})} \leqslant \|F - \min(\max(\tilde{F}, -1), 1)\|^2_{L^2(\mathbb{Z}/N\mathbb{Z})} - \sigma$$
$$\leqslant \|F - \tilde{F}\|^2_{L^2(\mathbb{Z}/N\mathbb{Z})} - \sigma$$

where the last line comes from the fact that F takes values in $[-1, 1]$ and the map $t \mapsto \min(\max(t, -1), 1)$ is a contraction. This gives the required claim in the case that (26) holds.

Now suppose that (27) holds for some $q \leqslant A^{D_*}$. If we write $g := \frac{1}{3}(F - \tilde{F})$, then g takes values in $[-1, 1]$, and

$$\|g\|_{U^{D_*}_{q[A^{-2D_*}M^{d_0}]}} > \delta/3.$$

This implies that

$$\langle g, \mathcal{D}^{D_*}_{q[A^{-2D_*}M^{d_0}]} g \rangle_{L^2(\mathbb{Z}/N\mathbb{Z})} > (\delta/3)^{2^{D_*}}$$

and thus

$$\langle F - \tilde{F}, \mathcal{D}^{D_*}_{q[A^{-2D_*}M^{d_0}]} g \rangle_{L^2(\mathbb{Z}/N\mathbb{Z})} > (\delta/3)^{2^{D_*}}/3.$$

If we then set

$$\tilde{F}_1 := \tilde{F} - \sigma^{2/3} \mathcal{D}^{D_*}_{q[A^{-2D_*}M^{d_0}]} g$$

(say) then from the cosine rule we have

$$\|F - \tilde{F}_1\|^2_{L^2(\mathbb{Z}/N\mathbb{Z})} \leqslant \|F - \tilde{F}\|^2_{L^2(\mathbb{Z}/N\mathbb{Z})} - 2\sigma$$

(say). We would like to take $\tilde{F}_1$, but again there is the slight problem that $\tilde{F}_1$ can take values a little bit outside of $[-2, 2]$. However, if one sets $\tilde{F} := P(\tilde{F}_1)$ where P is the polynomial constructed previously, one obtains the desired claim.   □

3. Averaged Gowers uniformity of a W-tricked von Mangoldt function

In this section we require the following parameters, chosen in the following order.

- First, we let $d, r, D \geqslant 1$ be arbitrary natural numbers. Let d_0 be an integer with $1 \leqslant d_0 \leqslant d$.

- Then, we let D_* be a natural number that is sufficiently large depending on d, r, D.

- Then, we choose $\kappa > 0$ to be a real number that is sufficiently small depending on d, r, D, D_*.

- We let N' be a large integer (going to infinity), and let $w = w(N')$ be a sufficiently slowly growing function of N' (the choice of w can depend on d, r, D, D_*, κ). We use $o(1)$ to denote any quantity that goes to zero as $N' \to \infty$ (keeping the previous parameters d, r, D, D_*, κ fixed). Thus for instance $1/w = o(1)$.

- We set
$$W := \prod_{p \leqslant w} p \qquad (32)$$
and $N := \lfloor N'/W \rfloor$. We also let $b \in [W]$ be a natural number coprime to W.

- We set
$$A := W^{1/\kappa}$$
and
$$R := N^\kappa. \qquad (33)$$

- We select a quantity M such that
$$\log^{1/\kappa} N \leqslant M \leqslant (AN)^{1/d}.$$

- Finally, for each $1 \leqslant j \leqslant J$, we select a polynomial $P_j \in \mathbb{Z}[\mathrm{h}_1, \ldots, \mathrm{h}_r]$ of degree between $d_0 - 1$ and $d - 1$ inclusive with coefficients that are integers of magnitude at most A.

The reader may wish to keep in mind the hierarchy
$$d, r, D \ll D_* \ll 1/\kappa \ll W \ll A \ll R \ll N$$
and
$$A, \log^{1/\kappa} N \ll M \ll N$$

where the quantities W, A, R, M, N go to infinity (at different rates) as $N \to \infty$, while d, r, D, D_*, κ do not depend on N.

For each $b \in [W]$ which is coprime to W, define the normalized von Mangoldt functions $\Lambda'_{b,W} : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ by the formula

$$\Lambda'_{b,W}(x) := \frac{\phi(W)}{W} \Lambda'(Wx + b) \tag{34}$$

for $x \in [R, N]$ (where we embed $[R, N]$ into $\mathbb{Z}/N\mathbb{Z}$), with $\Lambda'_{b,W}(x)$ equal to zero for other choices of x. Here $\Lambda'(n)$ is the restriction of the von Mangoldt function to primes, thus $\Lambda'(p) = \log p$ for primes p, and $\Lambda'(n) = 0$ for nonprime n.

The objective of this section is to establish the following bound:

THEOREM 11 (Averaged Gowers uniformity of von Mangoldt). *Let the notation and hypotheses be as above. If $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ is any function with the pointwise bound $0 \leqslant f \leqslant \Lambda'_{b,W}$, then there exists a function $f' : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ with the pointwise bounds $0 \leqslant f' \ll 1$ such that*

$$\mathbb{E}_{\vec{h} \in [M]^r} \| f - f' \|_{\square^D_{(P_i(\vec{h})[-M,M])_{i \in [D]}}} = o(1). \tag{35}$$

Furthermore, in the special case where $d_0 = d$, $M \geqslant (A^{-1} N)^{1/d}$, and $f = \Lambda'_{b,W}$, we may take $f' = 1$.

The approximating function f' is known as a *dense model* for f in the literature. The above theorem can be compared with the bound $\| f - f' \|_{U^D(\mathbb{Z}/N\mathbb{Z})} = o(1)$ established in [12], and the results in [13] which imply that f' can be taken to equal 1 when $f = \Lambda'_{b,W}$. Also, the arguments in [27] give a version of the first part of this theorem in the case that M is a very small power of N (in particular, much smaller than R), and the subsequent arguments in [28] extend this to cover the regime where M grows slower than any power of N but faster than any power of $\log N$. These restrictions on M arose from a certain 'clearing denominators' step encountered when dealing with products of many dual functions associated to averaged local Gowers norms; they are circumvented in this paper by application of the concatenation machinery to replace these norms with more traditional Gowers norms that do not require the 'clearing denominators' method in order to handle products of dual functions.

REMARK 1. In [28, Theorem 9], in the notation of the current paper, the slightly weaker bound

$$\mathbb{E}_{\vec{h} \in [M]^r} \| f - f' \|_{\square^D_{P_1(\vec{h})[-M,M],\dots,P_D(\vec{h})[-M,M]}} \leqslant \varepsilon + o(1)$$

was established for any fixed $\varepsilon > 0$ assuming that $M = \log^L N$ for some $L = L(\varepsilon)$ that was sufficiently large depending on ε. Theorem 11 allows one to now take $L = 1/\kappa$ independent of ε. By [28, Remark 4], this allows one to also take L independent of ε in [28, Theorem 5], which gives Theorem 6.

Morally speaking, Theorem 11 ought to follow from the results in [12, 13] mentioned above after applying Theorem 9, but we run into the familiar difficulty that the function $\Lambda'_{b,W} - 1$ is not bounded. In the final part of Theorem 11, the conditions on d_0 and M should be dropped, but this requires control on correlations of Λ with nilsequences on short intervals, and such control is not currently available in the literature.

We now begin the proof of Theorem 11. As in [12, 27], we envelop $\Lambda'_{b,W}$ (and hence f) by a pseudorandom majorant $\nu = \nu_b : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}^+$, defined as follows. Let $\chi : \mathbb{R} \to \mathbb{R}$ be a fixed smooth even function that vanishes outside of $[-1, 1]$, positive at 0, and obeys the normalization

$$\int_0^1 |\chi'(t)|^2 \, dt = 1.$$

We allow all implied constants to depend on χ. We then set

$$\nu(x) = \nu_b(x) := \frac{\phi(W)}{W} \log R \left(\sum_{m | Wx+b} \mu(m) \chi\left(\frac{\log m}{\log R}\right) \right)^2 \tag{36}$$

for all $x \in [N]$. Comparing this with (34) and (33) we conclude the pointwise bound

$$0 \leqslant f(x) \leqslant \Lambda'_{b,W}(x) \ll_\kappa \nu(x) \tag{37}$$

for all $x \in \mathbb{Z}/N\mathbb{Z}$, since $\Lambda'_{b,W}(x)$ is only nonvanishing when $x \in [R, N]$ and $Wx + b$ is prime, in which case the only nonzero summand in (36) comes from the $m = 1$ term.

It was observed in [12] that

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \nu(x) = 1 + o(1). \tag{38}$$

In fact we have many further 'pseudorandomness' properties of ν; roughly speaking, any 'nondegenerate' multilinear correlation of ν with itself (with reasonable bounds on coefficients) will be $1 + o(1)$, as long as the complexity of the correlation is small compared to $1/\kappa$. Here is one specific instance of the principle we need:

PROPOSITION 2 (Gowers uniformity of $\nu - 1$). *Let K be a natural number independent of N (but which can depend on d, r, D, D_*, κ). Let $q \leqslant A^{KD_*}$ be a natural number. Then we have*

$$\mathbb{E}_{\vec{h},\vec{h}' \in [M]^r} \|\nu - 1\|^{2^{D_*+2D}}_{\Box^{D_*+2D}_{q[A^{-3KD_*}M^{d_0}],\ldots,q[A^{-3KD_*}M^{d_0}],(P_i(\vec{h})[-M,M])_{i\in[D]},(P_i(\vec{h}')[-M,M])_{i\in[D]}}} = o(1). \quad (39)$$

In particular, by monotonicity (7), *permutation symmetry, and Hölder's inequality one has the estimates*

$$\|\nu - 1\|_{\Box^{D_*}_{q[A^{-3KD_*}M^{d_0}],\ldots,q[A^{-3KD_*}M^{d_0}]}} = o(1),$$

$$\mathbb{E}_{\vec{h}\in[M]^r} \|\nu - 1\|^{2^{D_*+D}}_{\Box^{D_*+D}_{q[A^{-3KD_*}M^{d_0}],\ldots,q[A^{-3KD_*}M^{d_0}],(P_i(\vec{h})[-M,M])_{i\in[D]}}} = o(1),$$

and

$$\mathbb{E}_{\vec{h},\vec{h}' \in [M]^r} \|\nu - 1\|^{2^{2D}}_{\Box^{2D}_{(P_i(\vec{h})[-M,M])_{i\in[D]},(P_i(\vec{h}')[-M,M])_{i\in[D]}}} = o(1).$$

Proof. See Appendix A. $\qquad \Box$

Next, we give a variant of the estimate in [**12**, Proposition 6.2].

PROPOSITION 3 ($\nu - 1$ orthogonal to dual functions). *Let the notation and hypotheses be as in Theorem* 11. *Let K, K' be natural numbers independent of N (but which can depend on d, r, D, D_*, κ). For each $j = 1, \ldots, K$, let $\vec{f}_j = (f_{j,\omega})_{\omega\in\{0,1\}^{D_*}\setminus\{0\}^{D_*}}$ be a tuple of functions $f_{j,\omega}: \mathbb{Z}/N\mathbb{Z} \to [-1,1]$, and $q_j \leqslant A^{D_*}$ be a natural number, and write*

$$F_j := \mathcal{D}^{D_*}_{q_j[A^{-2D_*}M^{d_0}]}(f_j).$$

For each $j' \in [K']$, let $\vec{g}_{j'} = (g_{j',\omega})_{\omega\in\{0,1\}^D\setminus\{0\}^D}$ be a tuple of functions $g_{j',\omega}: \mathbb{Z}/N\mathbb{Z} \to [-1,1]$, and write

$$G_{j'} := \mathbb{E}_{\vec{h}\in[M]^r} \mathcal{D}^D_{(P_i(\vec{h})[-M,M])_{i\in[D]}}(\vec{g}_{j'}).$$

Then

$$\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}(\nu(x) - 1)\left(\prod_{j=1}^K F_j(x)\right)\left(\prod_{j'=1}^{K'} G_{j'}(x)\right) = o(1). \quad (40)$$

A key technical point here (as in [**12**]) is that the parameters K, K' are allowed to be large compared to $1/\kappa$ (though K, K' will still be small compared to W, A, R, M, N). The potentially large nature of K, K' requires one to proceed

carefully, as the pseudorandomness properties of ν do not allow us to directly control averages involving a number of forms that are comparable or larger to $1/\kappa$. On the other hand, Proposition 3 is easier to prove in one respect than [**12**, Proposition 6.2], because the functions $f_{j,\omega}$, $g_{j',\omega}$ are assumed to be bounded in magnitude by 1 rather than $\nu + 1$. We are able to use this stronger hypothesis due to our use of the Conlon–Fox–Zhao densification machinery (which was not available at the time that [**12**] was written) from [**4**] later in this paper.

Proof. We first prove (40) for small values of K', specifically $K' = 0, 1, 2$, and then use Theorem 10 to conclude the case of general K'.

We begin with the $K' = 0$ case. In this case, we may adapt the arguments used to prove [**12**, Proposition 6.2]. First, we write the left-hand side of (40) as

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\nu(x) - 1) \prod_{j=1}^{K} \prod_{\omega \in \{0,1\}^{D_*} \setminus \{0\}^{D_*}} \mathbb{E}_{\vec{h}_j \in Q_j^{D_*}} f_{j,\omega}(x + \omega \cdot \vec{h}_j)$$

where Q_j is the multiset

$$Q_j := q_j[A^{-2D_*}M^{d_0}] - q_j[A^{-2D_*}M^{d_0}]$$

and $\cdot$ denotes the usual dot product:

$$(\omega_1, \ldots, \omega_{D_*}) \cdot (h_{j,1}, \ldots, h_{j,D_*}) := \sum_{i=1}^{D_*} \omega_i h_{j,i}.$$

We now 'clear denominators' by writing $q := \prod_{j=1}^{K} q_j$, and introducing the multiset

$$Q := q[A^{-3KD_*}M^{d_0}] - q[A^{-3KD_*}M^{d_0}].$$

Note that $q \leqslant A^{KD_*}$, which implies that

$$d_{\mathrm{TV}}(Q_j, Q_j + h) = o(1)$$

for all $h \in Q$, thus

$$\mathbb{E}_{\vec{h}_j \in Q_j^{D_*}} f_{j,\omega}(x + \omega \cdot \vec{h}_j) = \mathbb{E}_{\vec{H}_j \in Q_j^{D_*}} f_{j,\omega}(x + \omega \cdot \vec{H}_j + \omega \cdot \vec{h}) + o(1)$$

for all $1 \leqslant j \leqslant K$ and $\vec{h} \in Q^{D_*}$. From (38) we have $\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}\nu(x) + 1 = 2 + o(1)$, so we can write the left-hand side of (40) as

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\nu(x) - 1) \prod_{j=1}^{K} \prod_{\omega \in \{0,1\}^{D_*} \setminus \{0\}^{D_*}} \mathbb{E}_{\vec{H}_j \in Q_j^{D_*}} f_{j,\omega}(x + \omega \cdot \vec{H}_j + \omega \cdot \vec{h}) + o(1);$$

averaging over $\vec{h}$, we can also write this left-hand side as

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{\vec{h} \in Q^{D_*}} (\nu(x) - 1) \prod_{j=1}^{K} \prod_{\omega \in \{0,1\}^{D_*} \setminus \{0\}^{D_*}} \mathbb{E}_{\vec{H}_j \in Q_j^{D_*}} f_{j,\omega}(x + \omega \cdot \vec{H}_j + \omega \cdot \vec{h}) + o(1).$$

We can rewrite this as

$$\mathbb{E}_{\vec{H} \in \prod_{j=1}^{K} Q_j^{D_*}} \langle (f_{\omega;\vec{H}})_{\omega \in \{0,1\}^{D_*}} \rangle_{U_{q[A^{-3KD_*}M^{d_0}]}^{D_*}} + o(1)$$

where

$$f_{\{0\}^{D_*};\vec{H}}(x) := \nu(x) - 1 \tag{41}$$

and

$$f_{\omega;\vec{H}}(x) := \prod_{j=1}^{K} f_{j,\omega}(x + \omega \cdot \vec{H}_j) \tag{42}$$

for $\omega \in \{0,1\}^{D_*} \setminus \{0\}^{D_*}$ and $\vec{H} = (\vec{H}_1, \ldots, \vec{H}_k)$. By the Cauchy–Schwarz–Gowers inequality (6), we can bound this by

$$\mathbb{E}_{\vec{H} \in \prod_{j=1}^{K} Q_j^{D_*}} \prod_{\omega \in \{0,1\}^{D_*}} \| f_{\omega;\vec{H}} \|_{U_{q[A^{-3KD_*}M^{d_0}]}^{D_*}} + o(1).$$

From Proposition 2 we have

$$\| f_{\{0\}^{D_*};\vec{H}} \|_{U_{q[A^{-3KD_*}M^{d_0}]}^{D_*}} = \| \nu - 1 \|_{U_{q[A^{-3KD_*}M^{d_0}]}^{D_*}} = o(1)$$

while for $\omega \in \{0,1\}^{D_*} \setminus \{0\}^{D_*}$ we have $|f_{\omega;\vec{H}}| \leqslant 1$ and hence

$$\| f_{\omega;\vec{H}} \|_{U_{q[A^{-3KD_*}M^{d_0}]}^{D_*}} \leqslant 1.$$

The claim follows.

Now we turn to the $K' = 1$ case. This is effectively the same as the $K' = 0$ case, except that $\nu - 1$ is replaced by $(\nu - 1)G_1$. Thus, by repeating the above arguments, we reduce to showing that

$$\| (\nu - 1)G_1 \|_{U_{q[A^{-3KD_*}M^{d_0}]}^{D_*}} = o(1).$$

Expanding out G_1 and using the triangle inequality for Gowers norms, it suffices to show that

$$\mathbb{E}_{\vec{h} \in [M]^r} \| (\nu - 1) \mathcal{D}_{(P_i(\vec{h})[-M,M])_{i \in [D]}}^{D} (\vec{g}_1) \|_{U_{q[A^{-3KD_*}M^{d_0}]}^{D_*}} = o(1)$$

so by Hölder's inequality it suffices to show that

$$\mathbb{E}_{\vec{h}\in[M]^r}\|(\nu-1)\mathcal{D}^D_{(P_i(\vec{h})[-M,M])_{i\in[D]}}(\vec{g}_1)\|^{2^{D_*}}_{U^{D_*}_{q[A^{-3KD_*}M^{d_0}]}} = o(1).$$

We can expand the left-hand side as

$$\mathbb{E}_{\vec{h}\in[M]^r}\langle g_{\omega,\vec{h}}\rangle_{\square^{D_*+D}_{q[A^{-3KD_*}M^{d_0}],\dots,q[A^{-3KD_*}M^{d_0}],(P_i(\vec{h})[-M,M])_{i\in[D]}}}$$

where $q[A^{-3KD_*}M^{d_0}]$ appears D_* times in the Gowers norm, and

$$g_{\omega,\{0\}^D} := \nu - 1$$

and

$$g_{\omega,\omega'} := g_{1,\omega'}$$

for all $\omega \in \{0,1\}^{D_*}$ and $\omega' \in \{0,1\}^D\setminus\{0\}^D$. By the Gowers–Cauchy–Schwarz inequality, we can bound the above expression by

$$\mathbb{E}_{\vec{h}\in[M]^r}\|\nu-1\|^{2^{D_*}}_{\square^{D_*+D}_{q[A^{-3KD_*}M^{d_0}],\dots,q[A^{-3KD_*}M^{d_0}],(P_i(\vec{h})[-M,M])_{i\in D}}}$$

so by Hölder's inequality it suffices to show that

$$\mathbb{E}_{\vec{h}\in[M]^r}\|\nu-1\|^{2^{D_*+D}}_{\square^{D_*+D}_{q[A^{-3KD_*}M^{d_0}],\dots,q[A^{-3KD_*}M^{d_0}],(P_i(\vec{h})[-M,M])_{i\in[D]}}} = o(1).$$

But this follows from Proposition 2.

Next, we turn to the $K'=2$ case. Repeating the previous arguments, we reduce to showing that

$$\|(\nu-1)G_1G_2\|_{U^{D_*}_{q[A^{-3KD_*}M^{d_0}]}} = o(1).$$

But the product G_1G_2 can be written as a dual function

$$G_1G_2 = \mathcal{D}^{2D}_{(P_i(\vec{h})[-M,M])_{i\in[D]},(P_i(\vec{h})[-M,M])_{i\in[D]}}(\vec{g}_{12})$$

where $\vec{g}_{12} = (g_{12,\omega})_{\omega\in\{0,1\}^{2D}\setminus\{0\}^{2D}}$ is defined by setting

$$g_{12,(\omega,\{0\}^D)} := g_{1,\omega}$$

and

$$g_{12,(\{0\}^D,\omega')} := g_{2,\omega'}$$

and

$$g_{12,(\omega,\omega')} := 1$$

for $\omega, \omega' \in \{0, 1\}^D \backslash \{0\}^D$. One can then repeat the $K' = 1$ arguments (replacing D with $2D$, and duplicating the polynomials $P_1, \ldots, P_D$) to conclude this case.

Now we turn to the general case when K' is allowed to be large. Let $\varepsilon > 0$ be a parameter (which we initially take to be independent of N) to be chosen later. It will suffice to show that the left-hand side of (40) is $O_{K'}(\varepsilon) + o(1)$ for each fixed $\varepsilon > 0$, since the claim then follows by using a diagonalization argument to send ε slowly to zero.

Applying Theorem 10 to each $G_{j'}$, we see that for any $j' = 1, \ldots, K'$, one can find an approximation $\tilde{G}_{j'} : \mathbb{Z}/N\mathbb{Z} \to [-2, 2]$ to $G_{j'}$ with

$$\|G_{j'} - \tilde{G}_{j'}\|_{L^2(\mathbb{Z}/N\mathbb{Z})} \leqslant \varepsilon$$

such that $\tilde{G}_{j'}$ has a representation of the form (25) for some $k, k_i = O_{d,r,D,D_*,\varepsilon}(1)$ (depending on j). We then write

$$G_{j'} = \tilde{G}_{j'} + E_{j'}$$

and

$$G_1 \ldots G_{K'} = \tilde{G}_1 \ldots \tilde{G}_{K'} + E$$

for some error functions $E_{j'}, E$ with

$$\|E_{j'}\|_{L^2(\mathbb{Z}/N\mathbb{Z})} \leqslant \varepsilon \tag{43}$$

and

$$E \ll_{K'} |E_1| + \cdots + |E_{K'}|. \tag{44}$$

We may split the left-hand side of (40) as the sum of

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\nu(x) - 1)F_1 \ldots F_K(x)\tilde{G}_1 \ldots \tilde{G}_{K'}(x) \tag{45}$$

and

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\nu(x) - 1)F_1 \ldots F_K(x)E(x). \tag{46}$$

By (25), the expression (45) is a bounded linear combination of $O_{d,r,D,D_*,\varepsilon,K}(1)$ terms, each of which is $o(1)$ by the $K' = 0$ case of this proposition (with K replaced by various quantities of size $O_{d,r,D,D_*,\varepsilon,K}(1)$). Thus it suffices to show that the expression (46) is $O_{K'}(\varepsilon) + o(1)$. Since the F_j are bounded in magnitude by 1, we can use (44) to bound (46) in magnitude by

$$\ll_{K'} \sum_{j'=1}^{K'} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\nu(x) + 1)|E_{j'}(x)|.$$

By (38) we have $\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\nu(x)+1) = 2+o(1)$, so by Cauchy–Schwarz it suffices to show that

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\nu(x)+1)E_{j'}(x)^2 \ll_{K'} \varepsilon^2 + o(1)$$

for each j'. From (43) it suffices to show that

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\nu(x)-1)E_{j'}(x)^2 = o(1).$$

From the definition of E, we can expand the left-hand side as a bounded linear combination of the expressions

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\nu(x)-1)\tilde{G}_{j'}(x)\tilde{G}_{j'}(x),$$

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\nu(x)-1)\tilde{G}_{j'}(x)G_{j'}(x),$$

and

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\nu(x)-1)G_{j'}(x)G_{j'}(x).$$

But these are all equal to $o(1)$ thanks to the $K' = 0, 1, 2$ cases of this proposition respectively. The claim follows. $\qquad\square$

To use this proposition, we recall the *dense model theorem*:

THEOREM 12 (Dense model theorem). *Let $\nu \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}^+$ be a function satisfying $\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}\nu(x) = 1 + o(1)$, and let $\mathcal{F}$ be a collection of functions $g \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}^+$ bounded in magnitude by $1 + o(1)$. Suppose that for any fixed k (independent of N) and any $g_1, \ldots, g_k \in \mathcal{F}$, one has*

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}g_1(x)\ldots g_k(x)(\nu(x)-1) = o(1)$$

as $N \to \infty$ uniformly in the choice of $g_1, \ldots, g_k$. Then for any function $f \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ with $0 \leqslant f(x) \leqslant \nu(x)$ for all x, there exists a function $f' \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ with $0 \leqslant f'(x) \leqslant 1 + o(1)$ for all x such that

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(f(x)-f'(x))g(x) = o(1)$$

as $N \to \infty$ uniformly for all $g \in \mathcal{F}$.

Proof. See [**24**, Theorem 1.1], [**27**, Theorem 7.1], or [**10**, Theorem 4.8]; the formulation here is closest to that in [**24**]. (This theorem also appeared implicitly in [**12**].) $\qquad\square$

PROPOSITION 4 (A dense model for f). *Let the notation and hypotheses be as in Theorem 11. Then there exists a function $f' \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ with the pointwise bounds $0 \leqslant f' \ll_\kappa 1$ such that*

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(f-f')(x)\mathcal{D}^{D_*}_{q[A^{-2D_*}M^{d_0}]}(\vec{g})(x) = o(1)$$

and

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(f - f')(x)\mathbb{E}_{\vec{h} \in [M]^r}\mathcal{D}^D_{(P_i(\vec{h})[-M,M])_{i \in [D]}}(\vec{g}')(x) = o(1)$$

for any $q \leqslant A^{D_*}$ and any tuples $\vec{g} = (g_\omega)_{\omega \in \{0,1\}^{D_*}\setminus\{0\}^{D_*}}$ and $\vec{g}' = (g'_\omega)_{\omega \in \{0,1\}^D\setminus\{0\}^D}$ of functions $g_\omega, g'_\omega \colon \mathbb{Z}/N\mathbb{Z} \to [-1, 1]$.

Proof. If we let $c > 0$ be a sufficiently small quantity depending on κ, then by (37) we can bound cf pointwise in magnitude by ν. Applying the dense model theorem with $\mathcal{F}$ consisting of those functions of the form

$$\mathcal{D}^{D_*}_{q[A^{-2D_*}M^{d_0}]}(\vec{g})$$

or

$$\mathbb{E}_{\vec{h} \in [M]^r}\mathcal{D}^D_{(P_i(\vec{h})[-M,M])_{i \in [D]}}(\vec{g}')$$

for any $q \leqslant A^{D_*}$ and any tuples $\vec{g} = (g_\omega)_{\omega \in \{0,1\}^{D_*}\setminus\{0\}^{D_*}}$ and $\vec{g}' = (g'_\omega)_{\omega \in \{0,1\}^D\setminus\{0\}^D}$ of functions $g_\omega, g'_\omega \colon \mathbb{Z}/N\mathbb{Z} \to [-1, 1]$, and using Proposition 3, we obtain the claim with f replaced by cf. Dividing by c, the proposition follows. $\qquad\square$

The above properties of $f - f'$ are not directly useful in applications because of the requirement that the functions g_ω, g'_ω take values in $[-1, 1]$. However, we can use *densification* argument of Conlon *et al.* [4] to relax this hypothesis, to a hypothesis that each g_ω or g'_ω is bounded by either 1 or ν:

PROPOSITION 5 (Densification). *Let the notation and hypotheses be as in Theorem* 11. *Let f' be the function in Proposition* 4. *Then one has*

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(f - f')(x)\mathcal{D}^{D_*}_{q[A^{-2D_*}M^{d_0}]}(\vec{g})(x) = o(1) \tag{47}$$

and

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(f - f')(x)\mathbb{E}_{\vec{h} \in [M]^r}\mathcal{D}^D_{(P_i(\vec{h})[-M,M])_{i \in [D]}}(\vec{g}')(x) = o(1) \tag{48}$$

for any $q \leqslant A^{D_*}$ and any tuples $\vec{g} = (g_\omega)_{\omega \in \{0,1\}^{D_*}\setminus\{0\}^{D_*}}$ and $\vec{g}' = (g'_\omega)_{\omega \in \{0,1\}^D\setminus\{0\}^D}$ of functions $g_\omega, g'_\omega \colon \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$, each of which are bounded in magnitude either by 1 or by ν. (Thus, for instance, for $\omega \in \{0,1\}^{D_*}$, one either has $|g_\omega(x)| \leqslant 1$ for all $x \in \mathbb{Z}/N\mathbb{Z}$, or $|g_\omega(x)| \leqslant \nu(x)$ for all $x \in \mathbb{Z}/N\mathbb{Z}$.)

Proof. We just establish the claim (48), as the claim (47) is proven by an essentially identical argument.

Let J denote the number of indices $\omega' \in \{0,1\}^D\setminus\{0\}^D$ for which one is assuming the bound $|g'_\omega(x)| \leqslant \nu(x)$ (rather than $|g'_\omega(x)| \leqslant 1$), thus $0 \leqslant J \leqslant 2^D - 1$. We prove the claim by induction on J. For $J = 0$, the claim is immediate

from Proposition 4. Now suppose inductively that $J \geqslant 1$, and that the claim has already been proven for $J - 1$.

By hypothesis, we have a tuple ω_0 in $\{0, 1\}^D \setminus \{0\}^D$ such that g'_{ω_0} is bounded in magnitude by ν, with g_ω being bounded by ν for $J - 1$ of the other indices $\{0, 1\}^D \setminus \{0\}^D$ and by 1 for all remaining indices. We also adopt the convention $g'_{\{0\}^D} := f - f'$, so that we may write the left-hand side of (48) as

$$\mathbb{E}_{\vec{h} \in [M]^r} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{\vec{k} \in \prod_{i=1}^D P_i(\vec{h})[-M,M]} \prod_{\omega \in \{0,1\}^D} g'_\omega(x + \omega \cdot \vec{k}).$$

Making the change of variables $x \mapsto x - \omega_0 \cdot \vec{k}$, we may write this as

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} g'_{\omega_0}(x) F(x)$$

where

$$F(x) := \mathbb{E}_{\vec{h} \in [M]^r} \mathbb{E}_{\vec{k} \in \prod_{i=1}^D P_i(\vec{h})[-M,M]} \prod_{\omega \in \{0,1\}^D \setminus \{\omega_0\}} g'_\omega(x + (\omega - \omega_0) \cdot \vec{k}).$$

Since g'_{ω_0} is bounded in magnitude by ν, we see from (38) and Cauchy–Schwarz that it suffices to show that

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \nu(x) F(x)^2 = o(1).$$

We split this into two claims

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} (\nu(x) - 1) F(x)^2 = o(1) \tag{49}$$

and

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} F(x)^2 = o(1). \tag{50}$$

We begin with (49). We can expand the left-hand side as

$$\mathbb{E}_{\vec{h}, \vec{h}' \in [M]^r} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{\vec{k} \in \prod_{i=1}^D P_i(\vec{h})[-M,M] \times \prod_{i=1}^D P_i(\vec{h}')[-M,M]} (\nu(x) - 1)$$
$$\times \prod_{\omega \in \{0,1\}^D \setminus \{\omega_0\}} g'_\omega(x + (\omega - \omega_0, 0) \cdot \vec{k}) g'_\omega(x + (0, \omega - \omega_0) \cdot \vec{k}).$$

By the Gowers–Cauchy–Schwarz inequality, we can bound this by

$$\mathbb{E}_{\vec{h}, \vec{h}' \in [M]^r} \|\nu(x) - 1\|_{\square^{2D}_{(P_i(\vec{h})[-M,M])_{i \in [D]}, (P_i(\vec{h}')[-M,M])_{i \in [D]}}}$$
$$\times \prod_{\omega \in \{0,1\}^D \setminus \{\omega_0\}} \|g'_\omega\|_{\square^{2D}_{(P_i(\vec{h})[-M,M])_{i \in [D]}, (P_i(\vec{h}')[-M,M])_{i \in [D]}}}.$$

Since g'_ω is bounded in magnitude by $\nu + 1 = (\nu - 1) + 2$, it thus suffices by Hölder's inequality to show that

$$\mathbb{E}_{\vec{h},\vec{h}' \in [M]^r} \|\nu - 1\|_{\square^{2D}_{(P_i(\vec{h})[-M,M])_{i \in [D]},(P_i(\vec{h}')[-M,M])_{i \in [D]}}}^{2D} = o(1).$$

But this follows from Proposition 2.

Now we show (50). By Hölder's inequality, it suffices to show that

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} F(x)^4 \ll_{r,D} 1 + o(1) \tag{51}$$

and

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} |F(x)| = o(1). \tag{52}$$

To prove (51), we bound g'_ω by $\nu + 1$. The estimate then follows from the following claim, proven in Appendix A:

LEMMA 3. *We have*

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \left(\mathbb{E}_{\vec{h} \in [M]^r} \mathbb{E}_{\vec{k} \in \prod_{i=1}^{D} P_i(\vec{h})[-M,M]} \prod_{\omega \in \{0,1\}^D \setminus \{\omega_0\}} (1 + \nu(x + (\omega - \omega_0) \cdot \vec{k})) \right)^4$$
$$\ll_{r,D} 1 + o(1).$$

Finally, we show (52). It suffices to show that

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} g(x) F(x) = o(1)$$

whenever $g \colon \mathbb{Z}/N\mathbb{Z} \to [-1, 1]$ is a function. But this follows from the induction hypothesis, since the left-hand side is simply (48) with g'_{ω_0} replaced with g. This proves (48); the estimate (47) is proven similarly and is left to the reader. $\square$

COROLLARY 3. *Let the notation and hypotheses be as in Theorem* 11. *Let f' be the function in Proposition* 4. *Then one has*

$$\|f - f'\|_{U^{D_*}_{q[A^{-2D_*}M^{d_0}]}} = o(1) \tag{53}$$

for all $q \leqslant A^{D_*}$, and also

$$\mathbb{E}_{\vec{h} \in [M]^r} \|f - f'\|_{\square^{D}_{(P_i(\vec{h})[-M,M])_{i \in [D]}}} = o(1). \tag{54}$$

Proof. By linearity, the bounds (47), (48) continue to hold if the hypotheses that $|g_\omega|, |g'_\omega|$ are bounded by 1 or by ν are replaced by $|g_\omega|, |g'_\omega| \ll_\kappa \nu + 1$.

In particular, we can set all of the g_ω, g'_ω equal to $f - f'$. The bound (47) then gives (53), while (48) gives

$$\mathbb{E}_{\vec{h} \in [M]^r} \| f - f' \|^{2^D}_{\square^D_{(P_i(\vec{h})[-M,M])_{i \in [D]}}} = o(1)$$

which implies (54) by Hölder's inequality. $\qquad\square$

The bound (54) already gives the first claim of Theorem 11. It remains to prove the second claim. Thus we may now assume that $d_0 = d$, $M \geqslant (A^{-1}N)^{1/d}$, and $f = \Lambda'_{b,W}$.

We now invoke the following estimates on global uniformity norms of the von Mangoldt function, arising from the results in [13] (using the inverse Gowers and Möbius–nilsequences conjectures proven in [16], [14], respectively):

PROPOSITION 6 (Global Gowers uniformity). *Let the notation and hypotheses be as in Theorem 11, with $d_0 = d$, $M \geqslant (A^{-1}N)^{1/d}$, and $f = \Lambda'_{b,W}$. Then one has*

$$\| \Lambda'_{b,W} - 1 \|_{U^{D_*}_{q[A^{-2D_*}M^{d_0}]}} = o(1)$$

for all $q \leqslant A^{D_*}$.

Proof. Raising both sides to the power 2^{D_*} and expanding, it suffices to show that

$$\mathbb{E}_{\vec{h} \in [A^{-2D_*}M^{d_0}]^{D_*}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \prod_{\omega \in I} \Lambda'_{b,W}(x + q\omega \cdot \vec{h}) = 1 + o(1)$$

for all $I \subset \{0,1\}^{D_*}$. The contribution of those x for which $N - qD_*A^{-2D_*}M^{d_0} \leqslant x \leqslant N$ can be easily verified by standard upper bound sieves to be $o(1)$; replacing Λ' in (34) by the von Mangoldt function Λ and removing the restriction $x \geqslant R$ also contributes an error of $o(1)$. Thus it suffices to show that

$$\mathbb{E}_{\vec{h} \in [A^{-2D_*}M^{d_0}]^{D_*}} \mathbb{E}_{x \in [N]} \prod_{\omega \in I} \frac{\phi(W)}{W} \Lambda(W(x + q\omega \cdot \vec{h}) + b) = 1 + o(1). \qquad (55)$$

Suppose first that w (and hence W and A) is a fixed quantity independent of N. Since

$$A^{-2D_*-1}N \leqslant A^{-2D_*}M^{d_0} \leqslant A^{-2D_*+1}N$$

we see that the quantity $A^{-2D_*}M^{d_0}$ is now comparable to N. The expression (55) is now of a form that can be handled by the results in [13, Theorem 1.8] (note that a modification of Example 2 from that paper shows that the linear forms here have finite complexity). Using that theorem (as well as the results in

[**16**, Theorem 1.3], [**14**, Theorem 1.1] establishing the conditional hypotheses of that theorem), we see that the left-hand side of (55) is equal to

$$(1 + O(1/A))\left(\frac{\phi(W)}{W}\right)^{|I|}\prod_p \beta_p + o(1)$$

where for each prime p, β_p is the quantity

$$\beta_p := \mathbb{E}_{\vec{h}\in(\mathbb{Z}/p\mathbb{Z})^{D_*}}\mathbb{E}_{x\in\mathbb{Z}/p\mathbb{Z}}\prod_{\omega\in I}\frac{p}{p-1}1_{p\nmid W(x+q\omega\cdot\vec{h})+b}.$$

For $p \leqslant w$, the quantity β_p simplifies to $(p/(p-1))^{|I|}$, and hence the left-hand side of (55) simplifies to

$$(1 + O(1/A))\prod_{p>w}\beta_p + o(1).$$

For $p > w$ not dividing q, the linear forms $(x, \vec{h}) \mapsto W(x + q\omega \cdot \vec{h})$ are not scalar multiples of each other over $\mathbb{Z}/p\mathbb{Z}$, and one can then easily verify that $\beta_p = 1 + O_{D_*}(1/p^2) = \exp(O_{D_*}(1/p^2))$ in these cases, leading to a net multiplicative contribution of $\exp(O_{D_*}(1/w))$ to the product $\prod_{p>w}\beta_p$. For $p > w$ dividing q, we can crudely estimate β_p as $1 + O_{D_*}(1/p) = \exp(O_{D_*}(1/w))$; since $q \leqslant A^{D_*} = W^{D_*/\kappa} = \exp(O_{D_*,\kappa}(w))$, the number of such primes p is at most $O_{D_*,\kappa}(w/\log w)$. We conclude that the net contribution of these primes to $\prod_{p>w}\beta_p$ is $\exp(O_{D_*,\kappa}(1/\log w))$. We conclude that the left-hand side of (55) is

$$(1 + O(1/A))\exp(O_{D_*,\kappa}(1/\log w)) + o(1)$$

for fixed w. Letting w grow sufficiently slowly to infinity, we obtain the claim. $\qquad\square$

Combining Proposition 6 with Corollary 3 and Theorem 9, we can now finally establish the second part of Theorem 11. Let f' be the function in Proposition 4. From (53), Proposition 6, and the triangle inequality we have

$$\|f' - 1\|_{U^{D_*}_{q[A^{-2D_*}M^{d_0}]}} = o(1)$$

for all $q \leqslant A^{D_*}$. Since $f' - 1 = O_\kappa(1)$, we can divide by a constant depending only on κ and invoke Theorem 9 to conclude that

$$\mathbb{E}_{\vec{h}\in[M]^r}\|f' - 1\|_{\square^D_{(P_i(\vec{h})[-M,M])_{i\in[D]}}} = o(1).$$

Combining this with (54) and the triangle inequality, we obtain the second claim of Theorem 11.

4. Applying a generalized von Neumann inequality

We continue to use the parameters $d, r, d_0, D, \kappa, W, A, R, N$ from the previous section. The objective of this section is to establish the following bound:

THEOREM 13. *For each $i = 1, \ldots, D$, let $b_i \in [W]$ be coprime to W. Let M_0 be a quantity with $\log^{1/\kappa} N \leqslant M_0 \leqslant AN^{1/d}$, and set $M := A^{-1}M_0$. For $i = 1, \ldots, D$, we set f_i to be a function with the pointwise bounds $|f_i| \ll \Lambda'_{b_i, W} + 1$. For each $i = 1, \ldots, d$, let $R_i \in \mathbb{Z}[\mathrm{m}_1, \ldots, \mathrm{m}_r]$ be a polynomial of degree at most d, with all coefficients bounded in magnitude by A. Assume also that for $i = 2, \ldots, D$, $R_i - R_1$ has degree at least d_0. Then*

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}\mathbb{E}_{\vec{m} \in [M_0]^r} \prod_{i=1}^{D} f_i(x + R_i(\vec{m})) \ll_{d,r,D} \|f_1\|^c + o(1) \qquad (56)$$

for some $1 \ll_{d,r,D} c \ll_{d,r,D} 1$, where $\|f_1\|$ is short for a norm of the form

$$\|f_1\| := \mathbb{E}_{\vec{h} \in [M]^r} \|f_1\|_{\square_{(P_i(\vec{h})[-M,M])_{i \in [D']}}^{D'}} \qquad (57)$$

for some natural number $D' = O_{d,r,D}(1)$, and polynomials $P_1, \ldots, P_{D'} \in \mathbb{Z}[\mathrm{h}_1, \ldots, \mathrm{h}_r]$ of degree between $d_0 - 1$ and $d - 1$ and coefficients $O(A^{O_{d,r,D}(1)})$.

The arguments in [27, Section 5] (based on van der Corput's method, the Cauchy–Schwarz inequality, PET induction, and a 'polynomial forms' condition on the pseudorandom majorant ν) 'morally' permit one to establish Theorem 13. However, the setup here differs from that in [27] in several minor technical aspects, most notably the multidimensional nature of the parameter $\vec{m}$, the nonconstancy of the b_i parameter in i, and the much larger value of the scale parameter M_0. As such, we need to adapt the argument from [27] to the current setting.

For inductive purposes it is convenient to prove a more general form of Theorem 13. We need the following definitions (inspired by, though not absolutely identical to, analogous definitions in [27]):

DEFINITION 1 (Polynomial system). A *polynomial system* $\mathcal{S}$ consists of the following objects:

- an integer $D_{\mathcal{S}} \geqslant 0$, which we call the *number of fine degrees of freedom*;

- a nonempty finite index set $\mathcal{A}$ (the elements of which we call the *nodes* of the system);

- a polynomial $R_\alpha \in \mathbb{Z}[m_1, \ldots, m_r, h_1, \ldots, h_{D_S}]$ of degree at most d attached to each node $\alpha \in \mathcal{A}$;

- a distinguished node $\alpha_0 \in \mathcal{A}$;

- a (possibly empty) collection $\mathcal{A}' \subset \mathcal{A}\backslash\{\alpha_0\}$ of *inactive nodes*. The nodes in $\mathcal{A}\backslash\mathcal{A}'$ will be referred to as *active*, thus for instance the distinguished node α_0 is always active.

We say that a node α is *linear* if $R_\alpha - R_{\alpha_0}$ is at most linear in $m_1, \ldots, m_r$ (that is it has degree at most 1 when viewed as a polynomial in $m_1, \ldots, m_r$ with coefficients in $\mathbb{Z}[h_1, \ldots, h_{D_S}]$), thus for instance the distinguished mode α_0 is always linear. We say that $\mathcal{S}$ is linear if all active nodes are linear. We require polynomial systems to obey three additional axioms:

- if α, β are distinct nodes in $\mathcal{A}$, then $R_\alpha - R_\beta$ is not constant in $m_1, \ldots, m_r$, $h_1, \ldots, h_{D_S}$ (that is it does not lie in $\mathbb{Z}$);

- if $\alpha \in \mathcal{A}\backslash\{\alpha_0\}$, then $R_\alpha - R_{\alpha_0}$ has degree at least d_0 in $m_1, \ldots, m_r, h_1, \ldots, h_{D_S}$;

- if α, β are distinct *linear* nodes in $\mathcal{A}$, then $R_\alpha - R_\beta$ is not constant in $m_1, \ldots, m_r$ (that is it does not lie in $\mathbb{Z}[h_1, \ldots, h_{D_S}]$).

DEFINITION 2 (Realizations and averages). Let $\mathcal{S}$ be a polynomial system. A *realization* $\vec{f} = (f_\alpha)_{\alpha \in \mathcal{A}}$ of $\mathcal{S}$ is an assignment of functions $f_\alpha : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ to each node α with the following properties:

- for any node α, one has the pointwise bound $|f_\alpha| \ll \nu_{b_\alpha} + 1$ for some $b_\alpha \in [W]$ coprime to W;

- for any inactive node α, one has $f_\alpha = \nu_{b_\alpha} + 1$ for some $b_\alpha \in [W]$ coprime to W.

We define the *average* $\Lambda_{\vec{S}}(\vec{f})$ to be the quantity

$$\Lambda_{\vec{S}}(\vec{f}) := \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{\vec{m} \in [M_0]^r} \mathbb{E}_{\vec{h} \in [M]^{D_S}} \prod_{\alpha \in \mathcal{A}} f_\alpha(x + R_\alpha(\vec{m}, \vec{h})). \tag{58}$$

Theorem 13 is then a special case of the following more general statement.

THEOREM 14. *Let C_0 be a quantity depending only on d, r, D, and assume κ sufficiently large depending on C_0. Let C_1 be a quantity depending only on d, r, D, C_0, κ (in particular, C_0, C_1 are independent of N). Let $\mathcal{S}$ be a system with at most C_0 nodes and at most C_0 fine degrees of freedom, with all polynomials R_α*

associated to the system having coefficients bounded in magnitude by A^{C_1}. Let $\vec{f}$ be a realization of $\mathcal{S}$. Then

$$\Lambda_{\vec{S}}(\vec{f}) \ll_{d,r,D,C_0,C_1} \|f_{\alpha_0}\|^c + o(1) \tag{59}$$

for some $c > 0$ depending on d, r, D, C_0, C_1, and $\|f_1\|$ is a norm of the form (57) with $D' = O_{d,r,D,C_0,C_1}(1)$, and $P_1, \ldots, P_{D'} \in \mathbb{Z}[\mathrm{h}_1, \ldots, \mathrm{h}_r]$ of degree between $d_0 - 1$ and $d - 1$ with coefficients $A^{O_{d,r,D,C_0,C_1}(1)}$.

Indeed, Theorem 13 is the special case in which the system $\mathcal{S}$ consists of the nodes $\mathcal{A} = \{1, \ldots, D\}$ with distinguished node $\alpha_0 = 1$ and all nodes active, with $D_{\mathcal{S}} = 0$, and R_i and b_i as indicated by Theorem 13.

It remains to establish Theorem 14. This will follow the same three-step procedure used in [27].

4.1. Reduction to the linear case.

The first step is to use the van der Corput method and PET induction to reduce matters to the linear case. We need some further definitions, again essentially from [27].

Given two nodes α, β in a polynomial system $\mathcal{S}$, we define the *distance* $d(\alpha, \beta)$ between the nodes to be the degree in $\mathrm{m}_1, \ldots, \mathrm{m}_r$ of the polynomial $R_\alpha - R_\beta$. This distance is symmetric, reflexive, and obeys the ultrametric triangle inequality

$$d(\alpha, \gamma) \leqslant \max(d(\alpha, \beta), d(\beta, \gamma)) \tag{60}$$

for all nodes α, β, γ. We define the *diameter* $\mathrm{diam}(\mathcal{S})$ of the system to be the maximal value of $d(\alpha, \beta)$ for α, β ranging over active nodes, and define an *extreme* node to be an active node α such that $d(\alpha, \alpha_0)$ is equal to the diameter of $\mathcal{S}$; note from the ultrametric triangle inequality that there is always at least one such node.

Given a node α, we then call two nodes β, γ *equivalent* relative to α if $d(\beta, \gamma) < d(\beta, \alpha)$; by (60), this is an equivalence relation on nodes, and every equivalence class has a well-defined distance to α. We then define the *weight* $\vec{w}_\alpha(\mathcal{S})$ of $\mathcal{S}$ relative to α to be the vector $(w_1, \ldots, w_d) \in \mathbb{Z}_+^d$, where w_i is the number of equivalence classes relative to α at a distance i from α. Thus for instance $\mathcal{S}$ will be linear if and only if the weight $\vec{w}_\alpha(\mathcal{S})$ relative to a node α takes the form $(w_1, 0, \ldots, 0)$. We order weights lexicographically, thus $(w_1, \ldots, w_d) < (w_1', \ldots, w_d')$ if there is $1 \leqslant i \leqslant d$ such that $w_i < w_i'$ and $w_j = w_j'$ for all $i < j \leqslant d$.

For a given choice of constants C_0, C_1 and a weight vector $\vec{w}$, let $P(C_0, C_1, \vec{w})$ denote the assertion that Theorem 14 holds for the given choice of C_0, C_1 and for all polynomial systems of weight $\vec{w}$ relative to some extreme

node α. (The requirement that α be extreme was mistakenly omitted in our previous paper [27]; it is needed to force the PET induction to terminate at a linear system.) The key inductive claim is then

PROPOSITION 7 (PET induction step). *For any $C_0, C_1, \vec{w} = (w_1, \ldots, w_d)$ which is nonlinear in the sense that $w_i \neq 0$ for some $i = 2, \ldots, d$, there exist a finite collection of triples $(C_0', C_1', \vec{w}')$ with $\vec{w}' < \vec{w}$, such that if $P(C_0', C_1', \vec{w}')$ holds for all triples in this collection, then $P(C_0, C_1, \vec{w})$ holds.*

Since the number of weight vectors $\vec{w}$ that can arise from systems $\mathcal{S}$ of at most C_0 nodes is finite, and the collection of all weight vectors is well ordered, we conclude from Proposition 7 that if $P(C_0, C_1, \vec{w})$ holds for all linear $\vec{w}$, then it holds for all $\vec{w}$. This implies that to prove Theorem 14, it suffices to do so in the case when $\mathcal{S}$ is linear.

We now establish Proposition 7. Let $\mathcal{S}$ be a polynomial system with at most C_0 nodes and at most C_0 fine degrees of freedom, and of weight $\vec{w}$ relative to some extreme node α, and with all polynomials having coefficients bounded in magnitude by A^{C_1}. Since $\vec{w}$ is nonlinear, the diameter of $\mathcal{S}$ is at least two. By subtracting R_α from each of the other R_β (noting that this does not affect the metric d or the average $\Lambda_{\vec{S}}(\vec{f})$), we may assume that $R_\alpha = 0$ (at the cost of increasing the coefficient bound from A^{C_1} to $2A^{C_1}$). As α was extreme, we now note that the polynomial R_{α_0} associated to α_0 has maximal m-degree among all the polynomials associated to active nodes.

We split $\mathcal{A} = \mathcal{A}_0 \cup \mathcal{A}_1$, where $\mathcal{A}_0$ is the set of nodes β with $d(\alpha, \beta) = 0$, and $\mathcal{A}_1$ is the set of nodes β with $d(\alpha, \beta) \geqslant 1$; note that the distinguished node α_0 lies in $\mathcal{A}_1$. We can then factor

$$\Lambda_{\vec{S}}(\vec{f}) = \mathbb{E}_{\vec{h} \in [M]^{D_{\mathcal{S}}}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} F_{\vec{h}}(x) \mathbb{E}_{\vec{m} \in [M_0]^r} G_{\vec{m}, \vec{h}}(x)$$

where

$$F_{\vec{h}}(x) := \prod_{\beta \in \mathcal{A}_0} f_\beta(x + R_\beta(0, \vec{h}))$$

and

$$G_{\vec{m}, \vec{h}}(x) := \prod_{\beta \in \mathcal{A}_1} f_\beta(x + R_\beta(\vec{m}, \vec{h})).$$

By hypothesis, each f_β is bounded in magnitude by $v_{b_\beta} + 1$ for some $b_\beta \in [W]$ coprime to W. Thus we have the pointwise bound

$$|F_{\vec{h}}(x)| \leqslant H_{\vec{h}}(x)$$

where

$$H_{\vec{h}}(x) := \prod_{\beta \in \mathcal{A}_0} (\nu_{b_\beta} + 1)(x + R_\beta(0, \vec{h})).$$

We can pointwise bound $|G_{\vec{m}, \vec{h}}(x)| \leqslant K_{\vec{m}, \vec{h}}(x)$ where

$$K_{\vec{m}, \vec{h}}(x) := \prod_{\beta \in \mathcal{A}_1} (\nu_{b_\beta} + 1)(x + R_\beta(\vec{m}, \vec{h})).$$

In Appendix A we establish the following bounds:

LEMMA 4. *With the notation as above, we have*

$$\mathbb{E}_{\vec{h} \in [M]^{D_\mathcal{S}}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} H_{\vec{h}}(x) = 2^{|\mathcal{A}_0|} + o(1) \tag{61}$$

and

$$\mathbb{E}_{\vec{h} \in [M]^{D_\mathcal{S}}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} H_{\vec{h}}(x)(\mathbb{E}_{\vec{m} \in [M]^r + \vec{a}} K_{\vec{m}, \vec{h}}(x))^2 = 2^{|\mathcal{A}_0| + 2|\mathcal{A}_1|} + o(1) \tag{62}$$

uniformly for all $\vec{a} \in \mathbb{Z}^r$.

By (61) and the Cauchy–Schwarz inequality, we see that to show (59), it suffices to show that

$$\mathbb{E}_{\vec{h} \in [M]^{D_\mathcal{S}}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} H_{\vec{h}}(x)(\mathbb{E}_{\vec{m} \in [M_0]^r} G_{\vec{m}, \vec{h}}(x))^2 \ll_{d,r,D,C_0,C_1} \|f_{\alpha_0}\|^{2c} + o(1). \tag{63}$$

We now apply the van der Corput method. By covering the boundary of $[M_0]^r$ by about $O_r(A^{r-1})$ translates of $[M]^r$, and using the pointwise bound $|G_{\vec{m}, \vec{h}}(x)| \leqslant K_{\vec{m}, \vec{h}}(x)$, we see that there is a collection Σ of $O_r(A^{r-1})$ elements $\vec{a}$ of $\mathbb{Z}^r$ such that

$$|\mathbb{E}_{\vec{m} \in [M_0]^r + [M]^r} G_{\vec{m}, \vec{h}}(x) - \mathbb{E}_{\vec{m} \in [M_0]^r} G_{\vec{m}, \vec{h}}(x)| \ll_r A^{-r} \sum_{\vec{a} \in \Sigma} \mathbb{E}_{\vec{m} \in [M]^r + \vec{a}} K_{\vec{m}, \vec{h}}(x)$$

and hence by (62) and the triangle inequality (and noting that $A^{-r} \times A^{r-1} = o(1)$)

$$\mathbb{E}_{\vec{h} \in [M]^{D_\mathcal{S}}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} H_{\vec{h}}(x)(\mathbb{E}_{\vec{m} \in [M_0]^r + [M]^r} G_{\vec{m}, \vec{h}}(x) - \mathbb{E}_{\vec{m} \in [M_0]^r} G_{\vec{m}, \vec{h}}(x))^2 = o(1)$$

for any $\vec{h} \in [M]^r$. Using this bound and the triangle inequality in L^2 (with respect to the measure implied by the averaging $\mathbb{E}_{\vec{h} \in [M]^{D_\mathcal{S}}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} H_{\vec{h}}(x) \dots$), we see that the bound (63) is equivalent to the bound

$$\mathbb{E}_{\vec{h} \in [M]^{D_\mathcal{S}}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} H_{\vec{h}}(x)(\mathbb{E}_{\vec{m} \in [M_0]^r + [M]^r} G_{\vec{m}, \vec{h}}(x))^2 \ll_{d,r,D,C_0,C_1} \|f_{\alpha_0}\|^{2c} + o(1).$$

By Cauchy–Schwarz, we can bound the left-hand side by

$$\mathbb{E}_{\vec{h}\in[M]^{D_S}}\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}H_{\vec{h}}(x)\mathbb{E}_{\vec{m}\in[M_0]^r}(\mathbb{E}_{\vec{h}'\in[M]^r}G_{\vec{m}+\vec{h}',\vec{h}}(x))^2$$

which we may expand as

$$\mathbb{E}_{\vec{h}\in[M]^{D_S}}\mathbb{E}_{h',h''\in[M]^r}\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}\mathbb{E}_{\vec{m}\in[M_0]^r}H_{\vec{h}}(x)G_{\vec{m}+\vec{h}',\vec{h}}(x)G_{\vec{m}+\vec{h}'',\vec{h}}(x).$$

Comparing this with (58), we see that this expression can be written as

$$\Lambda_{\mathcal{S}'}(\vec{f}')$$

where the polynomial system $\mathcal{S}'$ and the realization $\vec{f}' = (f'_\beta)_{\beta\in\mathcal{A}'}$ are defined as follows.

- The number of fine degrees of freedom is $D_{\mathcal{S}'} := D_{\mathcal{S}} + 2r$.

- The set of nodes $\mathcal{A}'$ consists of the disjoint union of $\mathcal{A}_0$, $\mathcal{A}_1$, and another copy $\mathcal{A}'_1$ of $\mathcal{A}_1$.

- The polynomials

  $$R'_\beta \in \mathbb{Z}[m_1,\ldots,m_r.h_1,\ldots,h_{D_S},h'_1,\ldots,h'_r,h''_1,\ldots,h''_r] = \mathbb{Z}[\vec{m},\vec{h},\vec{h}',\vec{h}'']$$

  for $\beta \in \mathcal{A}' = \mathcal{A}_0 \cup \mathcal{A}_1 \cup \mathcal{A}'_1$ are defined by setting

  $$R'_\beta := R_\beta(0,\vec{h})$$

  for $\beta \in \mathcal{A}_0$,

  $$R'_\beta := R_\beta(\vec{m}+\vec{h}',\vec{h})$$

  for $\beta \in \mathcal{A}_1$, and

  $$R'_{\beta'} := R_\beta(\vec{m}+\vec{h}'',\vec{h})$$

  for $\beta' \in \mathcal{A}'_1$ the copy of an element $\beta \in \mathcal{A}_1$.

- The distinguished node stays at α_0.

- The inactive nodes consist of all the nodes in $\mathcal{A}_0$, together with all the previously inactive nodes β of $\mathcal{A}_1$, as well as their copies β' in $\mathcal{A}'_1$.

- The realizations f'_β for $\beta \in \mathcal{A}' = \mathcal{A}_0 \cup \mathcal{A}_1 \cup \mathcal{A}'_1$ are defined by setting

  $$f'_\beta := v_{b_\beta} + 1$$

  for $\beta \in \mathcal{A}_0$, and

  $$f'_\beta = f'_{\beta'} = f_\beta$$

  for $\beta \in \mathcal{A}_1$, where β' is the copy of β in $\mathcal{A}'_1$.

It is a routine matter to check that $\mathcal{S}'$ obeys the axioms required for a polynomial system; its number of nodes and fine degrees of freedom are bounded by a constant C_0' depending only on C_0, r, and the polynomials have coefficients bounded by $A^{C_1'}$ for some C_1' depending only on C_1. Similarly one verifies that $\vec{f}'$ is indeed a realization of $\mathcal{S}'$.

Let d_* be the minimal distance of an active node of $\mathcal{A}_1 \cup \mathcal{A}_1'$ to α (or equivalently, the minimal m-degree of R_β' as β ranges over active nodes in $\mathcal{A}_1 \cup \mathcal{A}_1'$); this is well defined since $\alpha_0 \in \mathcal{A}_1$ is active. Among all the active nodes in $\mathcal{A}_1 \cup \mathcal{A}_1'$ at distance d_* from α, let $\tilde{\alpha}$ be a node that maximizes its distance from α_0. We claim that $\tilde{\alpha}$ is an extreme node of $\mathcal{S}'$, or in other words that

$$d(\beta, \alpha_0) \leqslant d(\tilde{\alpha}, \alpha_0)$$

for all active nodes β in $\mathcal{A}_1 \cup \mathcal{A}_1'$. By construction, this already holds when $d(\beta, \alpha) = d_*$, and it is not possible for $d(\beta, \alpha)$ to be less than d_*, so it remains to handle the case when $d(\beta, \alpha) > d_*$. But as α was extreme, this implies that

$$d(\alpha, \tilde{\alpha}) = d_* < d(\beta, \alpha) \leqslant \operatorname{diam}(\mathcal{S}) = d(\alpha, \alpha_0)$$

and hence by (60)

$$d(\tilde{\alpha}, \alpha_0) = \operatorname{diam}(\mathcal{S}) \geqslant d(\beta, \alpha_0)$$

giving the claim in this case also. Thus $\tilde{\alpha}$ is an extreme node in $\mathcal{S}'$.

Observe that if β is a node in $\mathcal{A}_1$ and β' its copy in $\mathcal{A}_1'$, then $R_\beta' - R_{\beta'}'$ has m-degree strictly less than that of R_β', thus the distance between β and β' is less than that between β and α, and hence β and β' are equivalent relative to α. From this we see that for $i > d_*$, the number of equivalence classes in $\mathcal{S}'$ relative to $\tilde{\alpha}$ at distance i is equal to the number of equivalence classes in $\mathcal{S}$ relative to α, while for $i = d_*$, $\mathcal{S}'$ has one fewer equivalence class relative to $\tilde{\alpha}$ at distance d_* than $\mathcal{S}$ relative to α (since $\tilde{\alpha}$ contributes a class to the latter but not the former). Thus the weight vector $\vec{w}'$ of $\mathcal{S}'$ relative to $\tilde{\alpha}$ is less than the weight vector $\vec{w}$ of $\mathcal{S}$ relative to α. Furthermore, given the bounds on the number of nodes and fine degrees of freedom, the weight vector $\vec{w}'$ ranges in a finite set that depends on $d, r, C_0, C_1, \vec{w}$. Applying the hypothesis $P(C_0, C_1, \vec{w}')$, we obtain the claim.

4.2. Parallelopipedization.

It remains to establish the linear case of Theorem 14. As in [27], the next step is 'parallelopipedization', in which one repeatedly uses the Cauchy–Schwarz inequality to reduce matters to controlling a weighted averaged local Gowers norm.

Let the notation be as in Theorem 14, with $\mathcal{S}$ linear. We abbreviate $(\mathrm{h}_1, \ldots, \mathrm{h}_{D_{\mathcal{S}}})$ as $\vec{h}$. By subtracting R_{α_0} from all of the other polynomials

R_α, we may assume that $R_{\alpha_0} = 0$, so that all the active R_α have degree at most one in m. We write $\mathcal{A}_l$ for those nodes at distance one from α_0, and $\mathcal{A}_{nl}$ for all nodes at distance greater than one, thus $\mathcal{A}$ is partitioned into $\{\alpha_0\}, \mathcal{A}_l, \mathcal{A}_{nl}$, with $\mathcal{A}_{nl}$ consisting entirely of inactive nodes, so that for each $\alpha \in \mathcal{A}_{nl}$, f_α is equal to $\nu_{b_\alpha} + 1$ for some $b_\alpha \in [W]$ coprime to W. By the triangle inequality we may replace each of the f_α with either ν_{b_α} or 1; by deleting all nodes with $f_\alpha = 1$ we can assume that only the latter case $f_\alpha = \nu_{b_\alpha}$ occurs for $\alpha \in \mathcal{A}_{nl}$.

For each $\alpha \in \mathcal{A}_l$ one has

$$R_\alpha = \vec{a}_\alpha \cdot \vec{\mathrm{m}} + c_\alpha$$

for some nonzero $\vec{a}_\alpha \in \mathbb{Z}[\vec{h}]^r$ and some $c_\alpha \in \mathbb{Z}[\vec{h}]$, with $\cdot$ denoting the usual dot product; furthermore from the axioms of a linear system we see that the $\vec{a}_\alpha$ are distinct as α varies. We can then write $\Lambda_{\vec{s}}(\vec{f})$ as

$$\mathbb{E}_{\vec{h} \in [M]^{D_\mathcal{S}}, \vec{m} \in [M_0]^r} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f_{\alpha_0}(x) \left(\prod_{\alpha \in \mathcal{A}_{nl}} \nu_{b_\alpha}(x + R_\alpha(\vec{m}, \vec{h})) \right)$$

$$\times \left(\prod_{\alpha \in \mathcal{A}_l} f_\alpha(x + \vec{a}_\alpha(\vec{h}) \cdot \vec{\mathrm{m}} + c_\alpha(\vec{h})) \right).$$

We need to show that this expression is $o(1)$. Arguing as in the previous section, we may replace the set $[M_0]^r$ that $\vec{m}$ is being averaged over by the multiset

$$[M_0]^r + \sum_{\alpha \in \mathcal{A}_l} [M]^r$$

(that is to say, the sum of $[M_0]^r$ and $|\mathcal{A}_l|$ copies of $[M]^r$, counting multiplicity). Thus it suffices to show that the expression

$$\mathbb{E}_{\vec{h} \in [M]^{D_\mathcal{S}}} \mathbb{E}_{\vec{m} \in [M_0]^r} \mathbb{E}_{\vec{k}_\alpha \in [M]^r \forall \alpha \in \mathcal{A}_l} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f_{\alpha_0}(x)$$

$$\times \left(\prod_{\alpha \in \mathcal{A}_{nl}} \nu_{b_\alpha}\left(x + R_\alpha\left(\vec{m} + \sum_{\beta \in \mathcal{A}_l} \vec{k}_\beta\right), \vec{h}\right) \right)$$

$$\times \left(\prod_{\alpha \in \mathcal{A}_l} f_\alpha\left(x + \vec{a}_\alpha(\vec{h}) \cdot \vec{\mathrm{m}} + \sum_{\beta \in \mathcal{A}_l} \vec{a}_\alpha(\vec{h}) \cdot \vec{k}_\beta + c_\alpha(\vec{h})\right) \right) \quad (64)$$

is $\ll_{r,D,C_0,C_1} \|f_{\alpha_0}\|^c + o(1)$. We shift x by $-\sum_{\beta \in \mathcal{A}_l} \vec{a}_\beta(\vec{h}) \cdot \vec{k}_\beta$ to write this expression as

$$\mathbb{E}_{\vec{h} \in [M]^{D_\mathcal{S}}} \mathbb{E}_{\vec{m} \in [M_0]^r} \mathbb{E}_{\vec{k}_\alpha \in [M]^r \forall \alpha \in \mathcal{A}_l} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f_{\alpha_0, \vec{m}, \vec{h}, \vec{k}}(x) \prod_{\alpha \in \mathcal{A}_l} f_{\alpha, \vec{m}, \vec{h}, \vec{k}}(x)$$

where $\vec{k} := (\vec{k}_\alpha)_{\alpha \in \mathcal{A}_l}$, with

$$f_{\alpha_0, \vec{m}, \vec{h}, \vec{k}}(x) := f_{\alpha_0}\left(x - \sum_{\beta \in \mathcal{A}_l} \vec{a}_\beta(\vec{h}) \cdot \vec{k}_\beta\right)$$

$$\times \prod_{\alpha \in \mathcal{A}_{nl}} v_{b_\alpha}\left(x + R_\alpha\left(\vec{m} + \sum_{\beta \in \mathcal{A}_l} \vec{k}_\beta, \vec{h}\right) - \sum_{\beta \in \mathcal{A}_l} \vec{a}_\beta(\vec{h}) \cdot \vec{k}_\beta\right)$$

and

$$f_{\alpha, \vec{m}, \vec{h}, \vec{k}}(x) := f_\alpha\left(x + \vec{a}_\alpha(\vec{h}) \cdot \vec{m} + \sum_{\beta \in \mathcal{A}_l} (\vec{a}_\alpha(\vec{h}) - \vec{a}_\beta(\vec{h})) \cdot \vec{k}_\beta + c_\alpha(\vec{h})\right)$$

for $\alpha \in \mathcal{A}_l$. The key point here is that $f_{\alpha, \vec{m}, \vec{h}, \vec{k}}$ does not depend on the α component $\vec{k}_\alpha$ of $\vec{k}$. We also have the pointwise bound

$$|f_{\alpha, \vec{m}, \vec{h}, \vec{k}}(x)| \leqslant v_{\alpha, \vec{m}, \vec{h}, \vec{k}}(x)$$

for all $\alpha \in \mathcal{A}_l$, where $v_{\alpha, \vec{m}, \vec{h}, \vec{k}}(x)$ is either identically equal to 1, or is given by the formula

$$v_{\alpha, \vec{m}, \vec{h}, \vec{k}}(x) = v_{b_\alpha}\left(x + \vec{a}_\alpha(\vec{h}) \cdot \mathrm{m} + \sum_{\beta \in \mathcal{A}_l} (\vec{a}_\alpha(\vec{h}) - \vec{a}_\beta(\vec{h})) \cdot \vec{k}_\beta + c_\alpha(\vec{h})\right).$$

For sake of exposition we assume that the latter holds for all α, as this is the most difficult case. As with $f_{\alpha, \vec{m}, \vec{h}, \vec{k}}$, the quantity $v_{\alpha, \vec{m}, \vec{h}, \vec{k}}(x)$ does not depend on the α component of $\vec{k}$.

Applying the weighted Cauchy–Schwarz–Gowers inequality (see [**27**, Proposition A.2] or [**13**, Corollary B.4]), we can thus bound upper bound the absolute value of (**64**) by

$$\mathbb{E}_{\vec{h} \in [M]^{D_S}} \mathbb{E}_{\vec{m} \in [M_0]^r} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \| f_{\alpha_0, \vec{m}, \vec{h}, \cdot}(x) \|_{\square^{\mathcal{A}_l}(v)} \prod_{\alpha \in \mathcal{A}_l} \| v_{\alpha, \vec{m}, \vec{h}, \cdot}(x) \|_{\square^{\mathcal{A}_l \setminus \alpha}}^{1/2}$$

where

$$\| f_{\alpha_0, \vec{m}, \vec{h}, \cdot}(x) \|_{\square^{\mathcal{A}_l}(v)}^{2^{|\mathcal{A}_l|}} := \mathbb{E}_{\vec{k}^{(0)}, \vec{k}^{(1)} \in ([M]^r)^{\mathcal{A}_l}} \left[\prod_{\omega \in \{0,1\}^{\mathcal{A}_l}} f_{\alpha_0, \vec{m}, \vec{h}, \vec{k}^{(\omega)}}(x) \right]$$

$$\times \prod_{\alpha \in \mathcal{A}_l} \prod_{\omega \in \{0,1\}^{\mathcal{A}_l \setminus \{\alpha\}}} v_{\alpha, \vec{m}, \vec{h}, \vec{k}^{(\omega)}}(x)$$

and

$$\| v_{\alpha, \vec{m}, \vec{h}, \cdot}(x) \|_{\square^{\mathcal{A}_l \setminus \alpha}}^{2^{|\mathcal{A}_l|-1}} := \mathbb{E}_{\vec{k}^{(0)}, \vec{k}^{(1)} \in ([M]^r)^{\mathcal{A}_l \setminus \{\alpha\}}} \prod_{\omega \in \{0,1\}^{\mathcal{A}_l \setminus \{\alpha\}}} v_{\alpha, \vec{m}, \vec{h}, \vec{k}^{(\omega)}}(x),$$

where for $\vec{k}^{(0)} = (\vec{k}_\beta^{(0)})_{\beta \in \mathcal{A}_l}$, $\vec{k}^{(1)} = (\vec{k}_\beta^{(1)})_{\beta \in \mathcal{A}_l}$, and $\omega = (\omega_\beta)_{\beta \in \mathcal{A}_l}$ one has

$$\vec{k}^{(\omega)} := (\vec{k}_\beta^{(\omega_\beta)})_{\beta \in \mathcal{A}_l};$$

if $\vec{k}^{(0)}, \vec{k}^{(1)}$ is only given in $([M]^r)^{\mathcal{A}_l \setminus \{\alpha\}}$, we extend it arbitrarily to $([M]^r)^{\mathcal{A}_l}$ for the purposes of defining $\vec{k}^{(\omega)}$, and similarly if ω is only given in $\{0, 1\}^{\mathcal{A}_l \setminus \{\alpha\}}$ instead of $\{0, 1\}^{\mathcal{A}_l}$. This leaves the α component of $\vec{k}^{(\omega)}$ undefined, but this is irrelevant for the purposes of evaluating $v_{\alpha, \vec{m}, \vec{h}, \vec{k}^{(\omega)}}(x)$ because (as mentioned previously) this quantity does not depend on the α component of $\vec{k}^{(\omega)}$.

In Appendix A we establish the following estimate:

LEMMA 5. *With the notation as above, we have*

$$\mathbb{E}_{\vec{h} \in [M]^{D_S}} \mathbb{E}_{\vec{m} \in [M_0]^r} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \|v_{\alpha, \vec{m}, \vec{h}, \cdot}(x)\|_{\square^{\mathcal{A}_l \setminus \alpha}}^{2^{|\mathcal{A}_l| - 1}} = 1 + o(1). \tag{65}$$

Thus, by Hölder's inequality (and modifying c as necessary), to show that the expression in (64) is $o(1)$, it suffices to establish the bound

$$\mathbb{E}_{\vec{h} \in [M]^{D_S}} \mathbb{E}_{\vec{m} \in [M_0]^r} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \|f_{\alpha_0, \vec{m}, \vec{h}, \cdot}(x)\|_{\square^{\mathcal{A}_l}_{(v)}}^{2^{|\mathcal{A}_l|}} \ll \|f_{\alpha_0}\|^c + o(1). \tag{66}$$

This is a weighted version of (a special case of) the Cauchy–Schwarz–Gowers inequality (6), and will be deduced from that inequality by one final application of the polynomial forms condition.

4.3.   Final Cauchy–Schwarz.   We now finish the proof of (66). By definition, the left-hand side of this estimate expands as

$$\mathbb{E}_{\vec{h} \in [M]^{D_S}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{\vec{k}^{(0)}, \vec{k}^{(1)} \in ([M]^r)^{\mathcal{A}_l}} w(\vec{h}, \vec{k}^{(0)}, \vec{k}^{(1)}, x)$$
$$\times \prod_{\omega \in \{0,1\}^{\mathcal{A}_l}} f_{\alpha_0}\left(x - \sum_{\beta \in \mathcal{A}_l} \vec{a}_\beta(\vec{h}) \cdot \vec{k}_\beta^{(\omega_\beta)}\right)$$

where

$$w(\vec{h}, \vec{k}^{(0)}, \vec{k}^{(1)}, x)$$
$$:= \mathbb{E}_{\vec{m} \in [M_0]^r} \left(\prod_{\alpha \in \mathcal{A}_l} \prod_{\omega \in \{0,1\}^{\mathcal{A}_l \setminus \{\alpha\}}} v_{\alpha, \vec{m}, \vec{h}, \vec{k}^{(\omega)}}(x)\right)$$
$$\times \left(\prod_{\alpha \in \mathcal{A}_{nl}} \prod_{\omega \in \{0,1\}^{\mathcal{A}_l}} v_{b_\alpha}\left(x + R_\alpha\left(\vec{m} + \sum_{\beta \in \mathcal{A}_l} \vec{k}_\beta^{(\omega_\beta)}, \vec{h}\right)\right) - \sum_{\beta \in \mathcal{A}_l} \vec{a}_\beta(\vec{h}) \cdot \vec{k}_\beta\right).$$

(In the analogous expansion in [**27**, Section 5.19], the terms arising from $\alpha \in \mathcal{A}_l$ were mistakenly omitted.) On the other hand, if we identify $\mathcal{A}_l$ with $\{1, \dots, D\}$, then $D \leqslant C_0$ and the expression

$$\mathbb{E}_{\vec{h} \in [M]^{D_S}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{\vec{k}^{(0)}, \vec{k}^{(1)} \in ([M]^r)^{\mathcal{A}_l}} \prod_{\omega \in \{0,1\}^{\mathcal{A}_l}} f_{\alpha_0}\left(x - \sum_{\beta \in \mathcal{A}_l} \vec{a}_\beta(\vec{h}) \cdot \vec{k}_\beta^{(\omega_\beta)}\right)$$

can be bounded in magnitude using the Cauchy–Schwarz–Gowers inequality (6) by

$$\mathbb{E}_{\vec{h} \in [M]^{D_S}} \| f_{\alpha_0} \|_{\Box_{(a_{ij}(\vec{h})[M])_{1 \leqslant i \leqslant D; 1 \leqslant j \leqslant r}}^{rD}}^{2^{rD}} \tag{67}$$

where $a_{i1}, \dots, a_{ir} \in \mathbb{Z}[\vec{h}]$ are the components of a_i. From Proposition 2 (and the monotonicity properties of the Gowers norms) we have

$$\mathbb{E}_{\vec{h} \in [M]^{D_S}} \| \nu - 1 \|_{\Box_{(a_{ij}(\vec{h})[M])_{1 \leqslant i \leqslant D; 1 \leqslant j \leqslant r}}^{rD}}^{2^{rD+1}} = o(1)$$

which by the triangle inequality implies that

$$\mathbb{E}_{\vec{h} \in [M]^{D_S}} \| \nu + 1 \|_{\Box_{(a_{ij}(\vec{h})[M])_{1 \leqslant i \leqslant D; 1 \leqslant j \leqslant r}}^{rD}}^{2^{rD+1}} \ll_{r,D} 1 + o(1)$$

and hence

$$\mathbb{E}_{\vec{h} \in [M]^{D_S}} \| f_{\alpha_0} \|_{\Box_{(a_{ij}(\vec{h})[M])_{1 \leqslant i \leqslant D; 1 \leqslant j \leqslant r}}^{rD}}^{2^{rD+1}} \ll_{r,D} 1.$$

Thus by Hölder's inequality, the quantity (67) is bounded by

$$\ll_{r,D} \left(\mathbb{E}_{\vec{h} \in [M]^{D_S}} \| f_{\alpha_0} \|_{\Box_{(a_{ij}(\vec{h})[M])_{1 \leqslant i \leqslant D; 1 \leqslant j \leqslant r}}^{rD}} \right)^c$$

for some $c > 0$ depending on r, D. This is an expression of the form $\| f_\alpha \|^c$. Thus, by the triangle inequality, it suffices to show that

$$\mathbb{E}_{\vec{h} \in [M]^{D_S}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{\vec{k}^{(0)}, \vec{k}^{(1)} \in ([M]^r)^{\mathcal{A}_l}} (w(\vec{h}, \vec{k}^{(0)}, \vec{k}^{(1)}, x) - 1)$$
$$\times \prod_{\omega \in \{0,1\}^{\mathcal{A}_l}} f_{\alpha_0}\left(x - \sum_{\beta \in \mathcal{A}_l} \vec{a}_\beta(\vec{h}) \cdot \vec{k}_\beta^{(\omega_\beta)}\right) = o(1).$$

Bounding f_{α_0} in magnitude by $\nu_{b_{\alpha_0}} + 1$ and using Cauchy–Schwarz, it suffices to show that

$$\mathbb{E}_{\vec{h} \in [M]^{D_S}} \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{\vec{k}^{(0)}, \vec{k}^{(1)} \in ([M]^r)^{\mathcal{A}_l}} (w(\vec{h}, \vec{k}^{(0)}, \vec{k}^{(1)}, x) - 1)^2$$
$$\times \prod_{\omega \in \{0,1\}^{\mathcal{A}_l}} \left(\nu_{b_{\alpha_0}}\left(x - \sum_{\beta \in \mathcal{A}_l} \vec{a}_\beta(\vec{h}) \cdot \vec{k}_\beta^{(\omega_\beta)}\right) + 1 \right) = o(1).$$

Expanding out the square, it suffices to show that

$$\mathbb{E}_{\vec{h}\in[M]^{D_S}}\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}\mathbb{E}_{\vec{k}^{(0)},\vec{k}^{(1)}\in([M]^r)^{\mathcal{A}_l}}\,w(\vec{h},\vec{k}^{(0)},\vec{k}^{(1)},x)^j$$

$$\times\prod_{\omega\in\{0,1\}^{\mathcal{A}_l}}\left(v_{b_{\alpha_0}}\left(x-\sum_{\beta\in\mathcal{A}_l}\vec{a}_\beta(\vec{h})\cdot\vec{k}_\beta^{(\omega_\beta)}\right)+1\right)=2^{|\mathcal{A}_l|}+o(1)$$

for $j=0,1,2$.

We just treat the most difficult case $j=2$, as the other cases $j=0,1$ are similar (with fewer v-type factors). Expanding out the second product, it suffices to show that

$$\mathbb{E}_{\vec{h}\in[M]^{D_S}}\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}\mathbb{E}_{\vec{k}^{(0)},\vec{k}^{(1)}\in([M]^r)^{\mathcal{A}_l}}\,w(\vec{h},\vec{k}^{(0)},\vec{k}^{(1)},x)^2$$

$$\times\prod_{\omega\in\Omega}v_{b_{\alpha_0}}\left(x-\sum_{\beta\in\mathcal{A}_l}\vec{a}_\beta(\vec{h})\cdot\vec{k}_\beta^{(\omega_\beta)}\right)=1+o(1)$$

for any $\Omega\subset\{0,1\}^{\mathcal{A}_l}$. The left-hand side may be expanded as

$$\mathbb{E}_{\vec{h}\in[M]^{D_S}}\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}\mathbb{E}_{\vec{k}^{(0)},\vec{k}^{(1)}\in([M]^r)^{\mathcal{A}_l}}\mathbb{E}_{\vec{m}^{(0)},\vec{m}^{(1)}\in[M_0]^r}\prod_{i=0}^{1}$$

$$\times\left(\prod_{\alpha\in\mathcal{A}_{nl}}\prod_{\omega\in\{0,1\}^{\mathcal{A}_l}}v_{b_\alpha}\left(x+R_\alpha\left(\vec{m}^{(i)}+\sum_{\beta\in\mathcal{A}_l}\vec{m}_\beta^{(i)},\vec{h}\right)-\sum_{\beta\in\mathcal{A}_l}\vec{a}_\beta(\vec{h})\cdot\vec{k}_\beta\right)\right)$$

$$\times\prod_{\alpha\in\mathcal{A}_l}\prod_{\omega\in\{0,1\}^{\mathcal{A}_l\setminus\{\alpha\}}}v_{b_\alpha}\left(x+\vec{a}_\alpha(\vec{h})\cdot\vec{m}^{(i)}\right)$$

$$+\sum_{\beta\in\mathcal{A}_l}(\vec{a}_\alpha(\vec{h})-\vec{a}_\beta(\vec{h}))\cdot\vec{k}_\beta^{(\omega_\beta)}+c_\alpha(\vec{h})\bigg). \tag{68}$$

But this is $1+o(1)$ thanks to the polynomial forms property of the measures v_b (see Appendix A). This completes the proof of Theorem 14 and hence Theorem 13.

5. The W-trick

Theorem 13 has a particularly pleasant consequence in the setting where all polynomials involved are distinct to top order.

COROLLARY 4. *Let d,r,D,W,A,N be as in previous sections. For each $i=1,\ldots,D$, let $b_i\in[W]$ be coprime to W. For each $i=1,\ldots,d$, let $R_i\in\mathbb{Z}[\mathrm{m}_1,\ldots,\mathrm{m}_r]$ be a polynomial of degree at most d, with all coefficients bounded in magnitude by A. Assume also that for $1\leqslant i<j\leqslant D$, R_i-R_j has*

degree exactly d. Let M_0 be a quantity with $A^{-1}N^{1/d} \leqslant M_0 \leqslant AN^{1/d}$. Then

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}\mathbb{E}_{\vec{m} \in [M_0]^r} \prod_{i=1}^{D} \Lambda'_{b_i, W}(x + R_i(\vec{m})) = 1 + o(1).$$

Proof. The claim is equivalent to

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}\mathbb{E}_{\vec{m} \in [M_0]^r} \left(\prod_{i=1}^{D} \Lambda'_{b_i, W}(x + R_i(\vec{m})) \right) - 1 = o(1).$$

By writing $\Lambda'_{b_i, W} = 1 + (\Lambda'_{b_i, W} - 1)$, one can expand the left-hand side as the sum of $2^D - 1$ terms, each of which is bounded in magnitude by $O(\|\Lambda'_{b_i, W} - 1\|^c) + o(1)$ for some i and some fixed $c > 0$ by Theorem 13 (after permuting the indices), where the norm is of the form (57) with $d_0 = d$. On the other hand, from Theorem 11 we have $\|\Lambda'_{b_i, W} - 1\| = o(1)$. The claim follows. □

With this corollary and the 'W-trick' (as in [13, Section 5]), we can now prove Theorem 4. Let $d, r, k, P_1, \ldots, P_r, M()$ be as in Theorem 4. We let d, r, D, W, A, R, N, N' be as in previous sections, with d, r as previously chosen and D set equal to k. By replacing N by N', and using the convergence of the infinite product $\prod_p \beta_p$ and the fact that w goes to infinity, it suffices to show that

$$\mathbb{E}_{n' \in [N']}\mathbb{E}_{m' \in [M']^r} \Lambda(n' + P_1(m')) \ldots \Lambda(n' + P_k(m')) = \prod_{p \leqslant w} \beta_p + o(1)$$

where $M' := M(N')$.

The contribution to the left-hand side of the case when one of the $n' + P_i(m')$ is a prime power (rather than a prime) can easily be seen to be $o(1)$ (in fact one obtains a power savings in N'), so we may replace Λ by its restriction Λ' to the primes without loss of generality.

We split n' and m' into residue classes $n' = b\ (W)$ and $m' = c\ (W)$ for $b \in [W]$ and $c \in [W]^r$. Call a pair (b, c) *admissible* if $b + P_i(c)$ is coprime to W for all $i = 1, \ldots, k$. From the Chinese remainder theorem, we see that the number of admissible pairs is $W^{r+1}(\phi(W)/W)^k \prod_{p \leqslant w} \beta_p$. For inadmissible (b, c), the quantity $\Lambda'(n' + P_1(m')) \ldots \Lambda'(n' + P_k(m'))$ is only nonvanishing when one of the $n' + P_i(m')$ is a prime $p \leqslant w$. It is not difficult to see that the total contribution of such a case is $o(1)$ if w is sufficiently slowly growing with respect to N'. Thus we may restrict attention to admissible pairs (b, c). Approximating the arithmetic progression $\{n' \in [N'] : n' = b\ (W)\}$ by $\{Wn + b : n \in [N]\}$, and similarly approximating $\{m' \in [M']^r : m' = c\ (W)\}$ by $\{Wm + c : m \in [M]^r\}$ with $M := \lfloor M'/W \rfloor$ (using crude estimates to bound the error in this

approximation by multiplicative and additive errors of $o(1)$, assuming w sufficiently slowly growing), it will thus suffice to show that

$$\mathbb{E}_{n \in [N]} \mathbb{E}_{m \in [M]^r} \Lambda'(Wn + b + P_1(Wm + c)) \dots \Lambda'(Wn + b + P_k(Wm + c))$$
$$= \left(\frac{W}{\phi(W)}\right)^k (1 + o(1)) \tag{69}$$

uniformly for all admissible pairs (b, c).

As (b, c) is admissible, we can write

$$Wn + b + P_i(Wm + c) = W(n + R_i(m)) + b_i$$

for some $b_i \in [W]$ coprime to W, and some polynomial $R_i \in \mathbb{Z}[\mathrm{m}_1, \dots, \mathrm{m}_r]$ of degree d with all coefficients bounded in magnitude by A. Since the $P_i - P_j$ all had degree d, the $R_i - R_j$ do also. Recalling that $M = o(N^{1/d})$, we see that the quantities $n + R_i(m)$ will lie in the interval $[R, N]$ unless $n = o(N)$ or $n = N - o(N)$. The contribution of these latter cases to (69) can easily be verified to be $o((W/\phi(W))^k)$ by any standard upper bound sieve (for example the Selberg sieve, or the 'fundamental lemma of sieve theory', see for example [7, Theorem 6.12]). Using (34), we thus see that (69) is equivalent to the estimate

$$\mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{m \in [M]^r} \Lambda'_{b_1, W}(n + R_1(m)) \dots \Lambda'_{b_k, W}(n + R_k(m)) = 1 + o(1).$$

But this follows from Corollary 4 (noting that the lower bound on M' will imply that $M \geqslant A^{-1} N^{1/d}$ if ω is going to zero sufficiently slowly). This concludes the proof of Theorem 4.

We now adapt the above arguments to prove Theorem 5. Repeating the above arguments all the way up to (69), we arrive at the task of showing that

$$\mathbb{E}_{n \in [N], m \in [M]} \Lambda'(Wn + b) \Lambda'(Wn + b + (Wm + c)) \Lambda'(Wn + b + P_3(Wm + c))$$
$$= \left(\frac{W}{\phi(W)}\right)^3 (1 + o(1))$$

uniformly for $(b, c) \in [W]^2$ with $b, b + c, b + P_3(c)$ coprime to W, with $M = o(\sqrt{N})$ and $M \geqslant \omega(N)N$ for some function $\omega(N)$ that goes to zero sufficiently slowly. Continuing the above arguments, we then reduce to showing that

$$\mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{m \in [M]} \Lambda'_{b_1, W}(n) \Lambda'_{b_2, W}(n + R_2(m)) \Lambda'_{b_3, W}(n + R_3(m)) = 1 + o(1)$$

where $b_1, b_2, b_3 \in [W]$ are given by the congruences

$$b_1 = b(W)$$
$$b_2 = b + c(W)$$
$$b_3 = b + P_3(c)(W)$$

and R_2, R_3 are the polynomials

$$R_2 := \mathrm{m} + \frac{b + c - b_2}{W}$$

$$R_3 := \frac{P_3(W\mathrm{m} + c) - b_3}{W}.$$

Using Theorem 13, we see that

$$\mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{m \in [M]} \Lambda'_{b_1, W}(n) \Lambda'_{b_2, W}(n + R_2(m))(\Lambda'_{b_3, W}(n + R_3(m)) - 1) = o(1)$$

so it suffices to show that

$$\mathbb{E}_{n \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{m \in [M]} \Lambda'_{b_1, W}(n) \Lambda'_{b_2, W}(n + R_2(m)) = 1 + o(1)$$

which on reversing some of the arguments following (69) is equivalent to

$$\mathbb{E}_{n \in [N]} \mathbb{E}_{m \in [M]} \Lambda'(Wn + b) \Lambda'(Wn + b + (Wm + c)) = \left(\frac{W}{\phi(W)} \right)^2 (1 + o(1)).$$

Set $M_0 := \lfloor M \log^{-10} N \rfloor$. Using crude bounds on Λ', we may replace the average $[M]$ by $[M] - [M_0]$ with negligible error, and then by shifting n by an element of M_0 and incurring a further negligible error, we may reduce to showing that

$$\mathbb{E}_{n \in [N]} \mathbb{E}_{h \in [M_0]} \mathbb{E}_{m \in [M]} \Lambda'(W(n + h) + b) \Lambda'(Wn + b + (Wm + c))$$
$$= \left(\frac{W}{\phi(W)} \right)^2 (1 + o(1)).$$

The left-hand side factors as

$$\mathbb{E}_{n \in [N]} (\mathbb{E}_{h \in [M_0]} \Lambda'(W(n + h) + b))(\mathbb{E}_{m \in [M]} \Lambda'(Wn + b + (Wm + c))).$$

From the prime number theorem in arithmetic progressions we have

$$\mathbb{E}_{n \in [N]} \mathbb{E}_{m \in [M]} \Lambda'(Wn + b + (Wm + c)) = \left(\frac{W}{\phi(W)} \right)(1 + o(1))$$

if w is sufficiently slowly growing, so it suffices to show that

$$\mathbb{E}_{n \in [N]} \left(\mathbb{E}_{h \in [M_0]} \frac{\phi(W)}{W} \Lambda'(W(n + h) + b) - 1 \right)$$
$$\times \left(\frac{\phi(W)}{W} \mathbb{E}_{m \in [M]} \Lambda'(Wn + b + (Wm + c)) \right) = o(1).$$

From the Brun–Titchmarsh inequality, the expression $(\phi(W)/W)\mathbb{E}_{m\in[M]}\Lambda'$ $(Wn + b + (Wm + c))$ is bounded by $O(1)$, so by the Cauchy–Schwarz inequality it suffices to show that

$$\mathbb{E}_{n\in[N]}\left(\mathbb{E}_{h\in[M_0]}\frac{\phi(W)}{W}\Lambda'(W(n+h)+b)-1\right)^2 = o(1).$$

Since $d \leqslant 5$, we have $M_0 \geqslant N^{1/6+\varepsilon}$, and the above claim then follows from standard zero-density estimates (see for example [**19**, Theorem 1.1]). On the generalized Riemann hypothesis, one can obtain this claim for M_0 as low as N^ε, and then no restriction on d is necessary; again, see [**19**, Theorem 1.1]. (In fact it suffices to assume the generalized density hypothesis.)

REMARK 2. The same method lets us handle a triplet of polynomials P_1, P_2, $P_3 \in \mathbb{Z}[\mathrm{m}]$ in which $P_3 - P_1$ has degree d and $P_2 - P_1$ has degree k for some $1 \leqslant k < d$ with $k/d > 1/6$ (and the hypothesis $k/d > 1/6$ can be omitted on the generalized Riemann or density hypothesis). Indeed, the above arguments let us reduce to showing an estimate of the form

$$\mathbb{E}_{n\in[N]}\mathbb{E}_{m\in[M]}\Lambda'(Wn+b)\Lambda'(Wn+b+(P_2-P_1)(Wm+c)) = \left(\frac{W}{\phi(W)}\right)^2(1+o(1)),$$

and a standard application of the circle method (using some Fourier restriction theorem for the von Mangoldt function on short intervals) lets us control this expression in turn by averages of Λ on arithmetic progressions of spacing $O(W^{O(1)})$ and length roughly $N^{k/d}$, which can again be controlled by zero-density estimates as before. We leave the details to the interested reader.

One may in principle be able to handle some higher complexity patterns of this type, for example $P_1 = 0$, $P_2 = \mathrm{m}^k$, $P_3 = 2\mathrm{m}^k$, $P_4 = \mathrm{m}^d$ when $1 \leqslant k < d$ with k/d sufficiently close to 1. Morally speaking, after using some suitable adaptation of the arguments in this paper and the (nontrivial) fact that the pattern $n, n+m^k, n+2m^k$ has 'true complexity' 1 in the sense of Gowers and Wolf [**11**], the average corresponding to this set of polynomials should be controlled by an expression roughly of the form

$$\mathbb{E}_{n\in[N]}\sup_{\alpha\in\mathbb{R}/\mathbb{Z}}|\mathbb{E}_{m\in[M^k]}(\Lambda'(W(n+m)+b)-1)e(\alpha m)|$$

and one would expect to be able to control this quantity when k/d is large from existing analytic number theory methods; the best result in this direction we currently know of is by Zhan [**30**], who used moment bounds on L-functions to treat the case $k/d > 5/8$. Again, we will not pursue the details of these arguments further here.

Acknowledgements

Appendix A.   The polynomial forms condition

Throughout this appendix, the parameters $d, r, D, \kappa, W, A, R, N$ are as in Section 3, with all quantities below allowed to depend on d, r. We now prove the various polynomial forms conditions required on the functions ν_b, specifically Proposition 2, Lemmas 3–5, and showing that the quantity (68) is $1 + o(1)$.

Our starting point is the following estimate on the ν_b from [27, 28]:

THEOREM 15 (Polynomial forms condition). *Let k, s, C be natural numbers not depending on N, let $b_1, \ldots, b_s \in [W]$ be coprime to W, and let $P_1, \ldots, P_k \in \mathbb{Z}[\mathrm{m}_1, \ldots, \mathrm{m}_s]$ be distinct polynomials of degree at most d, with all nonconstant coefficients of size at most A^C in magnitude. Assume that κ is sufficiently small depending on k, s. Assume also that the $P_i - P_j$ are nonconstant for all $1 \leqslant i < j \leqslant s$. Let $M_1, \ldots, M_s$ be quantities with $\log^{1/\kappa} N \leqslant M_1, \ldots, M_s \leqslant N$. Then*

$$\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \mathbb{E}_{m_1 \in [M_1], \ldots, m_s \in [M_s]} \prod_{i=1}^{k} \nu_{b_i}(x + P_i(m_1, \ldots, m_s)) = 1 + o(1).$$

Proof. In the case $b_1 = \cdots = b_s$ (and when the constant coefficients of the P_i are also bounded in magnitude by A^C), this follows directly from [27, Corollary 11.2] in the case when the $M_1, \ldots, M_s$ are bounded below by (say) N^κ, and from [28, Proposition 3] in the general case. (Strictly speaking, [28, Proposition 3] only claims the case when $M_1 = \cdots = M_s = \log^L N$ for some sufficiently large L depending on k, s, but the arguments easily extend to larger values of $M_1, \ldots, M_S$.) The arguments in [27, Sections 10, 11] used to prove [27, Corollary 11.2] or [28, Proposition 3] can be easily modified to handle the case when the more general case when the b_i are permitted to be distinct and the constant coefficients are permitted to be large, after replacing every occurrence of $WP_j + b$ with $WP_j + b_j$ in these arguments. (The notion of a 'terrible' prime has to then be modified to be a prime $p > w$ that divides $(WP_i + b_i) - (WP_j + b_j)$ for some $1 \leqslant i < j \leqslant s$, rather than just dividing $P_i - P_j$; however, these polynomials $(WP_i + b_i) - (WP_j + b_j)$ are nonconstant with all nonconstant

coefficients $O(WA^C)$, and the arguments used to prove [27, Corollary 11.2] or [28, Proposition 3] still show that the total contribution of the terrible primes only contributes a multiplicative factor of $O(1)$ to the error.) □

Lemma 3 then follows from this theorem (in the $b_1 = \cdots = b_s = b$ case), as after expanding out the fourth power one obtains a sum of 4^{2^D-1} expressions, all of which are $1 + o(1)$ thanks to Theorem 15 (with $s = 4(r + D)$, and k at most $4(2^D - 1)$). Proposition 2 similarly follows from this theorem (in the $b_1 = \cdots = b_s = b$ case), since on expanding out the left-hand side of (39), one obtains an alternating sum of $2^{2^{D_*+2D}}$ terms, all of which are $1 + o(1)$ thanks to Theorem 15 (with $s = 2r + D_* + 2D$, and k at most 2^{D_*+2D}). In both of these cases, a direct inspection reveals that the polynomials P_i used in the invocation of Theorem 15 have nonconstant differences $P_i - P_j$.

In a similar vein, Lemma 5 follows from Theorem 15 (now with the b_i all distinct), as the left-hand side of (5) expands as an expression of the form considered by Theorem 15 (with $s = D_S + r + 2r(|\mathcal{A}_l| - 1)$ and $k = 2^{|\mathcal{A}_l|-1}$). Similarly, for Lemma 4, the left-hand side of (61) expands as $2^{|\mathcal{A}_0|}$ terms, all of which are $1 + o(1)$ by Theorem 15 (with $s = D_S$ and k at most $|\mathcal{A}_0|$), and the left-hand side of (62) similarly expands as the sum of $2^{|\mathcal{A}_0|+2|\mathcal{A}_1|}$ terms, which are again $1 + o(1)$ by Theorem 15 (with $s = D_S + 2r$ and k at most $|\mathcal{A}_0| + 2|\mathcal{A}_1|$); it is in this latter case that we need to permit the nonconstant coefficients of the polynomials P_i in Theorem 15 to be larger than A^C in magnitude. As before, an inspection of the polynomials involved (using the fact that the $R_\alpha - R_\beta$ are nonconstant) shows that the $P_i - P_j$ are nonconstant.

Finally, the expression (68) is $1 + o(1)$ by an application of Theorem 15 with $s = D_S + 2r|\mathcal{A}_l| + 2r$ and $k = |\mathcal{A}_{nl}|2^{|\mathcal{A}_l|} + |\mathcal{A}_l|2^{|\mathcal{A}_l|-1}$. (Again, by focusing on the behavior with respect to the $\vec{m}^{(i)}$ variables, setting all other variables to zero, one can use the hypotheses on the $R_\alpha - R_\beta$ to show that the polynomials $P_i - P_j$ are nonconstant.)

References

[1] P. Bateman and R. Horn, 'A heuristic asymptotic formula concerning the distribution of prime numbers', *Math. Comput.* **16** (1962), 363–367.

[2] V. Bergelson and A. Leibman, 'Polynomial extensions of van der Waerden's and Szemerédi's theorems', *J. Amer. Math. Soc.* **9**(3) (1996), 725–753.

[3] P.-Y. Bienvenu, 'Asymptotics for some polynomial patterns in the primes', Preprint, 2015, arXiv:1511.07317.

[4] D. Conlon, B. Fox and Y. Zhao, 'A relative Szemerédi theorem', *Geom. Funct. Anal.* **25**(3) (2015), 733–762.

[5] B. Cook and A. Magyar, 'Diophantine equations in the primes', *Invent. Math.* **198**(3) (2014), 701–737.

[6]   K. Ford, B. Green, S. Konyagin and T. Tao, 'Large gaps between consecutive prime numbers', *Ann. of Math.* (2) **183**(3) (2016), 935–974.

[7]   J. Friedlander and H. Iwaniec, *Opera de Cribro*, American Mathematical Society Colloquium Publications, 57 (American Mathematical Society, Providence, 2010).

[8]   W. T. Gowers, 'A new proof of Szemerédi's theorem for arithmetic progressions of length four', *Geom. Funct. Anal.* **8**(3) (1998), 529–551.

[9]   W. T. Gowers, 'A new proof of Szemerédi's theorem', *Geom. Funct. Anal.* **11**(3) (2001), 465–588.

[10] W. T. Gowers, 'Decompositions, approximate structure, transference, and the Hahn–Banach theorem', *Bull. Lond. Math. Soc.* **42**(4) (2010), 573–606.

[11] W. T. Gowers and J. Wolf, 'The true complexity of a system of linear equations', *Proc. Lond. Math. Soc.* (3) **100**(1) (2010), 155–176.

[12] B. Green and T. Tao, 'The primes contain arbitrarily long arithmetic progressions', *Ann. of Math.* (2) **167**(2) (2008), 481–547.

[13] B. Green and T. Tao, 'Linear equations in primes', *Ann. of Math.* (2) **171**(3) (2010), 1753–1850.

[14] B. Green and T. Tao, 'The Möbius function is strongly orthogonal to nilsequences', *Ann. of Math.* (2) **175**(2) (2012), 541–566.

[15] B. Green and T. Tao, 'The quantitative behaviour of polynomial orbits on nilmanifolds', *Ann. of Math.* (2) **175**(2) (2012), 465–540.

[16] B. Green, T. Tao and T. Ziegler, 'An inverse theorem for the Gowers $U^{s+1}[N]$-norm', *Ann. of Math.* (2) **176**(2) (2012), 1231–1372.

[17] 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.

[18] M. N. Huxley, 'On the difference between consecutive primes', *Invent. Math.* **15** (1972), 164–170.

[19] D. Koukoulopoulos, 'Primes in short arithmetic progressions', *Int. J. Number Theory* **11**(5) (2015), 1499–1521.

[20] T. H. Le, 'Intersective polynomials and the primes', *J. Number Theory* **130**(8) (2010), 1705–1717.

[21] T. H. Le and J. Wolf, 'Polynomial configurations in the primes', *Int. Math. Res. Not. IMRN* (23) (2014), 6448–6473.

[22] K. Matomäki and M. Radziwiłł, 'Multiplicative functions in short intervals', *Ann. of Math.* (2) **183**(3) (2016), 1015–1056.

[23] K. Matomäki, M. Radziwiłł and T. Tao, 'An averaged form of Chowla's conjecture', *Algebra Number Theory* **9**(9) (2015), 2167–2196.

[24] O. Reingold, L. Trevisan, M. Tulsiani and S. Vadhan, 'New proofs of the Green–Tao–Ziegler dense model theorem: an exposition', Preprint, 2008, arXiv:0806.0381.

[25] A. Schinzel and W. Sierpiński, 'Sur certaines hypothèses concernant les nombres premiers', *Acta Arith.* **4** (1958), 185–208. Erratum **5** (1959), 259.

[26] E. Szemerédi, 'On sets of integers containing no k elements in arithmetic progression', *Acta Arith.* **27** (1975), 199–245. Collection of articles in memory of Juriĭ Vladimirovič Linnik.

[27] T. Tao and T. Ziegler, 'The primes contain arbitrarily long polynomial progressions', *Acta Math.* **201** (2008), 213–305. Erratum, *Acta Math.* **210**(2) (2013), 403–404.

[28] T. Tao and T. Ziegler, 'Narrow progressions in the primes', in *Analytic Number Theory* (Springer, Cham, 2015), 357–379.

[29] T. Tao and T. Ziegler, 'Concatenation theorems for anti-Gowers-uniform functions and Host–Kra characteristic factors', *Discrete Anal.* (2016), Paper No. 13, 60 pp.

[30] T. Zhan, 'On the representation of large odd integer as a sum of three almost equal primes', *Acta Math. Sinica (N.S.)* **7**(3) (1991), 259–272.