THE PRIMES CONTAIN ARBITRARILY LONG ARITHMETIC PROGRESSIONS

BEN GREEN AND TERENCE TAO

Abstract. We prove that there are arbitrarily long arithmetic progressions of primes.

There are three major ingredients. The first is Szemerédi's theorem, which asserts that any subset of the integers of positive density contains progressions of arbitrary length. The second, which is the main new ingredient of this paper, is a certain transference principle. This allows us to deduce from Szemerédi's theorem that any subset of a sufficiently pseudorandom set of positive *relative* density contains progressions of arbitrary length. The third ingredient is a recent result of Goldston and Yildirim, which we reproduce here. Using this, one may place the primes (or more precisely, a large fraction of the primes) inside a pseudorandom set of "almost primes" with positive relative density.

1. Introduction

It is a well-known conjecture that there are arbitrarily long arithmetic progressions of prime numbers. The conjecture is best described as "classical", or maybe even "folklore". In Dickson's *History* it is stated that around 1770 Lagrange and Waring investigated how large the common difference of an arithmetic progression of L primes must be, and it is hard to imagine that they did not at least wonder whether their results were sharp for all L.

It is not surprising that the conjecture should have been made, since a simple heuristic based on the prime number theorem would suggest that there are $\gg N^2/\log^k N$ k-tuples of primes $p_1, \ldots, p_k$ in arithmetic progression, each p_i being at most N. Hardy and Littlewood [21], in their famous paper of 1923, advanced a very general conjecture which, as a special case, contains the hypothesis that the number of such k-term progressions is asymptotically $C_k N^2/\log^k N$ for a certain explicit numerical factor C_k (we do not come close to establishing this conjecture here, obtaining instead a lower bound $(\gamma(k)+o(1))N^2/\log^k N$ for some very small $\gamma(k) > 0$).

The first theoretical progress on these conjectures was made by van der Corput [37] (see also [7]) who, in 1939, used Vinogradov's method of prime number sums to establish the case $k = 3$, that is to say that there are infinitely many triples of primes in arithmetic progression. However, the question of longer arithmetic progressions seems to have remained completely open (except for upper bounds), even for $k = 4$. On the other hand, it has been known for some time that better results can be obtained if one

1991 *Mathematics Subject Classification.* 11N13, 11B25, 374A5.

The first author is a PIMS postdoctoral fellow at the University of British Columbia, Vancouver, Canada. The second author is a Clay Prize Fellow and is supported by a grant from the Packard Foundation.

replaces the primes with a slightly larger set of *almost primes*. The most impressive such result is due to Heath-Brown [22]. He showed that there are infinitely many 4-term progressions consisting of three primes and a number which is either prime or a product of two primes. In a somewhat different direction, let us mention the beautiful results of Balog [2, 3]. Among other things he shows that for any m there are m distinct primes $p_1, \ldots, p_m$ such that all of the averages $\frac{1}{2}(p_i + p_j)$ are prime.

The problem of finding long arithmetic progressions in the primes has also attracted the interest of computational mathematicians. At the time of writing the longest known arithmetic progression of primes is of length 23, and was found by in 2004 Markus Frind, Paul Underwood, and Paul Jobling:

$$56211383760397 + 44546738095860k; \quad k = 0, 1, \ldots, 22.$$

An earlier arithmetic progression of primes of length 22 was found by Moran, Pritchard and Thyssen [28]:

$$11410337850553 + 4609098694200k; \quad k = 0, 1, \ldots, 21.$$

Our main theorem resolves the above conjecture.

Theorem 1.1. *The prime numbers contain infinitely many arithmetic progressions of length k for all k.*

In fact, we can say something a little stronger:

Theorem 1.2 (Szemerédi's theorem in the primes)**.** *Let A be any subset of the prime numbers of positive relative upper density, thus* $\limsup_{N\to\infty} \pi(N)^{-1}|A \cap [1, N]| > 0$*, where $\pi(N)$ denotes the number of primes less than or equal to N. Then A contains infinitely many arithmetic progressions of length k for all k.*

If one replaces "primes" in the statement of Theorem 1.2 by the set of all positive integers $\mathbb{Z}^+$, then this is a famous theorem of Szemerédi [34]. The special case $k = 3$ of Theorem 1.2 was recently established by the first author [18] using methods of Fourier analysis. In contrast, our methods here have a more ergodic theory flavour and do not involve much Fourier analysis (though the argument does rely on Szemerédi's theorem which can be proven by either combinatorial, ergodic theory, or harmonic analysis arguments).

Acknowledgements The authors would like to thank Jean Bourgain, Enrico Bombieri, Tim Gowers, Bryna Kra, Elon Lindenstrauss, Imre Ruzsa, Roman Sasyk, Peter Sarnak and Kannan Soundararajan for helpful conversations. We are particularly indebted to Andrew Granville for drawing our attention to the work of Goldston and Yıldırım, and to Dan Goldston for making the preprint [14] available. We are also indebted to Jamie Radcliffe, Lior Silberman, and Mark Watkins for corrections to an earlier manuscript. Portions of this work were completed while the first author was visiting UCLA and Université de Montréal, and he would like to thank these institutions for their hospitality. He would also like to thank Trinity College, Cambridge for support over several years.

2. An outline of the proof

Let us start by stating Szemerédi's theorem properly. In the introduction we claimed that it was a statement about sets of integers with positive upper density, but there are other equivalent formulations. A "finitary" version of the theorem is as follows.

Proposition 2.1 (Szemerédi's theorem)**.** [33, 34] *Let N be a positive integer and let $\mathbb{Z}_N := \mathbb{Z}/N\mathbb{Z}$.*[1] *Let $\delta > 0$ be a fixed positive real number, and let $k \geqslant 3$ be an integer. Then there is a minimal $N_0(\delta, k) < \infty$ with the following property. If $N \geqslant N_0(\delta, k)$ and $A \subseteq \mathbb{Z}_N$ is any set of cardinality at least δN, then A contains an arithmetic progression of length k.*

Finding the correct dependence of N_0 on δ and k (particularly δ) is a famous open problem. It was a great breakthrough when Gowers [15, 16] showed that

$$N_0(\delta, k) \leqslant 2^{2^{\delta^{-c_k}}},$$

where c_k is an explicit constant (Gowers obtains $c_k = 2^{2^{k+9}}$). It is possible that a new proof of Szemerédi's theorem could be found, with sufficiently good bounds that Theorem 1.1 would follow immediately. To do this one would need something just a little weaker than

$$N_0(\delta, k) \leqslant 2^{c_k \delta^{-1}} \tag{2.1}$$

(there is a trick, namely passing to a subprogression of common difference $2 \times 3 \times 5 \times \cdots \times w(N)$ for appropriate $w(N)$, which allows one to consider the primes as a set of density essentially $\log \log N/ \log N$ rather than $1/ \log N$). In our proof of Theorem 1.2, we will need to use Szemerédi's theorem, but we will not need any quantitative estimimates on $N_0(\delta, k)$.

Let us state, for contrast, the best known lower bound which is due to Rankin [31] (see also Lacey-Łaba [27]):

$$N_0(\delta, k) \geqslant \exp(C(\log 1/\delta)^{1+\lfloor \log_2(k-1) \rfloor}).$$

At the moment it is clear that a substantial new idea would be required to obtain a result of the strength (2.1). In fact, even for $k = 3$ the best bound is $N_0(\delta, 3) \leqslant 2^{C\delta^{-2} \log(1/\delta)}$, a result of Bourgain [6]. The hypothetical bound (2.1) is closely related to the following very open conjecture of Erdős and Turán:

Conjecture 2.2 (Erdős-Turán)**.** [8] *Suppose that $A = \{a_1 < a_2 < \dots\}$ is an infinite sequence of integers such that $\sum 1/a_i = \infty$. Then A contains arbitrarily long arithmetic progressions.*

This would imply Theorem 1.1.

We do not make progress on any of these issues here. In one sentence, our argument can be described instead as a *transference principle* which allows us to deduce Theorems 1.1 and 1.2 from Szemerédi's theorem, regardless of what bound we know for $N_0(\delta, k)$; in fact we prove a more general statement in Theorem 3.5 below. Thus, in this paper, we must assume Szemerédi's theorem. However with this one (rather large!) caveat[2] our paper is self-contained.

[1] We will retain this notation throughout the paper. We always assume for convenience that N is prime. It is very convenient to work in $\mathbb{Z}_N$, rather than the more traditional $[-N, N]$, since we are free to divide by $2, 3, \dots, k$ and it is possible to make linear changes of variables without worrying about the ranges of summation.

[2] We will also require some standard facts from analytic number theory such as the prime number theorem, Dirichlet's theorem on primes in arithmetic progressions, and the classical zero-free region for the Riemann ζ-function (see Lemma 11.1).

Szemerédi's theorem can now be proved in several ways. The original proof of Szemerédi [33, 34] was combinatorial. In 1977, Furstenberg made a very important breakthrough by providing an ergodic-theoretic proof [9]. Perhaps surprisingly for a result about primes, our paper has at least as much in common with the ergodic theoretic approach as it does with the harmonic analysis approach of Gowers. We will use a language which suggests this close connection, without actually relying explicitly on any ergodic-theoretical concepts[3]. In particular we shall always remain in the finitary setting of $\mathbb{Z}_N$, in contrast to the standard ergodic theory framework in which one takes weak limits (invoking the axiom of choice) to pass to an infinite measure-preserving system. As will become clear in our argument, in the finitary setting one can still access many tools and concepts from ergodic theory, but often one must incur error terms of the form $o(1)$ when one does so.

Here is another form of Szemerédi's theorem which suggests the ergodic theory analogy more closely. We use the conditional expectation notation $\mathbb{E}(f|x_i \in B)$ to denote the average of f as certain variables x_i range over the set B, and $o(1)$ for a quantity which tends to zero as $N \to \infty$ (we will give more precise definitions later).

Proposition 2.3 (Szemerédi's theorem, again)**.** *Write $\nu_{\mathrm{const}} : \mathbb{Z}_N \to \mathbb{R}^+$ for the constant function $\nu_{\mathrm{const}} \equiv 1$. Let $0 < \delta \leqslant 1$ and $k \geqslant 1$ be fixed. Let N be a large integer parameter, and let $f : \mathbb{Z}_N \to \mathbb{R}^+$ be a non-negative function obeying the bounds*

$$0 \leqslant f(x) \leqslant \nu_{\mathrm{const}}(x) \text{ for all } x \in \mathbb{Z}_N \tag{2.2}$$

and

$$\mathbb{E}(f(x)|x \in \mathbb{Z}_N) \geqslant \delta. \tag{2.3}$$

Then we have

$$\mathbb{E}(f(x)f(x+r)\dots f(x+(k-1)r)|x, r \in \mathbb{Z}_N) \geqslant c(k,\delta) - o_{k,\delta}(1)$$

for some constant $c(k,\delta) > 0$ which does not depend on f or N.

Remark. Ignoring for a moment the curious notation for the constant function ν_{const}, there are two main differences between this and Proposition 2.1. One is the fact that we are dealing with functions rather than sets: however, it is easy to pass from sets to functions by probabilistic arguments. Another difference, if one unravels the $\mathbb{E}$ notation, is that we are now asserting the existence of $\gg N^2$ arithmetic progressions, and not just one. Once again, such a statement can be deduced from Proposition 2.1 with some combinatorial trickery (of a less trivial nature this time – the argument was first worked out by Varnavides [38]). A formulation of Szemerédi's theorem similar to this one was also used by Furstenberg [9]. Combining this argument with the one in Gowers gives an explicit bound on $c(k,\delta)$ of the form $c(k,\delta) \geqslant \exp(-\exp(\delta^{-c_k}))$ for some $c_k > 0$.

[3] It has become clear that there is a deep connection between harmonic analysis (as applied to solving linear equations in sets of integers) and certain parts of ergodic theory. Particularly exciting is the suspicion that the notion of a k-step nilsystem, explored in many ergodic-theoretical works (see e.g. [24, 25, 26, 39]), might be analogous to a kind of "higher order Fourier analysis" which could be used to deal with systems of linear equations that cannot be handled by conventional Fourier analysis (a simple example being the equations $x_1 + x_3 = 2x_2$, $x_2 + x_4 = 2x_3$, which define an arithmetic progression of length 4). We will not discuss such speculations any further here, but suffice it to say that much is left to be understood.

Now let us abandon the notion that ν is the constant function. We say that $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ is a *measure*[4] if $\mathbb{E}(\nu) = 1 + o(1)$. We are going to exhibit a class of measures, more general than the constant function ν_{const}, for which Proposition 2.3 still holds. These measures, which we will call *pseudorandom*, will be ones satisfying two conditions called the *linear forms condition* and the *correlation condition*. These are, of course, defined formally below, but let us remark that they are very closely related to the ergodic-theory notion of weak-mixing. It is perfectly possible for a "singular" measure - for instance, a measure for which $\mathbb{E}(\nu^2)$ grows like a power of $\log N$ - to be pseudorandom. Singular measures are the ones that will be of interest to us, since they generally support rather sparse sets. This generalisation of Proposition 2.3 is Proposition 3.5 below.

Once Proposition 3.5 is proved, we turn to the issue of finding primes in AP. A possible choice for ν would be Λ, the von Mangoldt function (this is defined to equal $\log p$ at p^m, $m = 1, 2, \dots$, and 0 otherwise). Unfortunately, verifying the linear forms condition and the correlation condition for the von Mangoldt function (or minor variants thereof) is strictly harder than proving that the primes contain long arithmetic progressions.

However, all we need is a measure ν which (after rescaling by at most a constant factor) majorises Λ pointwise. Then, (2.3) will be satisfied with $f = \Lambda$. Such a measure is provided to us[5] by recent work of Goldston and Yıldırım [14] concerning the size of gaps between primes. The proof that the linear forms condition and the correlation condition are satisfied is heavily based on their work, so much so that parts of the argument are placed in an appendix.

The idea of using a majorant to study the primes is by no means new – indeed in some sense sieve theory is precisely the study of such objects. For another use of a majorant in an additive-combinatorial setting, see [29, 30].

It is now timely to make a few remarks concerning the proof of Proposition 3.5. It is in the first step of the proof that our original investigations began, when we made a close examination of Gowers' arguments. If $f : \mathbb{Z}_N \to \mathbb{R}^+$ is a function then the normalised count of k-term arithmetic progressions

$$\mathbb{E}(f(x)f(x+r)\dots f(x+(k-1)r)|x, r \in \mathbb{Z}_N) \tag{2.4}$$

is closely controlled by certain norms $\|\cdot\|_{U^d}$, which we would like to call the *Gowers uniformity norms*[6]. They are defined in §5. The formal statement of this fact can be called a generalised von Neumann theorem. Such a theorem, in the case $\nu = \nu_{\mathrm{const}}$, was proved by Gowers [16] as a first step in his proof of Szemerédi's theorem, using $k-2$ applications of the Cauchy-Schwarz inequality. In Proposition 5.3 we will prove

[4]The term *probability density* might be more accurate here, but *measure* has the advantage of brevity.

[5]Actually, there is a slight extra technicality which is caused by the very irregular distribution of primes in arithmetic progressions to small moduli (there are no primes congruent to $4 (\mathrm{mod}\, 6)$, for example). We get around this using something which we refer to as the W-trick, which basically consists of restricting the primes to the arithmetic progression $n \equiv 1 (\mathrm{mod}\, W)$, where $W = \prod_{p<w(N)} p$ and $w(N)$ tends slowly to infinity with N. Although this looks like a trick, it is actually an extremely important feature of that part of our argument which concerns primes.

[6]Analogous objects have recently surfaced in the genuinely ergodic-theoretical work of Host and Kra [24, 25, 26] concerning non-conventional ergodic averages, thus enhancing the connection between ergodic theory and additive number theory.

a generalised von Neumann theorem relative to an arbitrary pseudorandom measure ν. Our main tool is again the Cauchy-Schwarz inequality. We will use the term *uniform* loosely to describe a function which is small in some U^d norm. This should not be confused with the term pseudorandom, which will be reserved for measures on $\mathbb{Z}_N$.

Sections 6 and 8 are devoted to concluding the proof of Proposition 3.5. Very roughly the strategy will be to decompose the function f under consideration into a uniform component plus a bounded "anti-uniform" object (plus a negligible error). The notion of anti-uniformity is captured using the dual norms $(U^d)^*$, whose properties are laid out in §6.

The contribution of the uniform part to the count (2.4) will be negligible[7] by the generalised von Neumann theorem. The contribution from the anti-uniform component will be bounded from below by Szemerédi's theorem in its traditional form, Proposition 2.3.

3. Pseudorandom measures

In this section we specify exactly what we mean by a pseudorandom measure on $\mathbb{Z}_N$. First, however, we set up some notation. We fix the length k of the arithmetic progressions we are seeking. $N = |\mathbb{Z}_N|$ will always be assumed to be prime and large (in particular, we can invert any of the numbers $1, \ldots, k$ in $\mathbb{Z}_N$), and we will write $o(1)$ for a quantity that tends to zero as $N \to \infty$. We will write $O(1)$ for a bounded quantity. Sometimes quantities of this type will tend to zero (resp. be bounded) in a way that depends on some other, typically fixed, parameters. If there is any danger of confusion as to what is being proved, we will indicate such dependence using subscripts. Since every quantity in this paper will depend on k, however, we will usually not bother indicating the k dependence throughout this paper. As is customary we often abbreviate $O(1)X$ and $o(1)X$ as $O(X)$ and $o(X)$ respectively for various non-negative quantities X.

If A is a finite non-empty set (for us A is usually just $\mathbb{Z}_N$) and $f : A \to \mathbb{R}$ is a function, we write $\mathbb{E}(f) := \mathbb{E}(f(x)|x \in A)$ for the average value of f, that is to say

$$\mathbb{E}(f) := \frac{1}{|A|} \sum_{x \in A} f(x).$$

Here, as is usual, we write $|A|$ for the cardinality of the set A. More generally, if $P(x)$ is any statement concerning an element of A which is true for at least one $x \in A$, we define

$$\mathbb{E}(f(x)|P(x)) := \frac{\sum_{x \in A : P(x)} f(x)}{|\{x \in A : P(x)\}|}.$$

This notation extends to functions of several variables in the obvious manner. We now define two notions of randomness for a measure, which we term the linear forms condition and the correlation condition.

[7] Using the language of ergodic theory, we are essentially claiming that the anti-uniform functions form a characteristic factor for the expression (2.4). The point is that even though f is not necessarily bounded uniformly, the fact that it is bounded pointwise by a pseudorandom measure ν allows us to conclude that the *projection* of f to the anti-uniform component is bounded, at which point we can invoke the standard Szemerédi theorem.

Definition 3.1 (Linear forms condition)**.** Let $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ be a measure. Let m_0, t_0 and L_0 be small positive integer parameters. Then we say that ν satisfies the (m_0, t_0, L_0)-linear forms condition if the following holds. Let $m \leqslant m_0$ and $t \leqslant t_0$ be arbitrary, and suppose that $(L_{ij})_{1 \leqslant i \leqslant m, 1 \leqslant j \leqslant t}$ are arbitrary rational numbers with numerator and denominator at most L_0 in absolute value, and that b_i, $1 \leqslant i \leqslant m$, are arbitrary elements of $\mathbb{Z}_N$. For $1 \leqslant i \leqslant m$, let $\psi_i : \mathbb{Z}_N^t \to \mathbb{Z}_N$ be the linear forms $\psi_i(\mathbf{x}) = \sum_{j=1}^t L_{ij} x_j + b_i$, where $\mathbf{x} = (x_1, \dots, x_t) \in \mathbb{Z}_N^t$, and where the rational numbers L_{ij} are interpreted as elements of $\mathbb{Z}_N$ in the usual manner (assuming N is prime and larger than L_0). Suppose that as i ranges over $1, \dots, m$, the t-tuples $(L_{ij})_{1 \leqslant j \leqslant t} \in \mathbb{Q}^t$ are non-zero, and no t-tuple is a rational multiple of any other. Then we have

$$\mathbb{E}\left(\nu(\psi_1(\mathbf{x})) \dots \nu(\psi_m(\mathbf{x})) \mid \mathbf{x} \in \mathbb{Z}_N^t\right) = 1 + o_{L_0, m_0, t_0}(1). \tag{3.1}$$

Note that the rate of decay in the $o(1)$ term is assumed to be uniform in the choice of $b_1, \dots, b_m$.

Remark. It is the parameter m_0, which controls the number of linear forms, that is by far the most important, and will be kept relatively small. It will eventually be set equal to $k \cdot 2^{k-1}$. Note that the $m = 1$ case of the linear forms condition recovers the measure condition $\mathbb{E}(\nu) = 1 + o(1)$. For the application to the primes, the measure ν will be constructed using truncated divisor sums, and the linear forms condition will be deduced from some arguments of Goldston and Yıldırım.

Definition 3.2 (Correlation condition)**.** Let $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ be a measure, and let m_0 be a positive integer parameter. We say that ν satisfies the m_0-correlation condition if for every $1 < m \leqslant m_0$ there exists a weight function $\tau = \tau_m : \mathbb{Z}_N \to \mathbb{R}^+$ which obeys the moment conditions

$$\mathbb{E}(\tau^q) = O_{m,q}(1) \tag{3.2}$$

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

$$\mathbb{E}(\nu(x + h_1)\nu(x + h_2) \dots \nu(x + h_m) | x \in \mathbb{Z}_N) \leqslant \sum_{1 \leqslant i < j \leqslant m} \tau(h_i - h_j) \tag{3.3}$$

for all $h_1, \dots, h_m \in \mathbb{Z}_N$ (not necessarily distinct).

Remarks. The condition (3.3) may look a little strange, since if ν were to be chosen randomly then we would expect such a condition to hold with $1 + o(1)$ on the right-hand side. The condition has been designed with the primes in mind, because in that case we must tolerate slight "arithmetic" nonuniformities. Observe, for example, that the number of $p \leqslant N$ for which $p - h$ is also prime is not bounded above by a constant times $N/\log^2 N$ if h contains a very large number of prime factors, although such exceptions will of course be very rare and one still expects to have moment conditions such as (3.2). It is phenomena like this which prevent us from assuming an L^∞ bound for τ. While m_0 will be restricted to be small (in fact, equal to 2^{k-1}), it will be important for us that there is no upper bound required on q (which we will eventually need to be a very large function of k, but still independent of N of course). Since the correlation condition is an upper bound rather than an asymptotic, it is fairly easy to obtain; we shall prove it using the arguments of Goldston and Yıldırım (since we are using those methods in any case to prove the linear forms condition), but these upper bounds could also be obtained by more standard sieve theory methods.

Definition 3.3 (Pseudorandom measures)**.** Let $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ be a measure. We say that ν is k-pseudorandom if it satisfies the $(k \cdot 2^{k-1}, 3k-4, k)$-linear forms condition and also the 2^{k-1}-correlation condition.

Remarks. The exact values of the parameters m_0, t_0, L_0 chosen here are not too important; in our application to the primes, any quantities which depend only on k would suffice. It can be shown that if $C = C_k > 1$ is any constant independent of N and if $S \subseteq \mathbb{Z}_N$ is chosen at random, each $x \in \mathbb{Z}_N$ being selected to lie in S independently at random with probability $1/\log^C N$, then the measure $\nu = \log^C N \mathbf{1}_S$ is k-pseudorandom, and it is conjectured that the Von Mangoldt function is essentially of this form (once one eliminates the obvious obstructions to pseudorandomness coming from small prime divisors). While we will not attempt to establish this conjecture here, in §9 we will construct pseudorandom measures which are *nearly* supported on the primes; this is of course consistent with the so-called "fundamental lemma of sieve theory", but we will need a rather precise variant of this lemma due to Goldston and Yıldırım.

The function $\nu_{\mathrm{const}} \equiv 1$ is clearly k-pseudorandom for any k. In fact the pseudorandom measures are star-shaped around the constant measure:

Lemma 3.4. *Let ν be a k-pseudorandom measure. Then $\nu_{1/2} := (\nu + \nu_{\mathrm{const}})/2 = (\nu+1)/2$ is also a k-pseudorandom measure.*

Proof. It is clear that $\nu_{1/2}$ is non-negative and has expectation $1 + o(1)$. To verify the linear forms condition (3.1), we simply replace ν by $(\nu+1)/2$ in the definition and expand as a sum of 2^m terms, divided by 2^m. Since each term can be verified to be $1+o(1)$ by the linear forms condition (3.1), the claim follows. The correlation condition is verified in a similar manner. (A similar result holds for $(1-\theta)\nu + \theta\nu_{\mathrm{const}}$ for any $0 \leqslant \theta \leqslant 1$, but we will not need to use this generalization). □

The following result is one of the main theorems of the paper. It asserts that for the purposes of Szemerédi's theorem (and ignoring $o(1)$ errors), there is no distinction between a k-pseudorandom measure ν and the constant measure ν_{const}.

Theorem 3.5 (Szemerédi's theorem relative to a pseudorandom measure)**.** *Let $k \geqslant 3$ and $0 < \delta \leqslant 1$ be fixed parameters. Suppose that $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ is k-pseudorandom. Let $f : \mathbb{Z}_N \to \mathbb{R}^+$ be any non-negative function obeying the bound*

$$0 \leqslant f(x) \leqslant \nu(x) \text{ for all } x \in \mathbb{Z}_N \tag{3.4}$$

and

$$\mathbb{E}(f) \geqslant \delta. \tag{3.5}$$

Then we have

$$\mathbb{E}(f(x)f(x+r)\dots f(x+(k-1)r) | x, r \in \mathbb{Z}_N) \geqslant c(k,\delta) - o_{k,\delta}(1) \tag{3.6}$$

where $c(k,\delta) > 0$ is the same constant which appears in Theorem 2.3.

The proof of this theorem (which is combinatorial and ergodic-theoretic in nature, rather than number-theoretic or Fourier-analytic) will occupy the next few sections, §4–8. From §9 onwards we will apply this theorem to the specific case of the primes, by establishing a pseudorandom majorant for (a modified version of) the von Mangoldt function.

4. Notation

We now begin the proof of Theorem 3.5. Thoughout this proof we fix the parameter $k \geqslant 3$, the large integer N, and the probability density ν appearing in Theorem 3.5. All our constants in the $O()$ and $o()$ notation are allowed to depend on k (with all future dependence on this parameter being suppressed), and are also allowed to depend on the bounds implicit in the right-hand sides of (3.1) and (3.2). We may take N to be sufficiently large with respect to k and δ since (3.6) is trivial otherwise.

We need some standard L^q spaces.

Definition 4.1. For every $1 \leqslant q \leqslant \infty$ and $f : \mathbb{Z}_N \to \mathbb{R}$, we define the L^q norms as

$$\|f\|_{L^q} := \mathbb{E}(|f|^q)^{1/q}$$

with the usual convention that $\|f\|_{L^\infty} := \sup_{x \in \mathbb{Z}_N} |f(x)|$. We let $L^q(\mathbb{Z}_N)$ be the Banach space of all functions from $\mathbb{Z}_N$ to $\mathbb{R}$ equipped with the L^q norm; of course since $\mathbb{Z}_N$ is finite these spaces are all equal to each other as vector spaces, but the norms are only equivalent up to powers of N. We also observe that $L^2(\mathbb{Z}_N)$ is a real Hilbert space with the usual inner product

$$\langle f, g \rangle := \mathbb{E}(fg).$$

If Ω is a subset of $\mathbb{Z}_N$, we use $\mathbf{1}_\Omega : \mathbb{Z}_N \to \mathbb{R}$ to denote the indicator function of Ω, thus $\mathbf{1}_\Omega(x) = 1$ if $x \in \Omega$ and $\mathbf{1}_\Omega(x) = 0$ otherwise. Similarly if $P(x)$ is a statement concerning an element $x \in \mathbb{Z}_N$, we write $\mathbf{1}_{P(x)}$ for $\mathbf{1}_{\{x \in \mathbb{Z}_N : P(x)\}}(x)$.

In our arguments we shall frequently be performing linear changes of variables and then taking expectations. To facilitate this we adopt the following definition. Suppose that A and B are finite non-empty sets and that $\Phi : A \to B$ is a map. Then we say that Φ is a *uniform cover of B by A* if Φ is surjective and all the fibers $\{\Phi^{-1}(b) : b \in B\}$ have the same cardinality (i.e. they have cardinality $|A|/|B|$). Observe that if Φ is a uniform cover of B by A, then for any function $f : B \to \mathbb{R}$ we have

$$\mathbb{E}(f(\Phi(a))|a \in A) = \mathbb{E}(f(b)|b \in B). \tag{4.1}$$

5. Uniformity norms, and a generalized von Neumann theorem

As mentioned in earlier sections, the proof of Theorem 3.5 relies on splitting the given function f into a uniform component and an anti-uniform component. We will come to this splitting in later sections, but for this section we focus on defining the notion of uniformity, which is due to Gowers [15, 16]. The main result of this section will be a generalized von Neumann theorem (Proposition 5.3), which basically asserts that uniform functions are negligible for the purposes of computing sums such as (3.6).

Definition 5.1. Let $d \geqslant 0$ be a dimension[8]. We let $\{0,1\}^d$ be the standard discrete d-dimensional cube, consisting of d-tuples $\omega = (\omega_1, \ldots, \omega_d)$ where $\omega_j \in \{0,1\}$ for $j = 0, 1$. If $h = (h_1, \ldots, h_d) \in \mathbb{Z}_N^d$ we define $\omega \cdot h := \omega_1 h_1 + \ldots + \omega_d h_d$. If $(f_\omega)_{\omega \in \{0,1\}^d}$ is a $\{0,1\}^d$-tuple of functions in $L^\infty(\mathbb{Z}_N)$, we define the *d-dimensional Gowers inner product*

[8] In practice, we will have $d = k - 1$ where k is the length of the arithmetic progressions under consideration.

$\langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{U^d}$ by the formula

$$\langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{U^d} := \mathbb{E}\Big(\prod_{\omega \in \{0,1\}^d} f_\omega(x + \omega \cdot h) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^d \Big). \tag{5.1}$$

Henceforth we shall refer to a configuration $\{x + \omega \cdot h : \omega \in \{0,1\}^d\}$ as a *cube of dimension d*.

We recall from [16] the positivity properties of the Gowers inner product (5.1) when $d \geqslant 1$ (the $d = 0$ case being trivial). First suppose that f_ω does not depend on the final digit ω_d of ω, thus $f_\omega = f_{\omega_1,\ldots,\omega_{d-1}}$. Then we may rewrite (5.1) as

$$\langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{U^d} = \mathbb{E}\Big(\prod_{\omega' \in \{0,1\}^{d-1}} f_{\omega'}(x + \omega' \cdot h') f_{\omega'}(x + h_d + \omega' \cdot h') \Big| x \in \mathbb{Z}_N, h' \in \mathbb{Z}_N^{d-1}, h_d \in \mathbb{Z}_N \Big),$$

where we write $\omega' := (\omega_1, \ldots, \omega_{d-1})$ and $h' := (h_1, \ldots, h_{d-1})$. This can be rewritten further as

$$\langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{U^d} = \mathbb{E}\Big(\big|\mathbb{E}\big(\prod_{\omega' \in \{0,1\}^{d-1}} f_{\omega'}(y + \omega' \cdot h') \mid y \in \mathbb{Z}_N\big)\big|^2 \Big| h' \in \mathbb{Z}_N^{d-1} \Big), \tag{5.2}$$

so in particular we have the positivity property $\langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{U^d} \geqslant 0$ when f_ω is independent of ω_d. This proves the positivity property

$$\langle (f)_{\omega \in \{0,1\}^d} \rangle_{U^d} \geqslant 0 \tag{5.3}$$

when $d \geqslant 1$. We can thus define the *Gowers uniformity norm* $\|f\|_{U^d}$ of a function $f : \mathbb{Z}_N \to \mathbb{R}$ by the formula

$$\|f\|_{U^d} := \langle (f)_{\omega \in \{0,1\}^d} \rangle_{U^d}^{1/2^d} = \mathbb{E}\Big(\prod_{\omega \in \{0,1\}^d} f(x + \omega \cdot h) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^d \Big)^{1/2^d}. \tag{5.4}$$

When f_ω does depend on ω_d, (5.2) must be rewritten as

$$\langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{U^d} = \mathbb{E}\Big(\mathbb{E}\big(\prod_{\omega' \in \{0,1\}^{d-1}} f_{\omega',0}(y + \omega' \cdot h') \mid y \in \mathbb{Z}_N\big) \times \times \mathbb{E}\big(\prod_{\omega' \in \{0,1\}^{d-1}} f_{\omega',1}(y + \omega' \cdot h') \mid y \in \mathbb{Z}_N\big) \Big| h' \in \mathbb{Z}_N^{d-1} \Big).$$

From the Cauchy-Schwarz inequality in the h' variables, we thus see that

$$|\langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{U^d}| \leqslant \langle (f_{\omega',0})_{\omega \in \{0,1\}^d} \rangle_{U^d}^{1/2} \langle (f_{\omega',1})_{\omega \in \{0,1\}^d} \rangle_{U^d}^{1/2}.$$

Similarly if we replace the role of the ω_d digit by any of the other digits. Applying this Cauchy-Schwarz inequality once in each digit, we obtain the *Gowers Cauchy-Schwarz inequality*

$$|\langle (f_\omega)_{\omega \in \{0,1\}^d} \rangle_{U^d}| \leqslant \prod_{\omega \in \{0,1\}^d} \|f_\omega\|_{U^d}. \tag{5.5}$$

From the multilinearity of the inner product, and the binomial formula, we then obtain the inequality

$$|\langle (f+g)_{\omega\in\{0,1\}^d}\rangle_{U^d}| \leqslant (\|f\|_{U^d} + \|g\|_{U^d})^{2^d}$$

whence we obtain the *Gowers triangle inequality*

$$\|f+g\|_{U^d} \leqslant \|f\|_{U^d} + \|g\|_{U^d}.$$

(cf. [16] Lemmas 3.8 and 3.9).

The U^d norms are clearly homogeneous, so we have shown that they form a semi-norm. They are also non-decreasing in d; indeed, since

$$\|\nu_{\mathrm{const}}\|_{U^d} = \|1\|_{U^d} = 1, \tag{5.6}$$

we see from (5.5) that

$$|\langle (f_\omega)_{\omega\in\{0,1\}^d}\rangle_{U^d}| \leqslant \|f\|_{U^d}^{2^{d-1}}$$

where $f_\omega := 1$ when $\omega_d = 1$ and $f_\omega := f$ when $\omega_d = 0$. But the left-hand side can easily be computed to be $\|f\|_{U^{d-1}}^{2^{d-1}}$, and thus we have the monotonicity relation

$$\|f\|_{U^{d-1}} \leqslant \|f\|_{U^d} \tag{5.7}$$

for all $d \geqslant 2$.

The U^1 norm is not actually a norm, since one can compute from (5.4) that $\|f\|_{U^1} = |\mathbb{E}(f)|$ and thus $\|f\|_{U^1}$ may vanish without f itself vanishing. However, the U^2 norm (and hence all higher norms, by (5.7)) is non-degenerate. To see this we introduce the Fourier transform[9]

$$\widehat{f}(\xi) := \mathbb{E}(f(x)e^{-2\pi i x\xi/N} | x \in \mathbb{Z}_N)$$

for any $\xi \in \mathbb{Z}_N$. Some standard computations using the Fourier inversion formula $f(x) = \sum_{\xi\in\mathbb{Z}_N} \widehat{f}(\xi)e^{2\pi i x\xi/N}$ then show that

$$\|f\|_{U^2} = \Big(\sum_{\xi\in\mathbb{Z}_N} |\widehat{f}(\xi)|^4\Big)^{1/4}$$

(cf. [16], Lemma 2.2); by the Fourier inversion formula again we thus see that $\|f\|_{U^2}$ can only vanish if f vanishes identically[10].

We now show that pseudorandom measures ν are close to the constant measure ν_{const} in the U^d norms; this is of course consistent with our philosophy of deducing Theorem 3.5 from Theorem 2.3.

Lemma 5.2. *Suppose that ν is k-pseudorandom (as defined in Definition 3.3). Then we have*

$$\|\nu - \nu_{\mathrm{const}}\|_{U^d} = \|\nu - 1\|_{U^d} = o(1) \tag{5.8}$$

for all $1 \leqslant d \leqslant k-1$.

[9] This will be the only place in the argument where we actually use the Fourier transform on $\mathbb{Z}_N$; it of course plays a hugely important rôle in the $k = 3$ theory, and provides some very useful intuition to then think about the higher k theory, but will not be used elsewhere in this paper except as motivation.

[10] Another proof of this fact, which avoids the Fourier transform, is to test f against a delta mass, using (6.4) and the fact that the dual function of a delta mass is a constant multiple of that delta mass.

Proof. By (5.7) it suffices to prove the claim for $d = k-1$. Raising to the power 2^{k-1}, it suffices from (5.4) to show that

$$\mathbb{E}\Big(\prod_{\omega \in \{0,1\}^{k-1}} (\nu(x + \omega \cdot h) - 1) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1} \Big) = o(1).$$

The left-hand side can be expanded as

$$\sum_{A \subseteq \{0,1\}^{k-1}} (-1)^{|A|} \mathbb{E}\Big(\prod_{\omega \in A} \nu(x + \omega \cdot h) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1} \Big). \tag{5.9}$$

Let us look at the expression

$$\mathbb{E}\Big(\prod_{\omega \in A} \nu(x + \omega \cdot h) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1} \Big) \tag{5.10}$$

for some fixed $A \subseteq \{0,1\}^{k-1}$. This is of the form

$$\mathbb{E}\left(\nu(\psi_1(\mathbf{x})) \ldots \nu(\psi_{|A|}(\mathbf{x})) \mid \mathbf{x} \in \mathbb{Z}_N^k \right),$$

where $\mathbf{x} := (x, h_1, \ldots, h_{k-1})$ and the $\psi_1, \ldots, \psi_{|A|}$ are some ordering of the $|A|$ linear forms $x + \omega \cdot h$, $\omega \in A$. It is clear that none of these forms is a rational multiple of any other. Thus we may invoke the $(2^{k-1}, k, 1)$-linear forms condition, which is a consequence of the fact that ν is k-pseudorandom, to conclude that the expression (5.10) is $1 + o(1)$.

Referring back to (5.9), one sees that the claim now follows from the binomial theorem $\sum_{A \subseteq \{0,1\}^{k-1}} (-1)^{|A|} = (1-1)^{k-1} = 0$. $\square$

It is now time to state and prove our "generalised von Neumann theorem", which explains how the expression (3.6), which counts k-term arithmetic progressions, is governed by the Gowers uniformity norms. All of this, of course, is relative to a pseudorandom measure ν.

Proposition 5.3 (Generalized von Neumann)**.** *Suppose that ν is k-pseudorandom. Let $f_0, \ldots, f_{k-1} \in L^1(\mathbb{Z}_N)$ be functions which are pointwise bounded by $\nu + \nu_{\mathrm{const}}$, or in other words*

$$|f_j(x)| \leqslant \nu(x) + 1 \text{ for all } x \in \mathbb{Z}_N, 0 \leqslant j \leqslant k-1. \tag{5.11}$$

Let $c_0, \ldots, c_{k-1}$ be any distinct elements of $\mathbb{Z}_N$ (in practice we will take $c_j := j$). Then

$$\mathbb{E}\Big(\prod_{j=0}^{k-1} f_j(x + c_j r) \Big| x, r \in \mathbb{Z}_N \Big) = O\big(\inf_{0 \leqslant j \leqslant k-1} \|f_j\|_{U^{k-1}} \big) + o(1).$$

This Proposition is standard when $\nu = \nu_{\mathrm{const}}$ (see for instance [16, Theorem 3.2] or, for an analogous result in the ergodic setting, [10, Theorem 3.1]). The novelty is thus the extension to the pseudorandom ν studied in Theorem 3.5. The techniques here are inspired by similar Cauchy-Schwarz arguments relative to pseudo-random hypergraphs in [17].

Proof. By replacing ν with $(\nu + 1)/2$ (and by dividing f_j by 2), and using Lemma 3.4, we see that we may in fact assume without loss of generality that we can improve (5.11) to

$$|f_j(x)| \leqslant \nu(x) \text{ for all } x \in \mathbb{Z}_N, 0 \leqslant j \leqslant k-1. \tag{5.12}$$

For similar reasons we may assume that ν is strictly positive everywhere.

By rearranging the f_j and c_j if necessary, we may assume that the infimum

$$\inf_{0 \leqslant j \leqslant k-1} \|f_j\|_{U^{k-1}}$$

is attained when $j = 0$. By shifting x by $c_0 r$ if necessary we may assume that $c_0 = 0$. Our task is thus to show

$$\mathbb{E}\Big(\prod_{j=0}^{k-1} f_j(x + c_j r) \Big| x, r \in \mathbb{Z}_N \Big) = O\big(\|f_0\|_{U^{k-1}}\big) + o(1). \tag{5.13}$$

The proof of this will fall into two parts. First of all we will use the Cauchy-Schwarz inequality $k-1$ times (as is standard in the proof of theorems of this general type). In this way we will bound the left hand side of (5.13) by a *weighted* sum of f_0 over $(k-1)$-dimensional cubes. After that, we will show using the linear forms condition that these weights are roughly 1 on average, which will enable us to deduce (5.13).

We shall first need to set up some notation in order to apply Cauchy-Schwarz in a reasonably painless guise. Suppose that $0 \leqslant d \leqslant k-1$, and that we have two vectors $y = (y_1, \ldots, y_{k-1}) \in \mathbb{Z}_N^{k-1}$ and $y' = (y'_{k-d}, \ldots, y'_{k-1}) \in \mathbb{Z}_N^d$ of length $k-1$ and d respectively. For any set $S \subset \{k-d, \ldots, k-1\}$, we define the vector $y^{(S)} = (y_1^{(S)}, \ldots, y_{k-1}^{(S)}) \in \mathbb{Z}_N^{k-1}$ as

$$y_i^{(S)} := \begin{cases} y_i & \text{if } i \notin S \\ y'_i & \text{if } i \in S. \end{cases}$$

The set S thus indicates which components of $y^{(S)}$ come from y' rather than y.

Lemma 5.4 (Cauchy-Schwarz)**.** *Let $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ be any measure. Let $\phi_0, \phi_1, \ldots, \phi_{k-1} : \mathbb{Z}_N^{k-1} \to \mathbb{Z}_N$ be functions of $k-1$ variables y_i, such that ϕ_i does not depend on y_i for $1 \leqslant i \leqslant k-1$. Suppose that $f_0, f_1, \ldots, f_{k-1} \in L^1(\mathbb{Z}_N)$ are functions satisfying $|f_i(x)| \leqslant \nu(x)$ for all $x \in \mathbb{Z}_N$ and for each i, $0 \leqslant i \leqslant k-1$. For each $0 \leqslant d \leqslant k-1$ and $1 \leqslant i \leqslant k-1$, define the quantities*

$$J_d := \mathbb{E}\Big(\prod_{S \subseteq \{k-d,\ldots,k-1\}} \big(\prod_{i=0}^{k-d-1} f_i(\phi_i(y^{(S)}))\big)\big(\prod_{i=k-d}^{k-1} \nu^{1/2}(\phi_i(y^{(S)}))\big) \Big| y \in \mathbb{Z}_N^{k-1}, y' \in \mathbb{Z}_N^d \Big) \tag{5.14}$$

and

$$P_d := \mathbb{E}\Big(\prod_{S \subseteq \{k-d,\ldots,k-1\}} \nu(\phi_{k-d-1}(y^{(S)})) \Big| y \in \mathbb{Z}_N^{k-1}, y' \in \mathbb{Z}_N^d \Big). \tag{5.15}$$

Then for any $0 \leqslant d \leqslant k-2$, we have the inequality

$$|J_d|^2 \leqslant P_d J_{d+1}. \tag{5.16}$$

Remarks. The appearance of $\nu^{1/2}$ in (5.14) may seem odd. Note, however, that since ϕ_i does not depend on the i^{th} variable, each factor of $\nu^{1/2}$ in (5.14) occurs twice.

Proof of Lemma 5.4 Consider the quantity J_d. Since ϕ_{k-d-1} does not depend on y_{k-d-1}, we may take all quantities depending on ϕ_{k-d-1} outside of the y_{k-d-1} average. This allows us to write

$$J_d = \mathbb{E}\big(G(y, y') H(y, y') \bigm| y_1, \ldots, y_{k-d-2}, y_{k-d}, \ldots, y_{k-1}, y'_{k-d}, \ldots, y'_{k-d} \in \mathbb{Z}_N\big),$$

where

$$G(y,y') := \prod_{S\subseteq\{k-d,\ldots,k-1\}} f_{k-d-1}(\phi_{k-d-1}(y^{(S)}))\nu^{-1/2}(\phi_{k-d-1}(y^{(S)}))$$

and

$$H(y,y') := \mathbb{E}\Big(\prod_{S\subseteq\{k-d,\ldots,k-1\}} \prod_{i=0}^{k-d-2} f_i(\phi_i(y^{(S)})) \prod_{i=k-d-1}^{k-1} \nu^{1/2}(\phi_i(y^{(S)})) \;\Big|\; y_{k-d-1}\in\mathbb{Z}_N\Big)$$

(note we have multiplied and divided by several factors of the form $\nu^{1/2}(\phi_{k-d-1}(y^{(S)}))$. Now apply Cauchy-Schwarz to give

$$|J_d|^2 \leqslant \mathbb{E}\big(|G(y,y')|^2 \;\big|\; y_1,\ldots,y_{k-d-2},y_{k-d},\ldots,y_{k-1},y'_{k-d},\ldots,y'_{k-1}\in\mathbb{Z}_N\big)\times$$
$$\times\mathbb{E}\big(|H(y,y')|^2 \;\big|\; y_1,\ldots,y_{k-d-2},y_{k-d},\ldots,y_{k-1},y'_{k-d},\ldots,y'_{k-1}\in\mathbb{Z}_N\big).$$

Since $|f_{k-d-1}(x)| \leqslant \nu(x)$ for all x, one sees from (5.15) that

$$\mathbb{E}\big(|G(y,y')|^2 \;\big|\; y_1,\ldots,y_{k-d-2},y_{k-d},\ldots,y_{k-1},y'_{k-d},\ldots,y'_{k-1}\in\mathbb{Z}_N\big) \leqslant P_d$$

(note that the y_{k-d-1} averaging in (5.15) is redundant since ϕ_{k-d-1} does not depend on this variable). Moreover, by writing in the definition of $H(y,y')$ and expanding out the square, replacing the averaging variable y_{k-d-1} with the new variables y_{k-d-1}, y'_{k-d-1}, one sees from (5.14) that

$$\mathbb{E}\big(|H(y,y')|^2 \;\big|\; y_1,\ldots,y_{k-d-2},y_{k-d},\ldots,y_{k-1},y'_{k-d},\ldots,y'_{k-1}\in\mathbb{Z}_N\big) = J_{d+1}.$$

The claim follows. □

Applying the above Lemma $k-1$ times, we obtain in particular that

$$|J_0|^{2^{k-1}} \leqslant J_{k-1}\prod_{d=0}^{k-2} P_d^{2^{k-2-d}}. \tag{5.17}$$

Observe from (5.14) that

$$J_0 = \mathbb{E}\Big(\prod_{i=0}^{k-1} f_i(\phi_i(y)) \;\Big|\; y\in\mathbb{Z}_N^{k-1}\Big) \tag{5.18}$$

Remark. There is a close connection between the inequality (5.17) and [17, Lemma 7.1].

Proof of Proposition 5.3. We will apply (5.17), observing that (5.18) can be used to count configurations $(x, x+c_1r,\ldots,x+c_{k-1}r)$ by making a judicious choice of the functions ϕ_i. Take

$$\phi_i(y) := \sum_{j=1}^{k-1}\Big(1-\frac{c_i}{c_j}\Big)y_j$$

for $i=0,\ldots,k-1$. Then $\phi_0(y) = y_1+\cdots+y_{k-1}$, $\phi_i(y)$ does not depend on y_i and, as one can easily check, for any y we have $\phi_i(y) = x + c_i r$ where

$$r = -\sum_{i=1}^{k-1}\frac{y_i}{c_i}.$$

Now the map $\Phi : \mathbb{Z}_N^{k-1} \to \mathbb{Z}_N^2$ defined by

$$\Phi(y) := (y_1 + \cdots + y_{k-1}, \frac{y_1}{c_1} + \frac{y_2}{c_2} + \cdots + \frac{y_{k-1}}{c_{k-1}})$$

is a uniform cover, and so

$$\mathbb{E}\Big(\prod_{j=0}^{k-1} f_j(x + c_j r) \Big| x, r \in \mathbb{Z}_N \Big) = \mathbb{E}\Big(\prod_{i=0}^{k-1} f_i(\phi_i(y)) \Big| y \in \mathbb{Z}_N^{k-1} \Big) = J_0 \tag{5.19}$$

thanks to (5.18). On the other hand we have $P_d = 1 + o(1)$ for each $0 \leqslant d \leqslant k-2$, since the k-pseudorandom hypothesis on ν implies the $(2^d, k-1+d, k)$-linear forms condition. Applying (5.17) we thus obtain

$$J_0^{2^{k-1}} \leqslant (1 + o(1)) J_{k-1}. \tag{5.20}$$

Fix y. As S ranges over all subsets of $\{1, \ldots, k-1\}$, $\phi_0(y^{(S)})$ ranges over a $(k-1)$-dimensional cube $\{x + \omega \cdot h : \omega \in \{0,1\}^{k-1}\}$ where $x = y_1 + \cdots + y_{k-1}$ and $h_i = y_i' - y_i$, $i = 1, \ldots, k-1$. Thus we may write

$$J_{k-1} = \mathbb{E}\Big(W(x,h) \prod_{\omega \in \{0,1\}^{k-1}} f_0(x + \omega \cdot h) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1} \Big) \tag{5.21}$$

where the weight function $W(x,h)$ is given by

$$\begin{aligned} W(x,h) &= \mathbb{E}\Big(\prod_{\omega \in \{0,1\}^{k-1}} \prod_{i=1}^{k-2} \nu^{1/2}(\phi_i(y + \omega h)) \times \\ &\qquad \times \nu^{1/2}(\phi_{k-1}(y + \omega h)) \Big| y_1, \ldots, y_{k-2} \in \mathbb{Z}_N \Big) \\ &= \mathbb{E}\Big(\prod_{i=1}^{k-2} \prod_{\substack{\omega \in \{0,1\}^{k-1} \\ \omega_i = 0}} \nu(\phi_i(y + \omega h)) \times \\ &\qquad \times \prod_{\substack{\omega \in \{0,1\}^{k-1} \\ \omega_{k-1} = 0}} \nu(\phi_{k-1}(y + \omega h)) \Big| y_1, \ldots, y_{k-2} \in \mathbb{Z}_N \Big). \end{aligned}$$

Here, $\omega h \in \mathbb{Z}_N^{k-1}$ is the vector with components $(\omega h)_j := \omega_j h_j$ for $1 \leqslant j \leqslant k-1$, and $y \in \mathbb{Z}_N^{k-1}$ is the vector with components y_j for $1 \leqslant j \leqslant k-2$ and $y_{k-1} := x - y_1 - \ldots - y_{k-2}$. Now by the definition of the U^{k-1} norm we have

$$\mathbb{E}\Big(\prod_{\omega \in \{0,1\}^{k-1}} f_0(x + \omega \cdot h) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1} \Big) = \|f_0\|_{U^{k-1}}^{2^{k-1}}.$$

To prove (5.13) it therefore suffices, by (5.19), (5.20) and (5.21), to prove that

$$\mathbb{E}\Big((W(x,h) - 1) \prod_{\omega \in \{0,1\}^{k-1}} f_0(x + \omega \cdot h) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1} \Big) = o(1).$$

Using (5.12), it suffices to show that

$$\mathbb{E}\Big(|W(x,h)-1| \prod_{\omega \in \{0,1\}^{k-1}} \nu(x+\omega\cdot h) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1} \Big) = o(1).$$

From Lemma 5.2 we have

$$\mathbb{E}\Big(\prod_{\omega \in \{0,1\}^{k-1}} \nu(x+\omega\cdot h) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1} \Big) = O(1)$$

so by Cauchy-Schwarz it will suffice to prove

Lemma 5.5 (ν covers its own cubes uniformly)**.** *We have*

$$\mathbb{E}\Big(|W(x,h)-1|^2 \prod_{\omega \in \{0,1\}^{k-1}} \nu(x+\omega\cdot h) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1} \Big) = o(1).$$

Proof. Expanding out the square, it then suffices to show that

$$\mathbb{E}\Big(W(x,h)^q \prod_{\omega \in \{0,1\}^{k-1}} \nu(x+\omega\cdot h) \Big| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1} \Big) = 1+o(1)$$

for $q = 0, 1, 2$. This can be achieved by three applications of the linear forms condition, as follows:

$q = 0$. Use the $(2^{k-1}, k, 1)$-linear forms property with variables $x, h_1, \dots, h_{k-1}$ and forms

$$x + \omega \cdot h \qquad \omega \in \{0,1\}^{k-1}.$$

$q = 1$. Use the $(2^{k-2}(k+1), 2k-2, k)$-linear forms property with variables x, $h_1, \dots, h_{k-1}$, $y_1, \dots, y_{k-2}$ and forms

$$\begin{aligned} \phi_i(y+\omega h) &\qquad \omega \in \{0,1\}^{k-1},\ \ \omega_i = 0, 1 \leqslant i \leqslant k-2; \\ \phi_{k-1}(y+\omega h) &\qquad \omega \in \{0,1\}^{k-1},\ \ \omega_{k-1} = 0; \\ x+\omega\cdot h, &\qquad \omega \in \{0,1\}^{k-1}. \end{aligned}$$

$q = 2$. Use the $(k \cdot 2^{k-1}, 3k-4, k)$-linear forms property with variables x, $h_1, \dots, h_{k-1}$, $y_1, \dots, y_{k-2}$, $y'_1, \dots, y'_{k-2}$ and forms

$$\begin{aligned} \phi_i(y+\omega h) &\qquad \omega \in \{0,1\}^{k-1},\ \ \omega_i = 0, 1 \leqslant i \leqslant k-2; \\ \phi_i(y'+\omega h) &\qquad \omega \in \{0,1\}^{k-1},\ \ \omega_i = 0, 1 \leqslant i \leqslant k-2; \\ \phi_{k-1}(y+\omega h) &\qquad \omega \in \{0,1\}^{k-1},\ \ \omega_{k-1} = 0; \\ \phi_{k-1}(y'+\omega h) &\qquad \omega \in \{0,1\}^{k-1},\ \ \omega_{k-1} = 0; \\ x+\omega\cdot h, &\qquad \omega \in \{0,1\}^{k-1}. \end{aligned}$$

Here of course we adopt the convention that $y'_{k-1} = x - y'_1 - \dots - y'_{k-2}$. This completes the proof of the Lemma, and hence of Proposition 5.3. □

6. Anti-uniformity

Having studied the U^{k-1} norm, we now introduce the dual $(U^{k-1})^*$ norm, defined in the usual manner as

$$\|g\|_{(U^{k-1})^*} := \sup\{|\langle f, g\rangle| : f \in U^{k-1}(\mathbb{Z}_N); \|f\|_{U^{k-1}} \leqslant 1\}. \tag{6.1}$$

We say that g is *anti-uniform* if $\|g\|_{(U^{k-1})^*} = O(1)$ and $\|g\|_{L^\infty} = O(1)$. If g is anti-uniform, and if $|\langle f, g\rangle|$ is large, then f cannot be uniform since

$$|\langle f, g\rangle| \leqslant \|f\|_{U^{k-1}}\|g\|_{(U^{k-1})^*}.$$

Thus anti-uniform functions can be thought of as "obstructions to uniformity". The $(U^{k-1})^*$ are well-defined norms for $k \geqslant 3$ since U^{k-1} is then a genuine norm (not just a seminorm). In this section we show how to generate a large class of anti-uniform functions, in order that we can decompose an arbitrary function f into a uniform part and a bounded anti-uniform part in the next section.

Remark. In the $k = 3$ case we have the explicit formula

$$\|g\|_{(U^2)^*} = \Big(\sum_{\xi\in\mathbb{Z}_N} |\widehat{g}(\xi)|^{4/3}\Big)^{3/4}.$$

We will not, however, require this fact.

A basic way to generate anti-uniform functions is the following. For each function $F \in L^1(\mathbb{Z}_N)$, define the *dual function* $\mathcal{D}F$ of F by

$$\mathcal{D}F(x) := \mathbb{E}\Bigg(\prod_{\omega\in\{0,1\}^{k-1}:\omega\neq 0} F(x + \omega\cdot h)\ \Bigg|\ h \in \mathbb{Z}_N^{k-1}\Bigg). \tag{6.2}$$

Remark. Such functions have arisen recently in work of Host and Kra [25] in the ergodic theory setting (see also [1]).

The next lemma, while simple, is fundamental to our entire approach; it asserts that if a function majorised by a pseudorandom measure ν is not uniform, then it correlates[11] with a bounded anti-uniform function. Boundedness is the key feature here. The idea in proving Theorem 3.5 will then be to project out the influence of these bounded anti-uniform functions (through the machinery of conditional expectation) until one is only left with a uniform remainder, which can be discarded by the generalized von Neumann theorem (Proposition 5.3).

Lemma 6.1 (Non-uniformity implies correlation)**.** *Let ν be a k-pseudorandom measure, and let $F \in L^1(\mathbb{Z}_N)$ be any function obeying the bounds*

$$|F(x)| \leqslant \nu(x) + 1 \text{ for all } x \in \mathbb{Z}_N.$$

Then we have the identities

$$\langle F, \mathcal{D}F\rangle = \|F\|_{U^{k-1}}^{2^{k-1}} \tag{6.3}$$

and

$$\|\mathcal{D}F\|_{(U^{k-1})^*} = \|F\|_{U^{k-1}}^{2^{k-1}-1} \tag{6.4}$$

[11] This idea was inspired by the proof of the Furstenberg structure theorem [9], [10]; a key point in that proof being that if a system is not (relatively) weakly mixing, then it must contain a non-trivial (relatively) almost periodic function, which can then be projected out via conditional expectation.

as well as the estimate

$$\|\mathcal{D}F\|_{L^\infty} \leqslant 2^{2^{k-1}-1} + o(1). \tag{6.5}$$

Proof. The identity (6.3) is clear just by expanding out both sides using (6.2), (5.4). To prove (6.4) we may of course assume F is not identically zero. By (6.1) and (6.3) it suffices to show that

$$|\langle f, \mathcal{D}F\rangle| \leqslant \|f\|_{U^{k-1}} \|F\|_{U^{k-1}}^{2^{k-1}-1}$$

for arbitrary functions f. But by (6.2) the left-hand side is simply the Gowers inner product $\langle (f_\omega)_{\omega\in\{0,1\}^{k-1}}\rangle_{U^{k-1}}$, where $f_\omega := f$ when $\omega = 0$ and $f_\omega := F$ otherwise. The claim then follows from the Gowers Cauchy-Schwarz inequality (5.5).

Finally, we observe that (6.5) is a consequence of the linear forms condition. Bounding F by $2(\nu+1)/2 = 2\nu_{1/2}$, it suffices to show that

$$\mathcal{D}\nu_{1/2}(x) \leqslant 1 + o(1)$$

uniformly in the choice of $x \in \mathbb{Z}_N$. The left-hand side can be expanded using (6.2) as

$$\mathbb{E}\Big(\prod_{\omega\in\{0,1\}^{k-1}:\omega\neq 0} \nu_{1/2}(x + \omega\cdot h) \Big| h \in \mathbb{Z}_N^{k-1} \Big).$$

By the linear forms condition (3.1) (and Lemma 3.4) this expression is $1 + o(1)$ (this is the only place in the paper that we appeal to the linear forms condition in the non-homogeneous case where some $b_i \neq 0$. Here, all the b_i are x). □

Remarks. Observe that if $P : \mathbb{Z}_N \to \mathbb{Z}_N$ is any polynomial on $\mathbb{Z}_N$ of degree at most $k-2$, and $F(x) = e^{2\pi i P(x)/N}$, then[12] $\mathcal{D}F = F$ (this is basically a reflection of the fact that taking $k-1$ successive differences of P yields the zero function). Hence by the above lemma $\|F\|_{(U^{k-1})^*} \leqslant 1$, and thus F is anti-uniform. One should keep these "polynomially quasiperiodic" functions $e^{2\pi i P(x)/N}$ in mind as model examples of functions of the form $\mathcal{D}F$, whilst bearing in mind that they are not the only examples[13]. For some further discussion on the role of such polynomials of degree $k-2$ in determining uniformity especially in the $k = 4$ case, see [15, 16]. Very roughly speaking, uniform functions are analogous to the notion of "weakly mixing" functions that appear in ergodic theory proofs of Szemerédi's theorem, whereas anti-uniform functions are somewhat analogous to the notion of "almost periodic" functions. When $k = 3$ there is a more precise relation with linear exponentials (which are the same thing as characters on $\mathbb{Z}_N$). When $\nu = 1$, for example, one has the explicit formula

$$\mathcal{D}F(x) = \sum_{\xi\in\mathbb{Z}_N} |\widehat{F}(\xi)|^2 \widehat{F}(\xi) e^{2\pi i x\xi/N}.$$

[12] To make this assertion precise, one has to generalize the notion of dual function to complex-valued functions by inserting an alternating sequence of conjugation signs; see [16].

[13] The situation again has an intriguing parallel with ergodic theory, in which the rôle of the anti-uniform functions of order $k-2$ appear to be played by $k-2$-step nilfactors (see [25], [26], [39]), which may contain polynomial eigenfunctions of order $k-2$, but can also exhibit slightly more general behaviour; see [11] for further discussion.

By splitting the set of frequencies $\mathbb{Z}_N$ into the sets $S := \{\gamma : |\widehat{F}(\xi)| \geqslant \epsilon\}$ and $\mathbb{Z}_N \backslash S$ one sees that it is possible to write

$$\mathcal{D}F(x) = \sum_{\xi \in S} a_\xi e^{2\pi i x\xi/N} + E(x),$$

where $|a_\xi| \leqslant 2$ and $\|E\|_{L^\infty} \leqslant 4\epsilon$. Also, we have $|S| \leqslant 4\epsilon^{-2}$. Thus $\mathcal{D}(F)$ is equal to a linear combination of a few characters plus a small error.

Once again, these remarks concerning the relation with harmonic analysis are included only for motivational purposes.

Let us refer to functions of the form $\mathcal{D}F$, where F is pointwise bounded by $\nu + 1$, as a *basic anti-uniform function.* Observe from (6.5) that if N is sufficiently large, then all basic uniform functions take values in the interval $I := [-2^{2^{k-1}}, 2^{2^{k-1}}]$.

The following is a statement to the effect that the measure ν is uniformly distributed with respect not just to each basic anti-uniform function (which is a special case of (6.4)), but also to the *algebra* generated by such functions.

Proposition 6.2 (Uniform distribution wrt to basic anti-uniform functions)**.** *Suppose that ν is k-pseudorandom. Let $K \geqslant 1$ be a fixed integer, let $\Phi : I^K \to \mathbb{R}$ be a fixed continuous function, let $\mathcal{D}F_1, \ldots, \mathcal{D}F_K$ be basic anti-uniform functions, and define the function $\psi : \mathbb{Z}_N \to \mathbb{R}$ by*

$$\psi(x) := \Phi(\mathcal{D}F_1(x), \ldots, \mathcal{D}F_K(x)).$$

Then we have the estimate

$$\langle \nu - 1, \psi \rangle = o_{K,\Phi}(1).$$

Furthermore if Φ ranges over a compact set $E \subset C^0(I^K)$ of the space $C^0(I^K)$ of continuous functions on I^K (in the uniform topology) then the bounds here are uniform in Φ (i.e. one can replace $o_{K,\Phi}(1)$ with $o_{K,E}(1)$ in this case).

Remark. In light of the previous remarks, we see in particular that ν is uniformly distributed with respect to any continuous function of polynomial phase functions such as $e^{2\pi i P(x)/N}$, where P has degree at most $k - 2$.

Proof. We will prove this result in two stages, first establishing the result for Φ a polynomial and then using an approximation argument to deduce the general case. Fix $K \geqslant 1$, and let $F_1, \ldots F_K \in L^1(\mathbb{Z}_N)$ be fixed functions obeying the bounds

$$F_j(x) \leqslant \nu(x) + 1 \text{ for all } x \in \mathbb{Z}_N, 1 \leqslant j \leqslant K.$$

By replacing ν by $(\nu + 1)/2$, dividing the F_j by two, and using Lemma 3.4 as before, we may strengthen this bound without loss of generality to

$$|F_j(x)| \leqslant \nu(x) \text{ for all } x \in \mathbb{Z}_N, 1 \leqslant j \leqslant K. \tag{6.6}$$

Lemma 6.3. *Let $d \geqslant 1$. For any polynomial P of K variables and degree d with real coefficients (independent of N), we have*

$$\|P(\mathcal{D}F_1, \ldots, \mathcal{D}F_K)\|_{(U^{k-1})^*} = O_{K,d,P}(1).$$

Remark. It may seem surprising that that there is no size restriction on K or d, since we are presumably going to use the linear forms condition and we are only assuming that condition with bounded parameters. However whilst we do indeed restrict the size of m in (3.3), we do not need to restrict the size of q in (3.2).

Proof. By linearity it suffices to prove this when P is a monomial. By enlarging K to dK and repeating the functions F_j as necessary, it in fact suffices to prove this for the monomial $P(x_1, \ldots, x_K) = x_1 \ldots x_K$. Recalling the definition of $(U^{k-1})^*$, we are thus required to show that

$$\big\langle f, \prod_{j=1}^K \mathcal{D}F_j \big\rangle = O_K(1)$$

for all $f : \mathbb{Z}_N \to \mathbb{R}$ satisfying $\|f\|_{U^{k-1}} \leqslant 1$. By (6.2) the left-hand side can be expanded as

$$\mathbb{E}\Big(f(x) \prod_{j=1}^K \mathbb{E}\big(\prod_{\omega \in \{0,1\}^{k-1} : \omega \neq 0} F_j(x + \omega \cdot h^{(j)}) \bigm| h^{(j)} \in \mathbb{Z}_N^{k-1} \big) \Bigm| x \in \mathbb{Z}_N \Big).$$

We can make the change of variables $h^{(j)} = h + H^{(j)}$ for any $h \in \mathbb{Z}_N^{k-1}$, and then average over h, to rewrite this as

$$\mathbb{E}\Big(f(x) \prod_{j=1}^K \mathbb{E}\big(\prod_{\omega \in \{0,1\}^{k-1} : \omega \neq 0} F_j(x + \omega \cdot H^{(j)} + \omega \cdot h) \bigm| H^{(j)} \in \mathbb{Z}_N^{k-1} \big) \Bigm| x \in \mathbb{Z}_N; h \in \mathbb{Z}_N^{k-1} \Big).$$

Expanding the j product and interchanging the expectations, we can rewrite this in terms of the Gowers inner product as

$$\mathbb{E}\big(\langle (f_{\omega,H})_{\omega \in \{0,1\}^{k-1}} \rangle_{U^{k-1}} \bigm| H \in (\mathbb{Z}_N^{k-1})^K \big)$$

where $H := (H^{(1)}, \ldots, H^{(K)})$, $f_{0,H} := f$, and $f_{\omega,H} := g_{\omega \cdot H}$ for $\omega \neq 0$, where $\omega \cdot H := (\omega \cdot H^{(1)}, \ldots, \omega \cdot H^{(K)})$ and

$$g_{u^{(1)},\ldots,u^{(K)}}(x) := \prod_{j=1}^K F_j(x + u^{(j)}) \text{ for all } u^{(1)}, \ldots, u^{(K)} \in \mathbb{Z}_N. \tag{6.7}$$

By the Gowers-Cauchy-Schwarz inequality (5.5) we can bound this as

$$\mathbb{E}\Big(\|f\|_{U^{k-1}} \prod_{\omega \in \{0,1\}^{k-1} : \omega \neq 0} \|g_{\omega \cdot H}\|_{U^{k-1}} \Bigm| H \in (\mathbb{Z}_N^{k-1})^K \Big)$$

so to prove the claim it will suffice to show that

$$\mathbb{E}\Big(\prod_{\omega \in \{0,1\}^{k-1} : \omega \neq 0} \|g_{\omega \cdot H}\|_{U^{k-1}} \Bigm| H \in (\mathbb{Z}_N^{k-1})^K \Big) = O_K(1).$$

By Hölder's inequality it will suffice to show that

$$\mathbb{E}\big(\|g_{\omega \cdot H}\|_{U^{k-1}}^{2^{k-1}-1} \bigm| H \in (\mathbb{Z}_N^{k-1})^K \big) = O_K(1)$$

for each $\omega \in \{0,1\}^{k-1} \backslash 0$.

Fix ω. Since $2^{k-1}-1 \leqslant 2^{k-1}$, another application of Hölder's inequality shows that it in fact suffices to show that

$$\mathbb{E}\big(\|g_{\omega\cdot H}\|_{U^{k-1}}^{2^{k-1}} \bigm| H \in (\mathbb{Z}_N^{k-1})^K\big) = O_K(1).$$

Since ω is non-zero, the map $H \mapsto \omega \cdot H$ is an uniform covering of $\mathbb{Z}_N^K$ by $(\mathbb{Z}_N^{k-1})^K$. Thus by (4.1) we can rewrite the left-hand side as

$$\mathbb{E}\big(\|g_{u^{(1)},\dots,u^{(K)}}\|_{U^{k-1}}^{2^{k-1}} \bigm| u^{(1)},\dots,u^{(K)} \in \mathbb{Z}_N\big).$$

Expanding this out using (5.4) and (6.7), we can rewrite the left-hand side as

$$\mathbb{E}\Big(\prod_{\tilde{\omega}\in\{0,1\}^{k-1}}\prod_{j=1}^{K} F_j(x+u^{(j)}+h\cdot\tilde{\omega}) \Bigm| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1}, u^{(1)},\dots,u^{(K)} \in \mathbb{Z}_N\Big).$$

This factorizes as

$$\mathbb{E}\Big(\prod_{j=1}^{K}\mathbb{E}\big(\prod_{\tilde{\omega}\in\{0,1\}^{k-1}} F_j(x+u^{(j)}+h\cdot\tilde{\omega}) \bigm| u^{(j)} \in \mathbb{Z}_N\big) \Bigm| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1}\Big).$$

Applying (6.6), we reduce to showing that

$$\mathbb{E}\Big(\mathbb{E}\big(\prod_{\tilde{\omega}\in\{0,1\}^{k-1}} \nu(x+u+h\cdot\tilde{\omega}) \bigm| u \in \mathbb{Z}_N\big)^K \Bigm| x \in \mathbb{Z}_N, h \in \mathbb{Z}_N^{k-1}\Big) = O_K(1).$$

We can make the change of variables $y := x+u$, and then discard the redundant x averaging, to reduce to showing that

$$\mathbb{E}\Big(\mathbb{E}\big(\prod_{\tilde{\omega}\in\{0,1\}^{k-1}} \nu(y+h\cdot\tilde{\omega}) \bigm| y \in \mathbb{Z}_N\big)^K \Bigm| h \in \mathbb{Z}_N^{k-1}\Big) = O_K(1).$$

Now we are ready to apply the correlation condition (Definition 3.2). This is, in fact, the only time we will use that condition. It gives

$$\mathbb{E}\Big(\prod_{\tilde{\omega}\in\{0,1\}^{k-1}} \nu(y+h\cdot\tilde{\omega}) \Bigm| y \in \mathbb{Z}_N\Big) \leqslant \sum_{\tilde{\omega},\tilde{\omega}'\in\{0,1\}^{k-1}:\tilde{\omega}\neq\tilde{\omega}'} \tau(h\cdot(\tilde{\omega}-\tilde{\omega}'))$$

where, recall, τ is a weight function satisfying $\mathbb{E}(\tau^q) = O_q(1)$ for all q. Applying the triangle inequality in $L^K(\mathbb{Z}_N)$, it thus suffices to show that

$$\mathbb{E}\big(\tau(h\cdot(\tilde{\omega}-\tilde{\omega}'))^K \bigm| h \in (\mathbb{Z}_N)^{k-1}\big) = O_K(1)$$

for all distinct $\tilde{\omega},\tilde{\omega}' \in \{0,1\}^{k-1}$. But the map $h \mapsto h\cdot(\tilde{\omega}-\tilde{\omega}')$ is a uniform covering of $\mathbb{Z}_N$ by $(\mathbb{Z}_N)^{k-1}$, so by (4.1) the left-hand side is just $\mathbb{E}(\tau^K)$, which is $O_K(1)$. □

Proof of Proposition 6.2. Let Φ, ψ be as in the Proposition, and let $\varepsilon > 0$ be arbitrary. From (6.5) we know that the basic anti-uniform functions $\mathcal{D}F_1,\dots,\mathcal{D}F_K$ take values in the compact interval $I := [-2^{2^{k-1}}, 2^{2^{k-1}}]$ introduced earlier. By the Weierstrass approximation theorem, we can thus find a polynomial P (depending only on K and ε) such that

$$\|\Phi(\mathcal{D}F_1,\dots,\mathcal{D}F_K) - P(\mathcal{D}F_1,\dots,\mathcal{D}F_K)\|_{L^\infty} \leqslant \varepsilon$$

and so (since $\mathbb{E}(\nu) = 1 + o(1)$)

$$|\langle \nu-1, \Phi(\mathcal{D}F_1,\dots,\mathcal{D}F_K) - P(\mathcal{D}F_1,\dots,\mathcal{D}F_K)\rangle| \leqslant (2+o(1))\varepsilon.$$

On the other hand, from Lemma 6.3 and (6.1) we have

$$\langle \nu - 1, P(\mathcal{D}F_1, \ldots, \mathcal{D}F_K)\rangle = o_{K,\varepsilon}(1)$$

since P depends on K and ε. Combining the two estimates we thus see that for N sufficiently large (depending on K and ε) we have

$$|\langle \nu - 1, \Phi(\mathcal{D}F_1, \ldots, \mathcal{D}F_K)\rangle| \leqslant 4\varepsilon$$

(for instance). Since $\varepsilon > 0$ was arbitrary, the claim follows. It is clear that this argument also gives the uniform bounds when Φ ranges over a compact set (by covering this compact set by finitely many balls of radius ε in the uniform topology). □

7. Generalized Bohr sets and σ-algebras

To use Proposition 6.2, we shall associate a σ-algebra to each basic anti-uniform function, such that the measurable functions in each such algebra can be approximated by a function of the type considered in Proposition 6.2. We begin by setting out our notation for σ-algebras.

Definition 7.1. A *σ-algebra* $\mathcal{B}$ in $\mathbb{Z}_N$ is any collection of subsets of $\mathbb{Z}_N$ which contains the empty set $\emptyset$ and the full set $\mathbb{Z}_N$, and is closed under complementation, unions and intersections. As $\mathbb{Z}_N$ is a finite set, we will not need to distinguish between countable and uncountable unions or intersections. We define the *atoms* of a σ-algebra to be the minimal non-empty elements of $\mathcal{B}$ (with respect to set inclusion); it is clear that the atoms in $\mathcal{B}$ form a partition of $\mathbb{Z}_N$, and $\mathcal{B}$ consists precisely of arbitrary unions of its atoms (including the empty union $\emptyset$). A function $f \in L^q(\mathbb{Z}_N)$ is said to be *measurable* with respect to a σ-algebra $\mathcal{B}$ if all the level sets $\{f^{-1}(\{x\}) : x \in \mathbb{R}\}$ of f lie in $\mathcal{B}$, or equivalently if f is constant on each of the atoms of $\mathcal{B}$.

We define $L^q(\mathcal{B}) \subseteq L^q(\mathbb{Z}_N)$ to be the subspace of $L^q(\mathbb{Z}_N)$ consisting of $\mathcal{B}$-measurable functions, equipped with the same L^q norm. We can then define the conditional expectation operator $f \mapsto \mathbb{E}(f|\mathcal{B})$ to be the orthogonal projection of $L^2(\mathbb{Z}_N)$ to $L^2(\mathcal{B})$; this is of course also defined on all the other $L^q(\mathbb{Z}_N)$ spaces since they are all the same vector space. An equivalent definition of conditional expectation is

$$\mathbb{E}(f|\mathcal{B})(x) := \mathbb{E}(f(y)|y \in A(x))$$

for all $x \in \mathbb{Z}_N$, where $A(x)$ is the unique atom in $\mathcal{B}$ which contains x. It is clear that conditional expectation is a linear self-adjoint orthogonal projection on $L^2(\mathbb{Z}_N)$, is a contraction on $L^q(\mathbb{Z}_N)$ for every $1 \leqslant q \leqslant \infty$, preserves non-negativity, and also preserves constant functions. Also, if $\mathcal{B}'$ is a subalgebra of $\mathcal{B}$ then $\mathbb{E}(\mathbb{E}(f|\mathcal{B})|\mathcal{B}') = \mathbb{E}(f|\mathcal{B}')$.

If $\mathcal{B}_1, \ldots, \mathcal{B}_K$ are σ-algebras, we use $\mathcal{B}_1 \vee \ldots \vee \mathcal{B}_K$ to denote the σ-algebra generated by these algebras, or in other words the algebra whose atoms are the intersections of atoms in $\mathcal{B}_1, \ldots, \mathcal{B}_K$.

We now construct the basic σ-algebras that we shall use. We view the basic anti-uniform functions as generalizations of complex exponentials, and the atoms of the σ-algebras we use can be thought of as "generalized Bohr sets".

Proposition 7.2 (Each function generates a σ-algebra)**.** *Let ν be a k-pseudorandom measure, let $0 < \varepsilon < 1$ and $0 < \sigma < 1/2$ be parameters, and let $G \in L^\infty(\mathbb{Z}_N)$ be function* *taking values in the interval* $I := [-2^{2^{k-1}}, 2^{2^{k-1}}]$*. Then there exists a* σ*-algebra* $\mathcal{B}_{\varepsilon,\sigma}(G)$ *with the following properties:*

- (G lies in its own σ-algebra) *For any* σ*-algebra* $\mathcal{B}$*, we have*
$$\|G - \mathbb{E}(G|\mathcal{B} \vee \mathcal{B}_{\varepsilon,\sigma}(G))\|_{L^\infty(\mathbb{Z}_N)} \leqslant \varepsilon. \quad (7.1)$$
- (Bounded complexity) $\mathcal{B}_{\varepsilon,\sigma}(G)$ *is generated by at most* $O(1/\varepsilon)$ *atoms.*
- (Approximation by continuous functions of G) *If* A *is any atom in* $\mathcal{B}_{\varepsilon,\sigma}(G)$*, then there exists a continuous function* $\Psi_A : I \to [0,1]$ *such that*
$$\|(1_A - \Psi_A(G))(\nu + 1)\|_{L^1(\mathbb{Z}_N)} = O(\sigma). \quad (7.2)$$
Furthermore, Ψ_A *lies in a fixed compact set* $E = E_{\varepsilon,\sigma}$ *of* $C^0(I)$ *(which is independent of* F*,* ν*,* N*, or* A*).*

Proof. Let $0 < \sigma < 1/2$ be a parameter to be chosen later (it will eventually tend slowly to zero as $N \to \infty$). Define a (discrete) non-negative measure μ on the unit circle $\mathbb{T} := \mathbb{R}/\mathbb{Z}$ by setting
$$\mu(A) := \mathbb{E}\big(\mathbf{1}_A(\pi(G(x)/\varepsilon)(\nu(x) + 1) \bigm| x \in \mathbb{Z}_N\big)$$
for all $A \subseteq \mathbb{T}$, where $\pi : \mathbb{R} \to \mathbb{T}$ is the canonical projection map. Observe that
$$\mu(\mathbb{T}) = \mathbb{E}(\nu + 1) = O(1).$$
By the pigeonhole principle, we can thus find an $\alpha \in \mathbb{T}$ such that
$$\mu([\alpha - \sigma, \alpha + \sigma]) = O(\sigma).$$
Let us fix α so that this property holds. By definition of μ, we thus have
$$\sum_{n \in \mathbb{Z}} \mathbb{E}\big(\mathbf{1}_{G(x) \in [\varepsilon(n-\sigma+\alpha), \varepsilon(n+\sigma+\alpha)]}(\nu(x) + 1) \bigm| x \in \mathbb{Z}_N\big) = O(\sigma). \quad (7.3)$$
We now set $\mathcal{B}_\varepsilon(G)$ to be the σ-algebra whose atoms are the sets $(G)^{-1}([\varepsilon(n + \alpha), \varepsilon(n + 1 + \alpha)))$ for $n \in \mathbb{Z} + \alpha$. This is well-defined since the intervals $[\varepsilon(n + \alpha), \varepsilon(n + 1 + \alpha))$ tile the real line.

It is clear that if $\mathcal{B}$ is an arbitrary σ-algebra, then on any atom of $\mathcal{B} \vee \mathcal{B}_\varepsilon(G)$, the function G takes values in an interval of diameter ε, which yields (7.1). Now we verify the approximation by continuous functions property. Let $A := (G)^{-1}([\varepsilon(n+\alpha), \varepsilon(n+1+\alpha)))$ be an atom. Since G takes values in I, we may assume that $n = O(1/\varepsilon)$, since A is empty otherwise; note that this already establishes the bounded complexity property. Let $\psi_\sigma : \mathbb{R} \to [0,1]$ be a fixed continuous cutoff function which equals 1 on $[\sigma, 1 - \sigma]$ and vanishes outside of $[-\sigma, 1+\sigma]$, and define $\Psi_A(x) := \psi_\sigma(\frac{x}{\varepsilon} - n - \alpha)$. Then it is clear that Ψ_A ranges over a compact subset $E_{\varepsilon,\sigma}$ of $C^0(I)$ (because n and α are bounded). Furthermore from (7.3) it is clear that we have (7.2). The claim follows. □

We now specialize to the case when the functions G are basic anti-uniform functions.

Proposition 7.3. *Let* ν *be a* k*-pseudorandom measure. Let* $K \geqslant 1$ *be an fixed integer and let* $\mathcal{D}F_1, \dots, \mathcal{D}F_K \in L^1(\mathbb{Z}_N)$ *be basic anti-uniform functions. Let* $0 < \varepsilon < 1$ *and* $0 < \sigma < 1/2$ *be parameters, and let* $\mathcal{B}_{\varepsilon,\sigma}(\mathcal{D}F_j)$ *for* $j = 1, \dots, K$ *be constructed as in Lemma 7.2. Let* $\mathcal{B} := \mathcal{B}_{\varepsilon,\sigma}(\mathcal{D}F_1) \vee \ldots \vee \mathcal{B}_{\varepsilon,\sigma}(\mathcal{D}F_K)$*. Then if* σ *is sufficiently small depending on* K, ε*, and* N *is sufficiently large depending on* K*,* ε*,* σ*, we have*
$$\|\mathcal{D}F_j - \mathbb{E}(\mathcal{D}F_j|\mathcal{B})\|_{L^\infty(\mathbb{Z}_N)} \leqslant \varepsilon \text{ for all } 1 \leqslant j \leqslant K. \quad (7.4)$$

Furthermore there exists a set Ω which lies in $\mathcal{B}$ such that

$$\mathbb{E}((\nu+1)\mathbf{1}_\Omega) = O_{K,\varepsilon}(\sigma^{1/2}) \tag{7.5}$$

and such that

$$\|(1-\mathbf{1}_\Omega)\mathbb{E}(\nu-1|\mathcal{B})\|_{L^\infty(\mathbb{Z}_N)} = O_{K,\varepsilon}(\sigma^{1/2}). \tag{7.6}$$

Remark. We recommend that the reader pretend that the exceptional set Ω is empty; in practice we shall be able to set σ small enough that the contribution of Ω to our calculations will be negligible.

Proof. The claim (7.4) follows immediately from (7.1). Now we prove (7.5) and (7.6). Since each of the $\mathcal{B}_{\varepsilon,\sigma}(\mathcal{D}F_j)$ are generated by $O(1/\varepsilon)$ atoms, we see that $\mathcal{B}$ is generated by $O_{K,\varepsilon}(1)$ atoms. Call an atom A of $\mathcal{B}$ *small* if $\mathbb{E}((\nu+1)\mathbf{1}_A) \leqslant \sigma^{1/2}$, and let Ω be the union of all the small atoms. Then clearly Ω lies in $\mathcal{B}$ and obeys (7.5). To prove the remaining claim (7.6), it suffices to show that

$$\frac{\mathbb{E}((\nu-1)\mathbf{1}_A)}{\mathbb{E}(\mathbf{1}_A)} = \mathbb{E}(\nu-1|A) = o_{K,\varepsilon,\sigma}(1) + O_{K,\varepsilon}(\sigma^{1/2}) \tag{7.7}$$

for all atoms A in B which are not small. However, by definition of "small" we have

$$\mathbb{E}((\nu-1)\mathbf{1}_A) + 2\mathbb{E}(\mathbf{1}_A) = \mathbb{E}((\nu+1)\mathbf{1}_A) \geqslant \sigma^{1/2}.$$

Thus to complete the proof of (7.7) it will suffice (since σ is small and N is large) to show that

$$\mathbb{E}((\nu-1)\mathbf{1}_A) = o_{K,\varepsilon,\sigma}(1) + O_{K,\varepsilon}(\sigma). \tag{7.8}$$

On the other hand, since A is the intersection of K atoms $A_1, \ldots, A_K$ from $\mathcal{B}_{\varepsilon,\sigma}(\mathcal{D}F_1), \ldots, \mathcal{B}_{\varepsilon,\sigma}(\mathcal{D}F_K)$ respectively, we see from Proposition 7.2 (and Hölder's inequality) that we can find a continuous function $\Psi_A : I^K \to [0,1]$ such that

$$\|(\nu+1)(1_A - \Psi_A(\mathcal{D}F_1,\ldots,\mathcal{D}F_K))\|_{L^1(\mathbb{Z}_N)} = O_K(\sigma),$$

so in particular

$$\|(\nu-1)(1_A - \Psi_A(\mathcal{D}F_1,\ldots,\mathcal{D}F_K))\|_{L^1(\mathbb{Z}_N)} = O_K(\sigma),$$

Furthermore Ψ_A lives in a compact set $E_{\varepsilon,\sigma,K}$ of $C^0(I^K)$. From this and Proposition 6.2 we have

$$\mathbb{E}((\nu-1)\Psi_A(\mathcal{D}F_1,\ldots,\mathcal{D}F_K))) = o_{K,\varepsilon,\sigma}(1) = O_{K,\varepsilon}(\sigma^{1/2})$$

since N is assuemd large depending in K, ε, σ, and the claim (7.8) now follows from the triangle inequality. □

Remarks. This σ-algebra $\mathcal{B}$ is closely related to the compact σ-algebras studied in the ergodic theory proof of Szemerédi's theorem, see for instance [9, 10]. In the case $k = 3$ they are closely connected to the Kronecker factor of an ergodic system, and for higher k they are related to $(k-2)$-step nilsystems, see e.g. [25, 39].

8. A Furstenberg tower, and the proof of Theorem 3.5

We now have enough machinery to deduce Theorem 3.5 from Theorem 2.3. The key proposition is the following decomposition, which splits an arbitrary function into uniform and anti-uniform components (plus a negligible error).

Proposition 8.1 (Generalized Koopman-von Neumann theorem). *Let ν be a k-pseudorandom measure, and let $f \in L^1(\mathbb{Z}_N)$ be a non-negative function satisfying $0 \leqslant f(x) \leqslant \nu(x)$ for all x. Let $0 < \varepsilon \ll 1$ be a small parameter, and assume N sufficiently large depending on ε. Then there exists a σ-algebra $\mathcal{B}$ and an exceptional set $\Omega \in \mathcal{B}$ obeying the smallness condition*

$$\mathbb{E}(\nu \mathbf{1}_\Omega) = o_\varepsilon(1) \tag{8.1}$$

and such that ν is uniformly distributed outside of Ω:

$$\|(1 - \mathbf{1}_\Omega)\mathbb{E}(\nu - 1|\mathcal{B})\|_{L^\infty} = o_\varepsilon(1). \tag{8.2}$$

Furthermore we have the uniformity estimate

$$\|(1 - \mathbf{1}_\Omega)(f - \mathbb{E}(f|\mathcal{B}))\|_{U^{k-1}} \leqslant \varepsilon^{1/2^k}. \tag{8.3}$$

Remarks. As in the previous section, the exceptional set Ω should be ignored. The ordinary Koopman-von Neumann theory in ergodic theory asserts, among other things, that any function f on a measure-preserving system $(X, \mathcal{B}, T, \mu)$ can be orthogonally decomposed into a "weakly mixing part" $f - \mathbb{E}(f|\mathcal{B})$ (in which $f - \mathbb{E}(f|\mathcal{B})$ is asymptotically orthogonal to its shifts $T^n(f - \mathbb{E}(f|\mathcal{B}))$ on the average) and an "almost periodic part" $\mathbb{E}(f|\mathcal{B})$ (whose shifts form a precompact set); here $\mathcal{B}$ is the *Kronecker factor*, i.e. the σ-algebra generated by the almost periodic functions (or equivalently, by the eigenfunctions of T). This is somewhat related to the $k = 3$ case of the above Proposition, hence our labeling of that proposition as a generalized Koopman-von Neumann theorem. A slightly more quantitative analogy for the $k = 3$ case would be the assertion that any function bounded by a pseudorandom measure can be decomposed into a uniform component with small Fourier coefficients, and an anti-uniform component which consists of only a few Fourier coefficients (and in particular is bounded). For related ideas see [5, 18].

Proof of Theorem 3.5 assuming Proposition 8.1. Let f, δ be as in Theorem 3.5, and let $0 < \varepsilon \ll \delta$ be a parameter to be chosen later. Let $\mathcal{B}$ be as in the above decomposition, and write $f_U := (1 - \mathbf{1}_\Omega)(f - \mathbb{E}(f|\mathcal{B}))$ and $f_{U^\perp} := (1 - \mathbf{1}_\Omega)\mathbb{E}(f|\mathcal{B})$ (the subscript U stands for uniform, and $U^\perp$ for anti-uniform). Observe from (8.1), (3.4), (3.5) and the measurability of Ω that

$$\mathbb{E}(f_{U^\perp}) = \mathbb{E}((1 - \mathbf{1}_\Omega)f) \geqslant \mathbb{E}(f) - \mathbb{E}(\nu \mathbf{1}_\Omega) \geqslant \delta - o_\varepsilon(1).$$

Also, by (8.2) we see that $f_{U^\perp}$ is bounded above by $1 + o_\varepsilon(1)$. Since f is non-negative, $f_{U^\perp}$ is also. We may thus[14] apply Theorem 2.3 to obtain

$$\mathbb{E}\big(f_{U^\perp}(x) f_{U^\perp}(x + r) \ldots f_{U^\perp}(x + (k-1)r) \;\big|\; x, r \in \mathbb{Z}_N\big) \geqslant c(k, \delta) - o_\varepsilon(1).$$

[14] There is an utterly trivial issue here which we have ignored here, which is that $f_{U^\perp}$ is not bounded above by 1 but by $1 + o_\varepsilon(1)$, and that the density is bounded below by $\delta - o_\varepsilon(1)$ rather than δ. One can easily get around this by modifying $f_{U^\perp}$ by $o_\varepsilon(1)$ before applying Theorem 2.3, incurring a net error of $o_\varepsilon(1)$ at the end since $f_{U^\perp}$ is bounded.

On the other hand, from (8.3) we have $\|f_U\|_{U^{k-1}} \leqslant \varepsilon^{1/2^k}$; since $(1-\mathbf{1}_\Omega)f$ is bounded by ν and $f_{U^\perp}$ is bounded by $1+o_\varepsilon(1)$, we thus see that f_U is pointwise bounded by $\nu+1+o_\varepsilon(1)$. Applying the generalized von Neumann theorem (Proposition 5.3) we thus see that

$$\mathbb{E}\big(f_0(x)f_1(x+r)\dots f_{k-1}(x+(k-1)r)\ \big|\ x,r\in\mathbb{Z}_N\big) = O(\varepsilon^{1/2^k})$$

whenever each f_j is equal to f_U or $f_{U^\perp}$, with at least one f_j equal to f_U. Adding these two estimates together we obtain

$$\mathbb{E}(\tilde{f}(x)\tilde{f}(x+r)\dots\tilde{f}(x+(k-1)r)|x,r\in\mathbb{Z}_N) \geqslant c(k,\delta) - O(\varepsilon^{1/2^k}) - o_\varepsilon(1),$$

where $\tilde{f} := f_U + f_{U^\perp} = (1-\mathbf{1}_\Omega)f$. But since $0 \leqslant (1-\mathbf{1}_\Omega)f \leqslant f$ we obtain

$$\mathbb{E}(f(x)f(x+r)\dots f(x+(k-1)r)|x,r\in\mathbb{Z}_N) \geqslant c(k,\delta) - O(\varepsilon^{1/2^k}) - o_\varepsilon(1).$$

Since ε can be made arbitrarily small, the claim follows. □

To complete the proof of Theorem 3.5, it suffices to prove Proposition 8.1. To construct the σ-algebra $\mathcal{B}$ required in the Proposition, we will use the philosophy laid out by Furstenberg in his ergodic structure theorem (see [9, 10]), which decomposes any measure-preserving system into a weakly-mixing extension of a tower of compact extensions. In our setting, the idea is roughly speaking as follows. We initialize $\mathcal{B}$ to be the trivial σ-algebra $\mathcal{B} = \{\emptyset, \mathbb{Z}_N\}$. If the function $f - \mathbb{E}(f|\mathcal{B})$ is already uniform (in the sense of (8.3)), then we can terminate the algorithm. Otherwise, we use the machinery of dual functions, developed in §6, to locate an anti-uniform function G_1 which has some non-trivial correlation with f, and add the level sets of G_1 to the σ-algebra $\mathcal{B}$; the non-trivial correlation property will ensure that the L^2 norm of $\mathbb{E}(f|\mathcal{B})$ increases by a non-trivial amount during this procedure, while the pseudorandomness of ν will ensure that $\mathbb{E}(f|\mathcal{B})$ remains uniformly bounded. We then repeat the above algorithm until $f - \mathbb{E}(f|\mathcal{B})$ becomes sufficiently uniform, at which point we terminate the algorithm. In the original ergodic theory arguments of Furstenberg this algorithm was not guaranteed to terminate, and indeed one required the axiom of choice (in the guise of Zorn's lemma) in order to conclude the structure theorem. However, in our setting we can (almost surely) terminate in a bounded number of steps (in fact in at most $2^{2^k}/\varepsilon$ steps), because there is a quantitative L^2-increment to the bounded function $\mathbb{E}(f|\mathcal{B})$ at each stage.

Such a strategy will be familiar to any reader acquainted with the proof of Szemerédi's regularity lemma [35]. This is no coincidence: there is in fact a close connection between regularity lemmas such as [17, 19, 35] and ergodic theory of the type we have brushed up against in this paper. Indeed there are strong analogies between all of the known proofs of Szemerédi's theorem, despite the fact that they superficially appear to use very different techniques.

We turn to the details. Fix ε, and let K_0 be the smallest integer greater than $2^{2^k}/\varepsilon$. We shall need a parameter $0 < \sigma \ll \varepsilon$ which we shall choose later, and then we shall assume N is sufficiently large depending on σ and ε.

To construct $\mathcal{B}$ and Ω we shall iteratively construct a sequence of basic anti-uniform functions $\mathcal{D}F_1, \ldots, \mathcal{D}F_K$ on $\mathbb{Z}_N$, exceptional sets $\Omega_0 \subseteq \Omega_1 \subseteq \ldots \subseteq \Omega_K \subseteq \mathbb{Z}_N$, and a nested sequence of σ-algebras $\mathcal{B}_0 \subseteq \ldots \subseteq \mathcal{B}_K$ for some $0 \leqslant K \leqslant K_0$ as follows.

- Step 0. Initialize $K = 0$, and define $\mathcal{B}_0 := \{\emptyset, \mathbb{Z}_N\}$ and $\Omega_0 := \emptyset$.
- Step 1. Set $F_{K+1} := (1 - \mathbf{1}_{\Omega_K})(f - \mathbb{E}(f|\mathcal{B}_K))$. If we have
$$\|F_{K+1}\|_{U^{k-1}} \leqslant \varepsilon^{1/2^k}$$
then we set $\Omega := \Omega_K$ and $\mathcal{B} = \mathcal{B}_K$, and successfully terminate the algorithm.
- Step 2. If instead we have
$$\|F_{K+1}\|_{U^{k-1}} > \varepsilon^{1/2^k}, \tag{8.4}$$
then we let $\mathcal{B}_{K+1} := \mathcal{B}_K \vee \mathcal{B}_{\varepsilon,\sigma}(\mathcal{D}F_{K+1})$. (Here we of course need $K \leqslant K_0$, but this will be guaranteed by Step 4 below).
- Step 3. Locate an exceptional set $\Omega_{K+1} \supset \Omega_K$ in $\mathcal{B}_{K+1}$ obeying the smallness condition
$$\mathbb{E}((\nu + 1)\mathbf{1}_{\Omega_{K+1}}) = O_{K,\varepsilon}(\sigma^{1/2}) \tag{8.5}$$
and such that we have the uniform bound
$$\|(1 - \mathbf{1}_{\Omega_{K+1}})\mathbb{E}(\nu - 1|\mathcal{B}_{K+1})\|_{L^\infty} = O_{K,\varepsilon}(\sigma^{1/2})$$
outside of the exceptional set. If such an exceptional set cannot be found, we terminate the algorithm with an error; otherwise, we move on to Step 4.
- Step 4. Increment K to $K + 1$. If $K > K_0$, then we terminate the algorithm with an error; otherwise, return to Step 1.

The integer K indexes the iteration number of the algorithm, thus we begin with the zeroth iteration when $K = 0$, then the first iteration when $K = 1$, etc. It is worth noting that apart from $O_{K,\varepsilon}(\sigma^{1/2})$ error terms, none of the bounds we will encounter while executing this algorithm will actually depend on K.

Assume for the moment that the above algorithm does not terminate with an error in Step 3 or Step 4. Then it is clear that after at most $K_0 + 1$ iterations (in particular, after finitely many iterations) of this algorithm, we will have generated a σ-algebra $\mathcal{B}$ and an exceptional set Ω with the desired properties required for Proposition 8.1, with error terms of $O_{K,\varepsilon}(\sigma^{1/2})$ instead of $o_\varepsilon(1)$, if N is sufficiently large depending on σ, K, ε. But by making σ decay sufficiently slowly to zero, we can replace the $O_{K,\varepsilon}(\sigma^{1/2})$ bounds by $o_\varepsilon(1)$; note that the dependence of the error terms on K will not be relevant since K is bounded by K_0, which depends only on ε.

To conclude the proof of Proposition 8.1 it will thus suffice to show that the above algorithm does not terminate with an error. Suppose as a hypothesis for induction on K_1, $0 \leqslant K_1 \leqslant K_0$, that the algorithm will either terminate without error, or else reach Step 2 of the K_1^{th} iteration without returning an error; note that this claim is trivial for $K_1 = 0$. Supposing that the claim has been proven for some $K_1 < K_0$, we wish to verify it for $K_1 + 1$. Without loss of generality we may assume that the algorithm has not yet terminated by the time it reaches Step 2 of the K_1^{th} iteration. At this stage the σ-algebras $\mathcal{B}_0, \ldots \mathcal{B}_{K_1+1}$, the basic anti-uniform functions $\mathcal{D}F_1, \ldots, \mathcal{D}F_{K_1+1}$, and exceptional sets $\Omega_0, \ldots, \Omega_{K_1}$ have already been constructed. We then claim the bounds

$$\|\mathcal{D}F_j\|_{L^\infty} \leqslant 2^{2^{k-1}-1} + O_{j,\varepsilon}(\sigma^{1/2}) \tag{8.6}$$

for all $1 \leqslant j \leqslant K_1 + 1$. To see this, observe from previous iterations of Step 3 in the construction (or by Step 0, when $j = 1$) that

$$\|(1 - \mathbf{1}_{\Omega_{j-1}})\mathbb{E}(\nu - 1|\mathcal{B}_{j-1})\|_{L^\infty} = O_{j,\varepsilon}(\sigma^{1/2}),$$

and thus

$$\mathbb{E}(\nu|\mathcal{B}_{j-1})(x) = 1 + O_{j-1,\varepsilon}(\sigma^{1/2}) \text{ for all } x \notin \Omega_{j-1}.$$

Since $0 \leqslant f(x) \leqslant \nu(x)$ we conclude the pointwise estimates

$$0 \leqslant (1 - \mathbf{1}_{\Omega_{j-1}}(x))\mathbb{E}(f|\mathcal{B}_{j-1})(x) \leqslant 1 + O_{j,\varepsilon}(\sigma^{1/2}) \tag{8.7}$$

and hence (by definition of F_j)

$$|F_j(x)| \leqslant (1 + O_{j,\varepsilon}(\sigma^{1/2}))(\nu(x) + 1). \tag{8.8}$$

Applying (6.5), the claim (8.6) follows. This shows in particular that $\mathcal{D}F_1, \ldots, \mathcal{D}F_{K_1+1}$ are basic anti-uniform functions (up to multiplicative errors of $1 + O_{K_1,\varepsilon}(\sigma^{1/2})$, which are negligible).

Now observe from construction that $\mathcal{B}_{K_1+1}$ is the σ-algebra generated by $\mathcal{B}_{\varepsilon,\alpha_1}(\mathcal{D}F_1)$, $\ldots, \mathcal{B}_{\varepsilon,\alpha_{K_1+1}}(\mathcal{D}F_{K_1+1})$. We apply Proposition 7.3 (absorbing the errors of $O_{K_1,\varepsilon}(\sigma^{1/2})$) to conclude that we may find a set Ω in $\mathcal{B}_{K_1+1}$ such that

$$\mathbb{E}((\nu + 1)\mathbf{1}_\Omega) = O_{K_1,\varepsilon}(\sigma^{1/2})$$

and

$$\|(1 - \mathbf{1}_\Omega)\mathbb{E}(\nu - 1|\mathcal{B}_{K+1})\|_{L^\infty} = O_{K_1,\varepsilon}(\sigma^{1/2}).$$

If one sets $\Omega_{K+1} := \Omega_K \cup \Omega$ we see (using either the previous iteration of Step 3, or using Step 0) that Ω_{K+1} will obey the required properties to execute Step 3 without terminating. Since $K_1 < K_0$, we can now execute the algorithm all the way until Step 2 of the $(K_1 + 1)^{\text{st}}$ iteration (or else terminate without error), which closes the induction and concludes the claim.

In light of the above claim, we now know that the algorithm either terminates without error, or gets all the way to the K_0^{th} iteration. We will now show that the latter case cannot actually occur (if ε, σ are sufficiently small and N sufficiently large). The key point is that if the algorithm actually reaches Step 3 of the K_0^{th} iteration without terminating, then we have the L^2 increment property

$$\|(1 - \mathbf{1}_{\Omega_j})\mathbb{E}(f|\mathcal{B}_j)\|_{L^2}^2 \geqslant \|(1 - \mathbf{1}_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1})\|_{L^2}^2 + 2^{2^k-2}\varepsilon - O_{j,\varepsilon}(\sigma^{1/2}) - O(\varepsilon^2) \tag{8.9}$$

for all $1 \leqslant j \leqslant K_0$ (assuming N sufficiently large depending on K_0 and ε). On the other hand, from (8.7) we have

$$0 \leqslant \|(1 - \mathbf{1}_{\Omega_j})\mathbb{E}(f|\mathcal{B}_j)\|_{L^2}^2 \leqslant 1 + O_{j,\varepsilon}(\sigma^{1/2}) \text{ for all } 0 \leqslant j \leqslant K_0. \tag{8.10}$$

Since K_0 is the smallest integer greater than than $2^{2^k}/\varepsilon$, we obtain a contradiction from the pigeonhole principle (taking ε sufficiently small, then σ sufficiently small depending on ε, and then N sufficiently large depending on ε, σ so that the $O_{j,\varepsilon}(\sigma^{1/2})$ and $O(\varepsilon)$ terms are negligible).

It remains to actually prove the increment property (8.9). To do this we exploit the failure of the algorithm to terminate at Step 2 of the $(j-1)^{\text{st}}$ iteration. This implies that

$$\|F_j\|_{U^{k-1}} \geqslant \varepsilon^{1/2^k}.$$

By Lemma 6.1 and the definition of F_j we thus have

$$|\langle (1-\mathbf{1}_{\Omega_{j-1}})(f-\mathbb{E}(f|\mathcal{B}_{j-1})), \mathcal{D}F_j\rangle| = \|F_j\|_{U^{k-1}}^{2^{k-1}} \geqslant \varepsilon^{1/2}.$$

On the other hand, from the bounds (8.6), (8.8), (8.5) we have

$$\langle (\mathbf{1}_{\Omega_j}-\mathbf{1}_{\Omega_{j-1}})(f-\mathbb{E}(f|\mathcal{B}_{j-1})), \mathcal{D}F_j\rangle = O_{j,\varepsilon}(\sigma^{1/2})$$

while from (8.8), (7.1) we have

$$\langle (1-\mathbf{1}_{\Omega_j})(f-\mathbb{E}(f|\mathcal{B}_{j-1})), \mathcal{D}F_j - \mathbb{E}(\mathcal{D}F_j|\mathcal{B}_j)\rangle = O(\varepsilon).$$

By the triangle inequality we thus have

$$|\langle (1-\mathbf{1}_{\Omega_j})(f-\mathbb{E}(f|\mathcal{B}_{j-1})), \mathbb{E}(\mathcal{D}F_j|B_j)\rangle| \geqslant \varepsilon^{1/2} - O_{j,\varepsilon}(\sigma^{1/2}) - O(\varepsilon).$$

But since $(1-\mathbf{1}_{\Omega_j})$, $\mathbb{E}(\mathcal{D}F_j|B_j)$, and $\mathbb{E}(f|\mathcal{B}_{j-1})$ are all measurable in $\mathcal{B}_j$, we can replace f by $\mathbb{E}(f|\mathcal{B}_j)$, thus

$$|\langle (1-\mathbf{1}_{\Omega_j})(\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1})), \mathbb{E}(\mathcal{D}F_j|\mathcal{B}_j)\rangle| \geqslant \varepsilon^{1/2} - O_{j,\varepsilon}(\sigma^{1/2}) - O(\varepsilon).$$

By the Cauchy-Schwarz inequality and (8.6) we thus have

$$\|(1-\mathbf{1}_{\Omega_j})(\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1}))\|_{L^2} \geqslant 2^{-2^{k-1}+1}\varepsilon^{1/2} - O_{j,\varepsilon}(\sigma^{1/2}) - O(\varepsilon). \tag{8.11}$$

Morally speaking, this implies (8.9) thanks to Pythagoras's theorem, but the presence of the exceptional sets Ω_j, Ω_{j-1} mean that we have to take a little care (especially since we have no L^2 control on ν). We first observe from (8.5) and (8.7) we have

$$\|(\mathbf{1}_{\Omega_j}-\mathbf{1}_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1})\|_{L^2} = O_{j,\varepsilon}(\sigma^{1/2}).$$

By the triangle inequality (and (8.10)) we thus see that to prove (8.9) it will suffice to prove that

$$\|(1-\mathbf{1}_{\Omega_j})\mathbb{E}(f|\mathcal{B}_j)\|_{L^2}^2 \geqslant \|(1-\mathbf{1}_{\Omega_j})\mathbb{E}(f|\mathcal{B}_{j-1})\|_{L^2}^2 + \varepsilon^{1/2} - O_{j,\varepsilon}(\sigma^{1/2}) - O(\varepsilon).$$

The left-hand side can be expanded using the cosine rule as

$$\begin{aligned}\|(1-\mathbf{1}_{\Omega_j})\mathbb{E}(f|\mathcal{B}_{j-1})\|_{L^2}^2 &+ \|(1-\mathbf{1}_{\Omega_j})(\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1}))\|_{L^2}^2 \\ &+ 2\langle (1-\mathbf{1}_{\Omega_j})\mathbb{E}(f|\mathcal{B}_{j-1}), (1-\mathbf{1}_{\Omega_j})(\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1}))\rangle\end{aligned}$$

and so by (8.11) it will suffice to show the approximate orthogonality relationship

$$\langle (1-\mathbf{1}_{\Omega_j})\mathbb{E}(f|\mathcal{B}_{j-1}), (1-\mathbf{1}_{\Omega_j})(\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1}))\rangle = O_{j,\varepsilon}(\sigma^{1/2}).$$

Since $(1-\mathbf{1}_{\Omega_j})^2 = (1-\mathbf{1}_{\Omega_j})$, this term can be rewritten as

$$\langle (1-\mathbf{1}_{\Omega_j})\mathbb{E}(f|\mathcal{B}_{j-1}), \mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1})\rangle.$$

Now note that $(1-\mathbf{1}_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1})$ is measurable with respect to $\mathcal{B}_{j-1}$, and hence orthogonal to $\mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1})$ (since $\mathcal{B}_{j-1}$ is a sub-σ-algebra of $\mathcal{B}_j$). Thus the above expression can be rewritten as

$$\langle (\mathbf{1}_{\Omega_j}-\mathbf{1}_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1}), \mathbb{E}(f|\mathcal{B}_j)-\mathbb{E}(f|\mathcal{B}_{j-1})\rangle.$$

Again, since the left-hand side is measurable with respect to $\mathcal{B}_{j-1}$ and hence to $\mathcal{B}_j$, we can rewrite this as

$$\langle(\mathbf{1}_{\Omega_j} - \mathbf{1}_{\Omega_{j-1}})\mathbb{E}(f|\mathcal{B}_{j-1}), f - \mathbb{E}(f|\mathcal{B}_{j-1})\rangle.$$

But this is $O_{j,\varepsilon}(\sigma^{1/2})$ by (3.4), (8.5) and (8.7).

We have proven (8.9). This concludes the proof that the algorithm to generate $\mathcal{B}$ and Ω terminates without error, which proves Proposition 8.1 and hence Theorem 3.5. □

9. A pseudorandom measure which majorises the primes

Having concluded the proof of Theorem 3.5, we are now ready to apply it to the specific situation of locating arithmetic progressions in the primes. As in almost any additive problem involving the primes, we begin by considering the von Mangoldt function Λ defined by $\Lambda(n) = \log p$ if $n = p^m$ and 0 otherwise. Actually, for us the higher prime powers $p^2, p^3, \ldots$ will play no rôle whatsoever and will be discarded very shortly.

From the prime number theorem we know that the average value of $\Lambda(n)$ is $1 + o(1)$. In order to prove Theorem 1.1 (or Theorem 1.2), it would suffice to exhibit a measure $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ such that $\nu(n) \geqslant c(k)\Lambda(n)$ for some $c(k) > 0$ depending only on k, and which is k-pseudorandom. Unfortunately, such a measure cannot exist because the primes (and the von Mangoldt function) are concentrated on certain residue classes. Specifically, for any integer $q > 1$, Λ is only non-zero on those $\phi(q)$ residue classes $a(\text{mod}\, q)$ for which $(a, q) = 1$, whereas a pseudorandom measure can easily be shown to be uniformly distributed across all q residue classes; here of course $\phi(q)$ is the Euler totient function. Since $\phi(q)/q$ can be made arbitrarily small, we therefore cannot hope to obtain a pseudorandom majorant with the desired property $\nu(n) \geqslant c(k)\Lambda(n)$.

To get around this difficulty we employ a device which we call the W-trick[15], which effectively removes the arithmetic obstructions to uniformity arising from the very small primes. Let $w = w(N)$ be any function tending slowly[16] to infinity with N, so that $1/w(N) = o(1)$, and let $W = \prod_{p \leqslant w(N)} p$ be the product of the primes up to $w(N)$. Define the modified von Mangoldt function $\tilde{\Lambda} : \mathbb{Z}^+ \to \mathbb{R}^+$ by

$$\widetilde{\Lambda}(n) := \begin{cases} \frac{\phi(W)}{W}\log(Wn+1) & \text{when } Wn+1 \text{ is prime} \\ 0 & \text{otherwise.} \end{cases}$$

Note that we have discarded the contribution of the prime powers since we ultimately wish to count arithmetic progressions in the primes themselves. This W-trick exploits

[15]The reader will observe some similarity between this trick and the use of σ-algebras in the previous section to remove non-uniformity from the system. Here, of course, the precise obstruction to non-uniformity in the primes is very explicit, whereas the exact structure of the σ-algebras constructed in the previous section are somewhat mysterious. In the specific case of the primes, we expect (through such conjectures as the Hardy-Littlewood prime tuple conjecture) that the primes are essentially uniform once the obstructions from small primes are removed, and hence the algorithm of the previous section should in fact terminate immediately at the $K = 0$ iteration. However we emphasize that our argument does not give (or require) any progress on this very difficult prime tuple conjecture, as we allow K to be non-zero.

[16]Actually, it will be clear at the end of the proof that we can in fact take w to be a sufficiently large number independent of N, depending only on k, however it will be convenient for now to make w slowly growing in N in order to take advantage of the $o(1)$ notation.

the trivial observation that in order to obtain arithmetic progressions in the primes, it suffices to do so in the modified primes $\{n \in \mathbb{Z} : Wn+1 \text{ is prime}\}$ (at the cost of reducing the number of such progressions by a polynomial factor in W at worst). We also remark that one could replace $Wn+1$ here by $Wn+b$ for any integer $1 \leqslant b < W$ coprime to W without affecting the arguments which follow.

Observe that if $w(N)$ is sufficiently slowly growing ($w(N) \ll \log\log N$ will suffice here) then by Dirichlet's theorem concerning the distribution of the primes in arithmetic progressions[17] such as $\{n : n \equiv 1(\mathrm{mod}\, W)\}$ we have $\sum_{n\leqslant N} \widetilde{\Lambda}(n) = N(1+o(1))$. With this modification, we can now majorize the primes by a pseudorandom measure as follows:

Proposition 9.1. *Write $\epsilon_k := 1/2^k(k+4)!$, and let N be a sufficiently large prime number. Then there is a k-pseudorandom measure $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ such that $\nu(n) \geqslant k^{-1}2^{-k-5}\widetilde{\Lambda}(n)$ for all $\epsilon_k N \leqslant n \leqslant 2\epsilon_k N$.*

Remark. The purpose of ϵ_k is to assist in dealing with wraparound issues, which arise from the fact that we are working on $\mathbb{Z}_N$ and not on $[-N, N]$. Standard sieve theory techniques (in particular the "fundamental lemma of sieve theory") can come very close to providing such a majorant, but the error terms on the pseudo-randomness are not of the form $o(1)$ but rather something like $O(2^{-2^{Ck}})$ or so. This unfortunately does not quite seem to be good enough for our argument, which crucially relies on $o(1)$ type decay, and so we have to rely instead of recent arguments of Goldston and Yıldırım.

Proof of Theorem 1.1 assuming Proposition 9.1. Let N be a large prime number. Define the function $f \in L^1(\mathbb{Z}_N)$ by setting $f(n) := k^{-1}2^{-k-5}\tilde{\Lambda}(n)$ for $\epsilon_k N \leqslant n \leqslant 2\epsilon_k N$ and $f(n) = 0$ otherwise. From Dirichlet's theorem we observe that

$$\mathbb{E}(f) = \frac{k^{-1}2^{-k-5}}{N} \sum_{\epsilon_k N \leqslant n \leqslant 2\epsilon_k N} \tilde{\Lambda}(n) = k^{-1}2^{-k-5}\epsilon_k(1+o(1)).$$

We now apply Proposition 9.1 and Theorem 3.5 to conclude that

$$\mathbb{E}\big(f(x)f(x+r)\dots f(x+(k-1)r) \bigm| x, r \in \mathbb{Z}_N\big) \geqslant c(k, k^{-1}2^{-k-5}\epsilon_k) - o(1).$$

Observe that the degenerate case $r = 0$ can only contribute at most $O(\frac{1}{N}\log^k N) = o(1)$ to the left-hand side and can thus be discarded. Furthermore, every progression counted by the expression on the left is not just a progression in $\mathbb{Z}_N$, but a genuine arithmetic progression of integers since $\epsilon_k < 1/k$. Since the right-hand side is positive (and bounded away from zero) for sufficiently large N, the claim follows from the definition of f and $\tilde{\Lambda}$. $\square$

Thus to obtain arbitrarily long arithmetic progressions in the primes, it will suffice to prove Proposition 9.1. This will be the purpose of the remainder of this section (with certain number-theoretic computations being deferred to §10 and the Appendix).

[17] In fact, all we need is that $\sum_{N\leqslant n\leqslant 2N} \widetilde{\Lambda}(n) \gg N$. One could avoid appealing to the theory of Dirichlet L-functions by replacing $n \equiv 1(\mathrm{mod}\, W)$ by $n \equiv b(\mathrm{mod}\, W)$, for some randomly chosen b coprime to W, if desired.

To obtain a majorant for $\tilde{\Lambda}(n)$, we begin with the well-known formula

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(n/d) = \sum_{d|n} \mu(d) \log(n/d)_+$$

for the von Mangoldt function, where μ is the Möbius function, and $\log(x)_+$ denotes the positive part of the logarithm, that is to say $\max(\log(x), 0)$. Here and in the sequel d is always understood to be a positive integer. Motivated by this, we define

Definition 9.2 (Goldston–Yıldırım)**.** Let R be a parameter (in applications it will be a small power of N). Define

$$\Lambda_R(n) := \sum_{\substack{d|n \\ d \leqslant R}} \mu(d) \log(R/d) = \sum_{d|n} \mu(d) \log(R/d)_+.$$

These truncated divisor sums have been studied in several papers, most notably the works of Goldston and Yıldırım [12, 13, 14] concerning the problem of finding small gaps between primes. We shall use a modification of their arguments for obtaining asymptotics for these truncated primes to prove that the measure ν defined below is pseudorandom.

Definition 9.3. Let $R := N^{k^{-1}2^{-k-4}}$, and let $\epsilon_k := 1/2^k(k+4)!$. We define the function $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ by

$$\nu(n) \ := \begin{cases} \frac{\phi(W)}{W} \frac{\Lambda_R(Wn+1)^2}{\log R} & \text{when } \epsilon_k N \leqslant n \leqslant 2\epsilon_k N \\ 1 & \text{otherwise} \end{cases}$$

for all $0 \leqslant n < N$, where we identify $\{0, \ldots, N-1\}$ with $\mathbb{Z}_N$ in the usual manner.

This ν will be our majorant for Proposition 9.1. We first verify that it is indeed a majorant.

Lemma 9.4. *$\nu(n) \geqslant 0$ for all $n \in \mathbb{Z}_N$, and furthermore we have $\nu(n) \geqslant k^{-1}2^{-k-5}\widetilde{\Lambda}(n)$ for all $\epsilon_k N \leqslant n \leqslant 2\epsilon_k N$ (if N is sufficiently large depending on k).*

Proof. The first claim is trivial. The second claim is also trivial unless $Wn+1$ is prime. From definition of R, we see that $Wn + 1 > R$ if N is sufficiently large. Then the sum over $d|Wn + 1$, $d \leqslant R$ in (9.2) in fact consists of just the one term $d = 1$. Therefore $\Lambda_R(Wn + 1) = \log R$, which means that $\nu(n) = \frac{\phi(W)}{W} \log R \geqslant k^{-1}2^{-k-5}\widetilde{\Lambda}(n)$ by construction of R and N (assuming $w(N)$ sufficiently slowly growing in N). □

We will have to wait a while to show that ν is actually a measure. The next proposition will be crucial in showing that ν has the linear forms property.

Proposition 9.5 (Goldston-Yıldırım)**.** *Let m, t be positive integers. For each $1 \leqslant i \leqslant m$, let $\psi_i(\mathbf{x}) := \sum_{j=1}^t L_{ij}x_j + b_i$, be linear forms with integer coefficients L_{ij} such that $|L_{ij}| \leqslant \sqrt{w(N)}/2$ for all $i = 1, \ldots m$ and $j = 1, \ldots, t$. We assume that the t-tuples $(L_{ij})_{j=1}^t$ are never identically zero, and that no two t-tuples are rational multiples of each other. Write $\theta_i = W\psi_i + 1$. Suppose that B is a product $\prod_{i=1}^t I_i \subset \mathbb{R}^t$ of t intervals I_i, each of which having length at least R^{10m}. Then (if the function $w(N)$ is sufficiently slowly growing in N)*

$$\mathbb{E}(\Lambda_R(\theta_1(\mathbf{x}))^2 \ldots \Lambda_R(\theta_m(\mathbf{x}))^2 | \mathbf{x} \in B) = (1 + o_{m,t}(1)) \left(\frac{W \log R}{\phi(W)}\right)^m.$$

Remarks. We have attributed this proposition to Goldston and Yıldırım, because it is a straightforward generalisation of [14, Proposition 2]. The W-trick makes much of the analysis of the so-called singular series (which is essentially just $(W/\phi(W))^m$ here) easier in our case, but to compensate we have the slight extra difficulty of dealing with forms in several variables.

To keep this paper as self-contained as possible, we give a proof of Proposition 9.5. In §10 the reader will find a proof which depends on an estimation of a certain contour integral involving the Riemann ζ-function. This is along the lines of [14, Proposition 2] but somewhat different in detail. The aforementioned integral is precisely the same as one that Goldston and Yıldırım find an asymptotic for. We recall their argument in the Appendix.

Much the same remarks apply to the next proposition, which will be of extreme utility in demonstrating that ν has the correlation property (Definition 3.2).

Proposition 9.6 (Goldston-Yıldırım)**.** *Let $m \geqslant 1$ be an integer, and let B be an interval of length at least R^{10m}. Suppose that $h_1, \dots, h_m$ are distinct integers satisfying $|h_i| \leqslant N^2$ for all $1 \leqslant i \leqslant m$, and let Δ denote the integer*

$$\Delta := \prod_{1 \leqslant i < j \leqslant m} |h_i - h_j|.$$

Then (for N sufficiently large depending on m, and assuming the function $w(N)$ sufficiently slowly growing in N)

$$\begin{aligned} &\mathbb{E}(\Lambda_R(W(x_1+h_1)+1)^2 \dots \Lambda_R(W(x_m+h_m)+1)^2 | x \in B) \\ &\leqslant (1+o_m(1)) \left(\frac{W \log R}{\phi(W)}\right)^m \prod_{p|\Delta} (1+O_m(p^{-1/2})). \end{aligned} \tag{9.1}$$

Here and in the sequel, p is always understood to be prime.

Assuming both Proposition 9.5 and Proposition 9.6, we can now conclude the proof of Proposition 9.1. We begin by showing that ν is indeed a measure.

Lemma 9.7. *The measure ν constructed in Definition 9.3 obeys the estimate $\mathbb{E}(\nu) = 1 + o(1)$.*

Proof. Apply Proposition 9.5 with $m := t := 1$, $\psi_1(x_1) := x_1$ and $B := [\epsilon_k N, 2\epsilon_k N]$ (taking N sufficiently large depending on k, of course). Comparing with Definition 9.3 we thus have

$$\mathbb{E}(\nu(x) \mid x \in [\epsilon_k N, 2\epsilon_k N]) = 1 + o(1).$$

But from the same definition we clearly have

$$\mathbb{E}(\nu(x) \mid x \in \mathbb{Z}_N \backslash [\epsilon_k N, 2\epsilon_k N]) = 1;$$

Combining these two results confirms the lemma. □

Now we verify the linear forms condition, which is proven in a similar spirit to the above lemma.

Proposition 9.8. *The measure ν satisfies the $(k \cdot 2^{k-1}, 3k-4, k)$-linear forms condition.*

Proof. Let $\psi_i(x) = \sum_{j=1}^t L_{ij}x_j + b_i$ be linear forms of the type which feature in Definition 3.1. That is to say, we have $m \leqslant k \cdot 2^{k-1}$, $t \leqslant 3k-4$, the L_{ij} are rational numbers with numerator and denominator at most k in absolute value, and none of the t-tuples $(L_{ij})_{j=1}^t$ is zero or is equal to a rational multiple of any other. We wish to show that

$$\mathbb{E}(\nu(\psi_1(\mathbf{x}))\dots\nu(\psi_m(\mathbf{x})) \mid \mathbf{x} \in \mathbb{Z}_N^m) = 1 + o(1). \tag{9.2}$$

We may clear denominators and assume that all the L_{ij} are integers, at the expense of increasing the bound on L_{ij} to $|L_{ij}| \leqslant (k+1)!$. Since $w(N)$ is growing to infinity in N, we may assume that $(k+1)! < \sqrt{w(N)}/2$ by taking N sufficiently large. This is required in order to apply Proposition 9.5 as we have stated it.

The two-piece definition of ν in Definition 9.3 means that we cannot apply Proposition 9.5 immediately.

We chop the range of summation in (9.2) into Q^t equal-sized boxes, where $Q = Q(N)$ is a slowly growing function of N to be chosen later. Thus let

$$B_{u_1,\dots,u_t} = \{\mathbf{x} : x_j \in [\lfloor u_jN/Q \rfloor, \lfloor (u_j+1)N/Q \rfloor), j = 1, \dots, t\},$$

where the u_j are to be considered $(\bmod\, Q)$. Observe that the left-hand side of (9.2) can be rewritten as

$$\mathbb{E}(\mathbb{E}(\nu(\psi_1(\mathbf{x}))\dots\nu(\psi_m(\mathbf{x}))|\mathbf{x} \in B_{u_1,\dots,u_t})|u_1,\dots,u_t \in \mathbb{Z}_Q).$$

Call a t-tuple $(u_1,\dots,u_t) \in \mathbb{Z}_Q^t$ *nice* if for every $1 \leqslant i \leqslant m$, the sets $\psi_i(B_{u_1,\dots,u_t})$ are either completely contained in the interval $[\epsilon_k N, 2\epsilon_k N]$ or are completely disjoint from this interval. From Proposition 9.5 and Definition 9.3 we observe that

$$\mathbb{E}(\nu(\psi_1(\mathbf{x}))\dots\nu(\psi_m(\mathbf{x}))|\mathbf{x} \in B_{u_1,\dots,u_t}) = 1 + o_{m,t}(Q^t)$$

whenever $(u_1,\dots,u_t)$ is nice, since we can replace each of the $\nu(\psi_i(x))$ factors by either $\frac{\phi(W)}{W\log R}\Lambda_R^2(\theta_i(\mathbf{x}))$ or 1. When $(u_1,\dots,u_t)$ is not nice, then we can crudely bound ν by $1 + \frac{\phi(W)}{W\log R}\Lambda_R^2(\theta_i(\mathbf{x}))$, multiply out, and apply Proposition 9.5 again to obtain

$$\mathbb{E}(\nu(\psi_1(\mathbf{x}))\dots\nu(\psi_m(\mathbf{x}))|\mathbf{x} \in B_{u_1,\dots,u_t}) = O_{m,t}(1) + o_{m,t}(Q^t)$$

We shall shortly show that the proportion of non-nice t-tuples $(u_1,\dots,u_t)$ in $\mathbb{Z}_Q^t$ is at most $O_{m,t}(1/Q)$, and thus the left-hand side of (9.2) is $1 + o_{m,t}(Q^t) + O_{m,t}(1/Q)$, and the claim follows by choosing Q sufficiently slowly growing in N.

It remains to verify the claim about the proportion of non-nice t-tuples. Suppose $(u_1,\dots,u_t)$ is not nice. Then there exists $1 \leqslant i \leqslant m$ and $\mathbf{x}, \mathbf{x}' \in B_{u_1,\dots,u_t}$ such that $\psi_i(\mathbf{x})$ lies in the interval $[\epsilon_k N, 2\epsilon_k N]$, but $\psi_i(\mathbf{x}')$ does not. But from definition of $B_{u_1,\dots,u_t}$(and the boundedness of the L_{ij}) we have

$$\psi_i(\mathbf{x}), \psi_i(\mathbf{x}') = \sum_{j=1}^t L_{ij}\lfloor Nu_j/Q \rfloor + b_i + O_{m,t}(N/Q).$$

Thus we must have

$$a\epsilon_k N = \sum_{j=1}^t L_{ij}\lfloor Nu_j/Q \rfloor + b_i + O_{m,t}(N/Q)$$

for either $a = 1$ or $a = 2$. Dividing by N/Q, we obtain

$$\sum_{j=1}^{t} L_{ij} u_j = a\epsilon_k Q + b_i Q/N + O_{m,t}(1) (\bmod Q).$$

Since $(L_{ij})_{j=1}^t$ is non-zero, the number of t-tuples $(u_1, \ldots, u_t)$ which satisfy this equation is at most $O_{m,t}(Q^{t-1})$. Letting a and i vary we thus see that the proportion of non-nice t-tuples is at most $O_{m,t}(1/Q)$ as desired (the m and t dependence is irrelevant since both are functions of k). □

In a short while we will use Proposition 9.6 to show that ν satisfies the correlation condition (Definition 3.2). Prior to that, however, we must look at the average size of the "arithmetic" factor $\prod_{p|\Delta}(1 + O_m(p^{-1/2}))$ appearing in that proposition.

Lemma 9.9. *Let $m \geqslant 1$ be a parameter. There is a weight function $\tau = \tau_m : \mathbb{Z} \to \mathbb{R}^+$ such that $\tau(n) \geqslant 1$ for all $n \neq 0$, and such that for all distinct $h_1, \ldots, h_j \in [\epsilon_k N, 2\epsilon_k N]$ we have*

$$\prod_{p|\Delta}(1 + O_m(p^{-1/2})) \leqslant \sum_{1 \leqslant i < j \leqslant m} \tau(h_i - h_j),$$

where Δ is defined in Proposition 9.6, and such that $\mathbb{E}(\tau^q(n) | 0 < |n| \leqslant N) = O_{m,q}(1)$ for all $0 < q < \infty$.

Proof. We observe that

$$\prod_{p|\Delta}(1 + O_m(p^{-1/2})) \leqslant \prod_{1 \leqslant i < j \leqslant m} \Bigg(\prod_{p | h_i - h_j} (1 + p^{-1/2}) \Bigg)^{O_m(1)}.$$

By the arithmetic mean-geometric mean inequality (absorbing all constants into the $O_m(1)$ factor) we can thus take $\tau_m(n) := O_m(1) \prod_{p|n}(1 + p^{-1/2})^{O_m(1)}$ for all $n \neq 0$. (The value of τ at 0 is irrelevant for this lemma since we are taking all the h_i to be distinct). To prove the claim, it thus suffices to show that

$$\mathbb{E}\Bigg(\prod_{p|n}(1 + p^{-1/2})^{O_m(q)} \,\Bigg|\, 0 < n \leqslant N \Bigg) = O_{m,q}(1) \text{ for all } 0 < q < \infty.$$

Since $(1 + p^{-1/2})^{O_m(q)}$ is bounded by $1 + p^{-1/4}$ for all but $O_{m,q}(1)$ many primes p, we have

$$\mathbb{E}\Bigg(\prod_{p|n}(1 + p^{-1/2})^{O_m(q)} \,\Bigg|\, 0 < n \leqslant N \Bigg) \leqslant O_{m,q}(1) \mathbb{E}\Bigg(\prod_{p|n}(1 + p^{-1/4}) \,\Bigg|\, 0 < n \leqslant N \Bigg).$$

But $\prod_{p|n}(1 + p^{-1/4}) \leqslant \sum_{d|n} d^{-1/4}$, and hence

$$\begin{aligned} \mathbb{E}\Bigg(\prod_{p|n}(1 + p^{-1/2})^{O_m(q)} \,\Bigg|\, 0 < n \leqslant N \Bigg) &\leqslant O_{m,q}(1) \frac{1}{N} \sum_{n=1}^{N} \sum_{d|n} d^{-1/4} \\ &\leqslant O_{m,q}(1) \frac{1}{N} \sum_{d=1}^{N} \frac{N}{d} d^{-1/4}, \end{aligned}$$

which is $O_{m,q}(1)$ as desired. □

We are now ready to verify the correlation condition.

Proposition 9.10. *The measure ν satisfies the 2^{k-1}-correlation condition.*

Proof. Let us begin by recalling what it is we wish to prove. For any $1 \leqslant m \leqslant 2^{k-1}$ and $h_1, \ldots, h_m \in \mathbb{Z}_N$ we must show a bound

$$\mathbb{E}\big(\nu(x+h_1)\nu(x+h_2)\ldots\nu(x+h_m) \bigm| x \in \mathbb{Z}_N\big) \leqslant \sum_{1 \leqslant i < j \leqslant m} \tau(h_i - h_j), \qquad (9.3)$$

where the weight function $\tau = \tau_m$ is bounded in L^q for all q.

Fix m, $h_1, \ldots, h_m$. We shall take the weight function constructed in Lemma 9.9 (identifying $\mathbb{Z}_N$ with the integers between $-N/2$ and $+N/2$), and set

$$\tau(0) := \exp(Cm \log N / \log\log N)$$

for some large absolute constant C. From the previous lemma we see that $\mathbb{E}(\tau^q) = O_{m,q}(1)$ for all q, since the addition of the weight $\tau(0)$ at 0 only contributes $o_{m,q}(1)$ at most.

We first dispose of the easy case when at least two of the h_i are equal. In this case we bound the left-hand side of (9.2) crudely by $\|\nu\|_{L^\infty}^m$. But from Definitions 9.2, 9.3 and by standard estimates for the maximal order of the divisor function $d(n)$ we have the crude bound $\|\nu\|_{L^\infty} \ll \exp(C \log N / \log\log N)$, and the claim follows thanks to our choice of $\tau(0)$.

Suppose then that the h_i are distinct. Since, in (9.3), our aim is only to get an upper bound, there is no need to subdivide $\mathbb{Z}_N$ into intervals as we did in the proof of Proposition 9.8. Write

$$g(n) := \frac{\phi(W)}{W} \frac{\Lambda_R^2(Wn+1)}{\log R} \mathbf{1}_{[\epsilon_k N, 2\epsilon_k N]}(n).$$

Then by construction of ν (Definition 9.3), we have

$$\mathbb{E}\big(\nu(x+h_1)\ldots\nu(x+h_m) \bigm| x \in \mathbb{Z}_N\big)$$
$$\leqslant \mathbb{E}\big((1+g(x+h_1))\ldots(1+g(x+h_m)) \bigm| x \in \mathbb{Z}_N\big).$$

The right-hand side may be rewritten as

$$\sum_{A \subseteq \{1,\ldots,m\}} \mathbb{E}\Big(\prod_{i \in A} g(x+h_i) \Bigm| x \in \mathbb{Z}_N\Big)$$

(cf. the proof of Lemma 3.4). Observe that for $i, j \in A$ we may assume $|h_i - h_j| \leqslant \epsilon_k N$, since the expectation vanishes otherwise. By Proposition 9.6 and Lemma 9.9, we therefore have

$$\mathbb{E}\Big(\prod_{i \in A} g(x+h_i) \Bigm| x \in \mathbb{Z}_N\Big) \leqslant \sum_{1 \leqslant i < j \leqslant m} \tau(h_i - h_j) + o_m(1).$$

Summing over all A, and adjusting the weights τ by a bounded factor (depending only on m and hence on k), we obtain the result. □

Proof of Proposition 9.1. This is immediate from Lemma 9.4, Lemma 9.7, Proposition 9.8, Proposition 9.10 and the definition of k-pseudorandom measure, which is Definition 3.3. □

10. Correlation estimates for Λ_R

To conclude the proof of Theorem 1.1 it remains to verify Propositions 9.5 and 9.6. That will be achieved in this section, assuming an estimate (Lemma 10.4) for a certain contour integral involving the ζ-function. The proof of that estimate is given in [14], and will be repeated in the Appendix for sake of completeness. The techniques of this section are also rather close to those in [14]. We are greatly indebted to Dan Goldston for sharing this preprint with us.

The linear forms condition for Λ_R. We begin by proving Proposition 9.5. Recall that for each $1 \leqslant i \leqslant m$ we have a linear form $\psi_i(\mathbf{x}) = \sum_{j=1}^t L_{ij}x_j + b_j$ in t variables $x_1, \dots, x_t$. The coefficients L_{ij} satisfy $|L_{ij}| \leqslant \sqrt{w(N)}/2$, where $w(N)$ is the function, tending to infinity with N, which we used to set up the W-trick. We assume that none of the t-tuples $(L_{ij})_{j=1}^t$ are zero or are rational multiples of any other. Define $\theta_i := W\psi_i + 1$.

Let $B := \prod_{j=1}^t I_j$ be a product of intervals I_j, each of length at least R^{10m}. We wish to prove the estimate

$$\mathbb{E}\big(\Lambda_R(\theta_1(\mathbf{x}))^2 \dots \Lambda_R(\theta_m(\mathbf{x}))^2 \bigm| \mathbf{x} \in B\big) = (1 + o_{m,t}(1)) \left(\frac{W \log R}{\phi(W)}\right)^m.$$

The first step is to eliminate the role of the box B. We can use Definition 9.2 to expand the left-hand side as

$$\mathbb{E}\bigg(\prod_{i=1}^m \sum_{\substack{d_i, d_i' \leqslant R \\ d_i, d_i' | \theta_i(\mathbf{x})}} \mu(d_i)\mu(d_i') \log\frac{R}{d_i} \log\frac{R}{d_i'} \biggm| \mathbf{x} \in B\bigg)$$

which we can rearrange as

$$\sum_{d_1,\dots,d_m,d_1',\dots,d_m' \leqslant R} \bigg(\prod_{i=1}^m \mu(d_i)\mu(d_i') \log\frac{R}{d_i} \log\frac{R}{d_i'}\bigg) \mathbb{E}\bigg(\prod_{i=1}^m \mathbf{1}_{d_i, d_i' | \theta_i(\mathbf{x})} \biggm| \mathbf{x} \in B\bigg). \qquad (10.1)$$

Because of the presence of the Möbius functions we may assume that all the d_i, d_i' are square-free. Write $D := [d_1, \dots, d_m, d_1', \dots, d_m']$ to be the least common multiple of the d_i and d_i', thus $D \leqslant R^{2m}$. Observe that the expression $\prod_{i=1}^m \mathbf{1}_{d_i, d_i' | \theta_i(\mathbf{x})}$ is periodic with period D in each of the components of $\mathbf{x}$, and can thus can be safely defined on $\mathbb{Z}_D^t$. Since B is a product of intervals of length at least R^{10m}, we thus see that

$$\mathbb{E}\bigg(\prod_{i=1}^m \mathbf{1}_{d_i, d_i' | \theta_i(\mathbf{x})} \biggm| \mathbf{x} \in B\bigg) = \mathbb{E}\bigg(\prod_{i=1}^m \mathbf{1}_{d_i, d_i' | \theta_i(\mathbf{x})} \biggm| \mathbf{x} \in \mathbb{Z}_D^t\bigg) + O_{m,t}(R^{-8m}).$$

The contribution of the error term $O_m(R^{-8m})$ to (10.1) can be crudely estimated by $O_{m,t}(R^{-6m}\log^{2m} R)$, which is easily acceptable. Our task is thus to show that

$$\sum_{d_1,\dots,d_m,d'_1,\dots,d'_m \leqslant R} \Big(\prod_{i=1}^m \mu(d_i)\mu(d'_i)\log\frac{R}{d_i}\log\frac{R}{d'_i}\Big)\mathbb{E}\Big(\prod_{i=1}^m \mathbf{1}_{d_i,d'_i|\theta_i(\mathbf{x})}\Bigm| \mathbf{x}\in\mathbb{Z}_D^t\Big)$$
$$= (1+o_{m,t}(1))\left(\frac{W\log R}{\phi(W)}\right)^m. \qquad (10.2)$$

To prove (10.2), we shall perform a number of standard manipulations (as in [14]) to rewrite the left-hand side as a contour integral of an Euler product, which in turn can be rewritten in terms of the Riemann ζ-function and some other simple factors. We begin by using the Chinese remainder theorem (and the square-free nature of d_i, d'_i) to rewrite

$$\mathbb{E}\Big(\prod_{i=1}^m \mathbf{1}_{d_i,d'_i|\theta_i(\mathbf{x})}\Bigm| \mathbf{x}\in\mathbb{Z}_D^t\Big) = \prod_{p|D}\mathbb{E}\Big(\prod_{i:p|d_id'_i}\mathbf{1}_{\theta_i(\mathbf{x})=0(\mathrm{mod}\,p)}\Bigm| \mathbf{x}\in\mathbb{Z}_p^t\Big).$$

Note that the restriction that p divides D can be dropped since the multiplicand is 1 otherwise. In particular, if we write $X_{d_1,\dots,d_m}(p) := \{1\leqslant i\leqslant m : p|d_i\}$ and

$$\omega_X(p) := \mathbb{E}\Big(\prod_{i\in X}\mathbf{1}_{\theta_i(\mathbf{x})=0(\mathrm{mod}\,p)}\Bigm| \mathbf{x}\in\mathbb{Z}_p^t\Big) \qquad (10.3)$$

for each subset $X\subseteq\{1,\dots,m\}$, then we have

$$\mathbb{E}\Big(\prod_{i=1}^m \mathbf{1}_{d_i,d'_i|\theta_i(\mathbf{x})}\Bigm| \mathbf{x}\in\mathbb{Z}_D^t\Big) = \prod_p \omega_{X_{d_1,\dots,d_m}(p)\cup X_{d'_1,\dots,d'_m}(p)}(p).$$

We can thus write the left-hand side of (10.2) as

$$\sum_{d_1,\dots,d_m,d'_1,\dots,d'_m\in\mathbb{Z}^+}\Big(\prod_{i=1}^m \mu(d_i)\mu(d'_i)(\log\frac{R}{d_i})_+(\log\frac{R}{d'_i})_+\Big)\prod_p \omega_{X_{d_1,\dots,d_m}(p)\cup X_{d'_1,\dots,d'_m}(p)}(p).$$

To proceed further, we need to express the logarithms in terms of multiplicative functions of the d_i, d'_i. To this end, we introduce the vertical line contour Γ_1 parameterized by

$$\Gamma_1(t) := \frac{1}{\log R} + it; \quad -\infty < t < +\infty \qquad (10.4)$$

and observe the contour integration identity

$$\frac{1}{2\pi i}\int_{\Gamma_1}\frac{x^z}{z^2}\,dz = (\log x)_+$$

valid for any real $x>0$. The choice of $\frac{1}{\log R}$ for the real part of Γ_1 is not currently relevant, but will be convenient later when we estimate the contour integrals that emerge (in particular, R^z is bounded on Γ_1, while $1/z^2$ is not too large). Using this identity, we can rewrite the left-hand side of (10.2) as

$$(2\pi i)^{-2m}\int_{\Gamma_1}\dots\int_{\Gamma_1} F(z,z')\prod_{j=1}^m \frac{R^{z_j+z'_j}}{z_j^2 z_j'^2}\,dz_j dz'_j \qquad (10.5)$$

where there are $2m$ contour integrations in the variables $z_1, \ldots, z_m, z'_1, \ldots, z'_m$ on Γ_1, $z := (z_1, \ldots, z_m)$ and $z' := (z'_1, \ldots, z'_m)$, and

$$F(z, z') := \sum_{d_1,\ldots,d_m,d'_1,\ldots,d'_m \in \mathbb{Z}^+} \Big(\prod_{j=1}^{m} \frac{\mu(d_j)\mu(d'_j)}{d_j^{z_j} d_j'^{z'_j}} \Big) \prod_p \omega_{X_{d_1,\ldots,d_m}(p) \cup X_{d'_1,\ldots,d'_m}(p)}(p). \quad (10.6)$$

We have changed the indices from i to j to avoid conflict with the square root of -1. Observe that the summand in (10.6) is a multiplicative function of the d_j and d'_j, and thus we have (formally, at least) the Euler product representation $F(z, z') = \prod_p E_p(z, z')$, where

$$E_p(z, z') := \sum_{X, X' \subseteq \{1,\ldots,m\}} \frac{(-1)^{|X|+|X'|} \omega_{X \cup X'}(p)}{p^{\sum_{j \in X} z_j + \sum_{j \in X'} z'_j}}. \quad (10.7)$$

From (10.3) we have $\omega_\emptyset(p) = 1$ and $\omega_X(p) \leqslant 1$, and so $E_p(z, z') = 1 + O_\sigma(1/p^\sigma)$ when $\Re(z_j), \Re(z'_j) > \sigma$ (we obtain more precise estimates below). Thus this Euler product is absolutely convergent to $F(z, z')$ in the domain $\Re(z_j), \Re(z'_j) > 1$ at least.

To proceed further we need to exploit the hypothesis that the linear parts of $\psi_1, \ldots, \psi_m$ are non-zero and not rational multiples of each other. This shall be done via the following elementary estimates on $\omega_X(p)$.

Lemma 10.1. *If $p \leqslant w(N)$, then $\omega_X(p) = 0$ for all non-empty X; in particular, $E_p = 1$ when $p \leqslant w(N)$. If instead $p > w(N)$, then $\omega_X(p) = p^{-1}$ when $|X| = 1$ and $\omega_X(p) \leqslant p^{-2}$ when $|X| \geqslant 2$.*

Proof. The first statement is clear, since the maps $\theta_j : \mathbb{Z}_p^t \to \mathbb{Z}_p$ are identically 1 when $p \leqslant w(N)$. The second statement (when $p > w(N)$ and $|X| = 1$) is similar since in this case θ_j uniformly covers $\mathbb{Z}_p$. Now suppose $p > w(N)$ and $|X| = 2$. We claim that none of the s pure linear forms $W(\psi_i - b_i)$ is a multiple of any other $(\bmod\, p)$. Indeed, if this were so then we should have $L_{ij} L_{i'j}^{-1} \equiv \lambda (\bmod\, p)$ for some λ, and for all $j = 1, \ldots, t$. But if a/q and a'/q' are two rational numbers in lowest terms, with $q, q' < \sqrt{w(N)}/2$, then clearly $a/q \not\equiv a'/q' (\bmod\, p)$ unless $a = a'$, $q = q'$. It follows that the two pure linear forms $\psi_i - b_i$ and $\psi_{i'} - b_{i'}$ are rational multiples of one another, contrary to assumption. Thus the set of $\mathbf{x} \in (\mathbb{Z}/p\mathbb{Z})^t$ for which $\theta_i(\mathbf{x}) \equiv 0 (\bmod\, p)$ for all $i \in X$ is contained in the intersection of two skew affine subspaces of $(\mathbb{Z}/p\mathbb{Z})^t$, and as such has cardinality at most p^{t-2}. □

This lemma implies, comparing with (10.7), that

$$E_p(z, z') = 1 - \mathbf{1}_{p > w(N)} \sum_{j=1}^{m} (p^{-1-z_j} + p^{-1-z'_j} - p^{-1-z_j-z'_j})$$

$$+ \mathbf{1}_{p > w(N)} \sum_{\substack{X, X' \subseteq \{1,\ldots,m\} \\ |X \cup X'| \geqslant 2}} \frac{O(1/p^2)}{p^{\sum_{j \in X} z_j + \sum_{j \in X'} z'_j}}, \quad (10.8)$$

where the $O(1/p^2)$ numerator does not depend on z, z'. To take advantage of this expansion, we factorize $E_p = E_p^{(1)} E_p^{(2)} E_p^{(3)}$, where

$$E_p^{(1)}(z,z') := \frac{E_p(z,z')}{\prod_{j=1}^m (1 - \mathbf{1}_{p>w(N)} p^{-1-z_j})(1 - \mathbf{1}_{p>w(N)} p^{-1-z'_j})(1 - \mathbf{1}_{p>w(N)} p^{-1-z_j-z'_j})^{-1}}$$

$$E_p^{(2)}(z,z') := \prod_{j=1}^m (1 - \mathbf{1}_{p\leqslant w(N)} p^{-1-z_j})^{-1}(1 - \mathbf{1}_{p\leqslant w(N)} p^{-1-z'_j})^{-1}(1 - \mathbf{1}_{p\leqslant w(N)} p^{-1-z_j-z'_j})$$

$$E_p^{(3)}(z,z') := \prod_{j=1}^m (1 - p^{-1-z_j})(1 - p^{-1-z'_j})(1 - p^{-1-z_j-z'_j})^{-1}.$$

Writing $G_j := \prod_p E_p^{(j)}$, one thus has $F = G_1 G_2 G_3$ (at least for $\Re(z_j), \Re(z'_j)$ sufficiently large). If we introduce the Riemann ζ-function $\zeta(s) := \prod_p (1 - \frac{1}{p^s})^{-1}$ then we have

$$G_3(z,z') = \prod_{j=1}^m \frac{\zeta(1 + z_j + z'_j)}{\zeta(1 + z_j)\zeta(1 + z'_j)} \tag{10.9}$$

so in particular G_3 can be continued meromorphically to the entire complex plane. As for the other two factors, we have the following estimates which allow us to continue these factors a little bit to the left of the imaginary axes.

Definition 10.2. For any $\sigma > 0$, let $\mathcal{D}^m_\sigma \subseteq \mathbb{C}^{2m}$ denote the domain

$$\mathcal{D}^m_\sigma := \{z_j, z'_j : -\sigma < \Re(z_j), \Re(z'_j) < 100, j = 1, \ldots, m\}.$$

If $G = G(z,z')$ is an analytic function of $2m$ complex variables on $\mathcal{D}^m_\sigma$, we define the $C^k(\mathcal{D}^m_\sigma)$ norm of G for any integer $k \geqslant 0$ as

$$\|G\|_{C^k(\mathcal{D}^m_\sigma)} := \sup_{a_1+\ldots+a_m+a'_1+\ldots+a'_m \leqslant k} \Big\| \Big(\frac{\partial}{\partial z_1}\Big)^{a_1} \ldots \Big(\frac{\partial}{\partial z_m}\Big)^{a_m} \Big(\frac{\partial}{\partial z'_1}\Big)^{a'_1} \ldots \Big(\frac{\partial}{\partial z'_m}\Big)^{a'_m} G \Big\|_{L^\infty(\mathcal{D}^m_\sigma)}$$

where $a_1, \ldots, a'_m$ ranges over all non-negative integers with total sum at most k.

Lemma 10.3. *The Euler products $\prod_p E_p^{(j)}$ for $j = 1, 2$ are absolutely convergent in the domain $\mathcal{D}^m_{1/6m}$. In particular, G_1, G_2 can be continued analytically to this domain. Furthermore, we have the estimates*

$$\|G_1\|_{C^m(\mathcal{D}^m_{1/6m})} \leqslant O_m(1)$$
$$\|G_2\|_{C^m(\mathcal{D}^m_{1/6m})} \leqslant C(m, w(N))$$
$$G_1(0,0) = 1 + o_m(1)$$
$$G_2(0,0) = (W/\phi(W))^m.$$

Remark. The choice $\sigma = 1/6m$ is of course not best possible, but in fact any small positive quantity depending on m would suffice for our argument here. The dependence of $C(m, w(N))$ on $w(N)$ is not important, but one can easily obtain (for instance) growth bounds of the form $w(N)^{O_m(w(N))}$.

Proof. First consider $j = 1$. From (10.8) and Taylor expansion we have the crude bound $E_p^{(1)}(z,z') = 1 + O_m(p^{-1-1/6m})$ in $\mathcal{D}^m_{1/6m}$, which gives the desired convergence and also the $C^m(\mathcal{D}^m_{1/6m})$ bound on G_1; the bound on $G_1(0,0)$ also follows since the Euler factors $E_p^{(1)}(z,z')$ are identically 1 when $p \leqslant w(N)$. The bound for G_2 are easy since this is just a finite Euler product involving at most $w(N)$ terms; the formula for $G_2(0,0)$ follows from direct calculation since $\frac{\phi(W)}{W} = \prod_{p<w(n)}(1-\frac{1}{p})$. $\square$

To estimate (10.5), we now invoke the following contour integration lemma.

Lemma 10.4. [14] *Let R be a positive real number. Let $G = G(z, z')$ be an analytic function of $2m$ complex variables on the domain $\mathcal{D}_\sigma^m$ for some $\sigma > 0$, and suppose that*

$$\|G\|_{C^m(\mathcal{D}_\sigma^m)} = \exp(O_{m,\sigma}(\log^{1/3} R)). \tag{10.10}$$

Then

$$\frac{1}{(2\pi i)^{2m}} \int_{\Gamma_1} \cdots \int_{\Gamma_1} G(z,z') \prod_{j=1}^m \frac{\zeta(1+z_j+z_j')}{\zeta(1+z_j)\zeta(1+z_j')} \frac{R^{z_j+z_j'}}{z_j^2 z_j'^2}\, dz_j dz_j'$$

$$= G(0,\ldots,0)\log^m R + \sum_{j=1}^m O_{m,\sigma}(\|G\|_{C^j(\mathcal{D}_\sigma^m)} \log^{m-j} R) + O_{m,\sigma}(e^{-\delta\sqrt{\log R}})$$

for some $\delta = \delta(m) > 0$.

Proof. While this Lemma is essentially in [14], we shall give a complete proof in the Appendix for sake of completeness. $\square$

We apply this lemma with $G := G_1 G_2$ and $\sigma := 1/6m$. From Lemma 10.3 and the Leibnitz rule we have the bounds

$$\|G\|_{C^j(\mathcal{D}_{1/6m}^m)} \leqslant C(j, m, w(N)) \text{ for all } 0 \leqslant j \leqslant m,$$

and in particular we obtain (10.10) by choosing $w(N)$ to grow sufficiently slowly in N. Also we have $G(0,0) = (1+o_m(1))(\frac{W}{\phi(W)})^m$ from that lemma. We conclude (again taking $w(N)$ sufficiently slowly growing in N) that the quantity in (10.5) is $(1+o_m(1))(\frac{W \log R}{\phi(W)})^m$, as desired. This concludes the proof of Proposition 9.5. $\square$

Higher order correlations for Λ_R. We now prove Proposition 9.6, which is proven by a very similar argument to Proposition 9.5. Note that the main differences here are that the number of variables t is just equal to 1, but on the other hand all the linear forms are equal to each other, $\psi_i(x_1) = x_1$. In particular, these linear forms are now rational multiples of each other and so Lemma 10.1 no longer applies. However, the arguments before that Lemma are still valid; thus we can still write the left-hand side of (9.1) as an expression of the form (10.5) plus an acceptable error, where F is again defined by (10.6) and E_p is defined by (10.7); the difference now is that $\omega_X(p)$ is the quantity

$$\omega_X(p) := \mathbb{E}\bigg(\prod_{i \in X} \mathbf{1}_{W(x+h_i)+1=0(\bmod p)} \,\bigg|\, x \in \mathbb{Z}_p\bigg).$$

Again we have $\omega_\emptyset(p) = 1$ for all p. The analogue of Lemma 10.1 is as follows.

Lemma 10.5. *If $p \leqslant w(N)$, then $\omega_X(p) = 0$ for all non-empty X; in particular, $E_p = 1$ when $p \leqslant w(N)$. If instead $p > w(N)$, then $\omega_X(p) = p^{-1}$ when $|X| = 1$ and $\omega_X(p) \leqslant p^{-1}$ when $|X| \geqslant 2$. Furthermore, if $|X| \geqslant 2$ then $\omega_X(p) = 0$ unless p divides $\Delta := \prod_{1 \leqslant i < j \leqslant s} |h_i - h_j|$.*

Proof. When $p \leqslant w(N)$ then $W(x+h_i)+1 \equiv 1 (\operatorname{mod} p)$ and the claim follows. When $p > w(N)$ and $|X| \geqslant 1$, $\omega_X(p)$ is equal to $1/p$ when the residue classes $\{h_i(\operatorname{mod} p) : i \in X\}$ are all equal, and zero otherwise, and the claim again follows. □

In light of this lemma, the analogue of (10.8) is now

$$E_p(z,z') = 1 - \mathbf{1}_{p>w(N)} \sum_{j=1}^m (p^{-1-z_j} + p^{-1-z'_j} - p^{-1-z_j-z'_j}) + \mathbf{1}_{p>w(N),p|\Delta} \lambda_p(z,z') \quad (10.11)$$

where $\lambda_p(z,z')$ is an expression of the form

$$\lambda_p(z,z') = \sum_{\substack{X,X' \subseteq \{1,\dots,m\} \\ |X \cup X'| \geqslant 2}} \frac{O(1/p)}{p^{\sum_{j\in X} z_j + \sum_{j \in X'} z'_j}}$$

and the $O(1/p)$ quantities do not depend on z, z'. We can thus factorize

$$E_p = E_p^{(0)} E_p^{(1)} E_p^{(2)} E_p^{(3)},$$

where

$$E_p^{(0)} = 1 + \mathbf{1}_{p>w(N),p|\Delta} \lambda_p(z,z')$$

$$E_p^{(1)} = \frac{E_p}{E_p^{(0)} \prod_{j=1}^m (1 - \mathbf{1}_{p>w(N)} p^{-1-z_j})(1 - \mathbf{1}_{p>w(N)} p^{-1-z'_j})(1 - \mathbf{1}_{p>w(N)} p^{-1-z_j-z'_j})^{-1}}$$

$$E_p^{(2)} = \prod_{j=1}^m (1 - \mathbf{1}_{p\leqslant w(N)} p^{-1-z_j})^{-1} (1 - \mathbf{1}_{p \leqslant w(N)} p^{-1-z'_j})^{-1} (1 - \mathbf{1}_{p\leqslant w} p^{-1-z_j-z'_j})$$

$$E_p^{(3)} = \prod_{j=1}^m (1 - p^{-1-z_j})(1-p^{-1-z'_j})(1-p^{-1-z_j-z'_j})^{-1}.$$

Write $G_j = \prod_p E_p^{(j)}$. Then, as before, $F = G_0 G_1 G_2 G_3$ and G_3 is given by (10.9) as before. As for G_0, G_1, G_2, we have the following analogue of Lemma 10.3.

Lemma 10.6. *Let $0 < \sigma < 1/6m$. Then the Euler products $\prod_p E_p^{(j)}$ for $j = 0,1,2$ are absolutely convergent in the domain $\mathcal{D}_\sigma^m$. In particular, G_0, G_1, G_2 can be continued analytically to this domain. Furthermore, we have the estimates*

$$\|G_0\|_{C^j(\mathcal{D}^m_{1/6m})} \leqslant O_m\left(\frac{\log R}{\log\log R}\right)^j \prod_{p|\Delta} (1 + O_m(p^{2m\sigma - 1})) \quad \textit{for } 0 \leqslant j \leqslant m \quad (10.12)$$

$$\|G_0\|_{C^m(\mathcal{D}^m_{1/6m})} \leqslant \exp(O_m(\log^{1/3} R)) \quad (10.13)$$

$$\|G_1\|_{C^m(\mathcal{D}^m_{1/6m})} \leqslant O_m(1)$$

$$\|G_2\|_{C^m(\mathcal{D}^m_{1/6m})} \leqslant C(m, w(N))$$

$$G_0(0,0) = \prod_{p|\Delta} (1 + O_m(p^{-1/2})) \quad (10.14)$$

$$G_1(0,0) = 1 + o_m(1)$$

$$G_2(0,0) = (W/\phi(W))^m.$$

Proof. The estimates for G_1 and G_2 proceed exactly as in Lemma 10.3 (the additional factors of $\lambda_p(z,z')$ which appear on both the numerator and denominator of $E_p^{(1)}$ cancel to first order, and thus do not present any new difficulties); it is the estimates for G_0 which are the most interesting.

We begin by proving (10.12). Fix j. First observe that $G_0 = \prod_{p|\Delta} E_p^{(0)}$. Now the number of primes dividing Δ is at most $O(\log\Delta/\log\log\Delta)$. Using the crude bound

$$\Delta = \prod_{1\leqslant i<j\leqslant m} |h_i - h_j| \leqslant N^{m^2} \leqslant R^{O_m(1)}, \tag{10.15}$$

we thus see that the number of factors in the Euler product for G_0 is $O_m(\frac{\log R}{\log\log R})$. Upon differentiating j times for any $0 \leqslant j \leqslant m$ using the Leibnitz rule, one gets a sum of $O_m((\log R/\log\log R)^j)$ terms, each of which consists of $O_m(\log R/\log\log R)$ factors, each of which is equal to some derivative of $1 + \lambda_p(z,z')$ of order between 0 and j. On $\mathcal{D}_\sigma^m$, each factor is bounded by $1 + O_m(p^{-1/2})$ (in fact, the terms containing a non-zero number of derivatives will be much smaller since the constant term 1 is eliminated). This gives (10.12).

Now we prove (10.13). In light of (10.12), it suffices to show that

$$\prod_{p|\Delta}(1 + O_m(p^{2m\sigma-1})) \leqslant \exp(O_m(\log^{1/3} R)).$$

Taking logarithms and using the hypothesis $\sigma < 1/6m$ (and (10.15)), we reduce to showing

$$\sum_{p|\Delta} p^{-2/3} \leqslant O(\log^{1/3}\Delta).$$

But there are at most $O(\log\Delta/\log\log\Delta)$ primes dividing Δ, hence the left-hand side can be crudely bounded by

$$\sum_{1\leqslant n\leqslant O(\log\Delta/\log\log\Delta)} n^{-2/3} = O(\log^{1/3}\Delta)$$

as desired.

The bound (10.14) now follows from the crude estimate $E_p^{(0)}(z,z') = 1 + O_m(p^{-1/2})$. □

We now apply Lemma 10.4 with $\sigma := 1/6m$ and $G := G_0G_1G_2$. Again by the Leibnitz rule we have the bound (10.10), and furthermore

$$\|G\|_{C^j(\mathcal{D}_\sigma^m)} \leqslant O_m(1)C(m,w(N))\left(\frac{\log R}{\log\log R}\right)^j$$

for all $0 \leqslant j \leqslant m$. From Lemma 10.6 and Lemma 10.4 we can then estimate (10.5) as

$$\leqslant O_m\left(\frac{W}{\phi(W)}\right)^m \log^m R \prod_{p|\Delta}(1 + O_m(p^{-1/2})) + C(m,w(N))\frac{\log^m R}{\log\log R} + O_m(e^{-\delta\sqrt{\log R}}).$$

The claim (9.1) then follows by choosing $w(N)$ (and hence W) sufficiently slowly growing in N (and hence in R). Proposition 9.6 follows. □

Remark. It should be clear that the above argument not only gives an upper bound for the left-hand side of (9.1), but in fact gives an asymptotic, by working out $G_0(0,0)$ more carefully; this is worked out in detail (in the $W = 1$ case) in [14].

11. Further remarks

In this section we discuss some extensions and refinements of our main result. First of all, notice that our proof actually shows that that there is some constant $\gamma(k)$ such that the number of k-term progressions of primes, all less than N, is at least $(\gamma(k) + o(1))N^2/\log^k N$. This is because the error term in (3.6) does not actually need to be $o(1)$, but merely less than $\frac{1}{2}c(k,\delta) + o(1)$ (for instance). Working backwards through the proof, this eventually reveals that the quantity $w(N)$ does not actually need to be growing in N, but can instead be a fixed number depending only on k (although this number will be very large because our final bounds $o(1)$ decayed to zero extremely slowly). Thus W can be made independent of N, and so the loss incurred by the W-trick when passing from primes to primes equal to 1 mod W is bounded uniformly in N. Nevertheless the bound we obtain on $\gamma(k)$ is extremely poor, in part because of the growth of constants in the best known bounds $c(k,\delta)$ on Szemerédi's theorem in [16], but also because we have not attempted to optimize the decay rate of the $o(1)$ factors and hence will need to take $w(N)$ to be extremely large. In the other direction, standard sieve theory arguments show that the number of k-term progressions of primes all less than N are at most $O_k(N^2/\log^k N)$, and so the lower bounds are only off by a constant depending on k.

As we remarked earlier, our method also extends to prove Theorem 1.2, namely that any subset of the primes with positive relative upper density contains a k-term arithmetic progression. The only significant change[18] to the proof is that one must use the pigeonhole principle to replace the residue class $n \equiv 1(\operatorname{mod} W)$ by a more general residue class $n \equiv b(\operatorname{mod} W)$ for some b coprime to W, since the set A in Theorem 1.2 does not need to obey a Dirichlet-type theorem in these residue classes. However it is easy to verify that this does not significantly affect the rest of the argument, and we leave the details to the reader.

Applying Theorem 1.2 to the set of primes $p \equiv 1(\operatorname{mod} 4)$, we obtain the previously unknown fact that there are arbitrarily long progressions consisting of numbers which are the sum of two squares. For this problem, more satisfactory results were known for small k than was the case for the primes. Let S be the set of sums of two squares. It is a simple matter to show that there are infinitely many 4-term arithmetic progressions in S. Indeed, Heath-Brown [23] observed that the numbers $(n-1)^2+(n-8)^2$, $(n-7)^2+(n+4)^2$, $(n+7)^2+(n-4)^2$ and $(n+1)^2+(n+8)^2$ always form such a progression; in fact, he was able to prove much more, in particular finding an asymptotic for the number of 4-term progressions in S, all of whose members are at most N.

[18]Also, since we are only assuming positivity of the upper density and not the lower density, we only have good density control for an infinite sequence $N_1, N_2, \ldots \to \infty$ of integers, which may not be prime. However one can easily use Bertrands postulate (for instance) to make the N_j prime, giving up a factor of $O(1)$ at most.

It is reasonably clear that our method will produce long arithmetic progressions for many sets of primes for which one can give a lower bound which agrees with some upper bound coming from a sieve, up to a multiplicative constant. Invoking Chen's famous theorem to the effect that there are $\gg N/\log^2 N$ primes $p \leqslant N$ for which $p+2$ is a prime or a product of two primes, it ought to be a simple matter to adapt our arguments to show that there are arbitrarily long arithmetic progressions $p_1, \ldots, p_k$ of primes, such that each $p_i + 2$ is either prime or the product of two primes; indeed there should be $N/\log^{2k} N$ such progressions with entries less than N. Whilst we do not plan to write a detailed proof of this fact, we will in [20] give a proof of the case $k = 3$ using harmonic analysis.

Another possible extension, which would require more significant modification to our argument, would be a Bergelson-Leibman type result (cf. [4]) for primes. That is, one could hope to show that if $F_i : \mathbb{N} \to \mathbb{N}$ are polynomials with $F(0) = 0$, then there are infinitely many configurations $(a + F_1(d), \ldots, a + F_k(d))$ in which all k elements are prime. We may address this issue at greater length in a future paper.

Appendix: Proof of Lemma 10.4

In this appendix we prove Lemma 10.4. This Lemma was essentially proven in [14], but for the sake of self-containedness we provide a complete proof here (following very closely the approach in [14]).

Throughout this section, $R \geqslant 2$, $m \geqslant 1$, and $\sigma > 0$ will be fixed. We shall use $\delta > 0$ to denote various small constants, which may vary from line to line (the previous interpretation of δ as the average value of a function f will now be irrelevant). We begin by recalling the classical zero-free region for the Riemann ζ function.

Lemma 11.1 (Zero-free region)**.** *Define the classical zero free region $\mathcal{Z}$ to be the closed region*

$$\mathcal{Z} := \{s \in \mathbb{C} : 10 \geqslant \Re s \geqslant 1 - \frac{\beta}{\log(|\Im s| + 2)}\}$$

for some small $0 < \beta < 1$. Then if β is sufficiently small, ζ is non-zero and meromorphic in $\mathcal{Z}$ with a simple pole at 1 and no other singularities. Furthermore we have the bounds

$$\zeta(s) - \frac{1}{s-1} = O(\log(|\Im s| + 2)); \quad \frac{1}{\zeta(s)} = O(\log(|\Im s| + 2))$$

for all $s \in \mathcal{Z}$.

Proof. See Titchmarsh [36, Chapter 3]. □

Fix β in the above lemma; we may take β to be small enough that $\mathcal{Z}$ is contained in the region where $1 - \sigma < \Re(s) < 101$. We will allow all our constants in the $O()$ notation to depend on β and σ, and omit explicit mention of these dependencies from our subscripts.

In addition to the contour Γ_1 defined in (10.4), we will need the two further contours Γ_0 and Γ_2, defined by

$$\begin{aligned} \Gamma_0(t) &:= -\frac{\beta}{\log(|t|+2)} + it, \qquad -\infty < t < \infty \\ \Gamma_2(t) &:= 1 + it, \quad -\infty < t < \infty. \end{aligned} \tag{11.1}$$

Thus Γ_0 is the left boundary of $\mathcal{Z} - 1$ (which thus lies to the left of the origin), while Γ_1 and Γ_2 are vertical lines to the right of the origin. The usefulness of Γ_2 for us lies in the simple observation that $\zeta(1 + z + z')$ has no poles when $z \in \mathcal{Z} - 1$ and $z' \in \Gamma_2$, but we will not otherwise attempt to estimate any integrals on Γ_2.

We observe the following elementary integral estimates.

Lemma 11.2. *Let B be some fixed constant. Then we have the bounds.*

$$\int_{\Gamma_0} \log^B(|z|+2) \left|\frac{R^z dz}{z^2}\right| \leqslant O_B(e^{-\delta\sqrt{\log R}}); \tag{11.2}$$

$$\int_{\Gamma_1} \log^B(|z|+2) \left|\frac{R^z dz}{z^2}\right| \leqslant O_B(\log R). \tag{11.3}$$

Here $\delta = \delta(B, \beta) > 0$ is a constant independent of R.

Proof. We first bound the left-hand side of (11.2). Substitute in the parametrisation (11.1). Since $\Gamma_0'(t) = O(1)$ and $|z| \gg |t| + \beta$ we have, for any $T \geqslant 2$,

$$\begin{aligned} \int_{\Gamma_0} \log^B(|z|+2) \left|\frac{R^z dz}{z^2}\right| &\leqslant O_B(1) \int_0^\infty R^{-\beta/(\log(|t|+2))} \frac{\log^B(|t|+2)}{(|t|+\beta)^2} dt \\ &\leqslant O_B(1)(\log^B T \int_0^T R^{-\beta/\log(t+2)} dt + \int_T^\infty \frac{\log^B t}{t^2} dt) \\ &\leqslant O_B(1)(T \log^B T \exp(-\beta \log R/\log T) + T^{-1}\log^B T). \end{aligned}$$

Choosing $T = \exp(\frac{1}{2}\sqrt{\beta \log R})$, so that the two expressions here are equal, one sees that this is bounded above by

$$(\beta \log R)^{B/2} \exp(-\tfrac{1}{2}\sqrt{\beta \log R}) = O_B(e^{-\delta\sqrt{\log R}}).$$

The bound (11.3) is much simpler, and can be obtained by noting that R^z is bounded on Γ_1, and substituting in (10.4) splitting the integrand up into the ranges $|t| \leqslant 1/\log R$ and $|t| > 1/\log R$. □

The next lemma is closely related to the case $m = 1$ of Lemma 10.4.

Lemma 11.3. *Let $f(z, z')$ be analytic in $\mathcal{D}^1_\sigma$ and suppose that*

$$|f(z, z')| \leqslant \exp(O_m(\log^{1/3} R))$$

uniformly in this domain. Then the integral

$$I := \frac{1}{(2\pi i)^2} \int_{\Gamma_1}\int_{\Gamma_1} f(z, z') \frac{\zeta(1+z+z')}{\zeta(1+z)\zeta(1+z')} \frac{R^{z+z'}}{z^2 z'^2} dz dz'$$

obeys the estimate

$$I = f(0,0)\log R + \frac{\partial f}{\partial z'}(0,0) + \frac{1}{2\pi i}\int_{\Gamma_1} f(z,-z) \frac{dz}{\zeta(1+z)\zeta(1-z)z^4} + O_m(e^{-\delta\sqrt{\log R}})$$

for some $\delta = \delta(\sigma, \beta) > 0$ independent of R.

Proof. We observe from Lemma 11.1 that we have enough decay of the integrand in the domain $\mathcal{D}^1_\sigma$ to interchange the order of integration, and to shift contours in either one of the variables z, z' while keeping the other fixed, without any difficulties when $\Im(z), \Im(z') \to \infty$; the only issue is to keep track of when the contour passes through a pole of the integrand. In particular we can shift the z' contour from Γ_1 to Γ_2, since we do not encounter any of the poles of the integrand while doing so. Let us look at the integrand for each fixed $z' \in \Gamma_2$, viewing it as an analytic function of z. We now attempt to shift the z contour of integration to Γ_0. In so doing the contour passes just one pole, a simple one at $z = 0$. The residue there is $\frac{1}{2\pi i}\int_{\Gamma_2} f(0,z')\frac{R^{z'}}{z'^2}dz'$, and so we have $I = I_1 + I_2$, where

$$I_1 := \frac{1}{2\pi i}\int_{\Gamma_2} f(0,z')\frac{R^{z'}}{z'^2}dz'$$
$$I_2 := \frac{1}{(2\pi i)^2}\int_{\Gamma_2}\int_{\Gamma_0} f(z,z')\frac{\zeta(1+z+z')R^{z+z'}}{\zeta(1+z)\zeta(1+z')z^2z'^2}dzdz'.$$

To evaluate I_1, we shift the z' contour of integration to Γ_0. Again there is just one pole, a double one at $z' = 0$. The residue there is $f(0,0)\log R + \frac{\partial f}{\partial z'}(0,0)$, and so

$$\begin{aligned} I_1 &= f(0,0)\log R + \frac{\partial f}{\partial z'}(0,0) + \frac{1}{2\pi i}\int_{\Gamma_1} f(0,z')\frac{R^{z'}}{z'^2}dz' \\ &= f(0,0)\log R + \frac{\partial f}{\partial z'}(0,0) + O_m(e^{-\delta\sqrt{\log R}}), \end{aligned}$$

for some $\delta > 0$, the latter step being a consequence of our bound on f and (11.2) (in the case $B = 0$).

To estimate I_2, we first swap the order of integration and, for each fixed z, view the integrand as an analytic function of z'. We move the z' contour from Γ_2 to Γ_0, this again being allowed since we have sufficient decay in vertical strips as $|\Im z'| \to \infty$. In so doing we pass exactly two simple poles, at $z' = -z$ and $z' = 0$. The residue at the first is exactly

$$\frac{1}{2\pi i}\int_{\Gamma_0} f(z,-z)\frac{dz}{\zeta(1+z)\zeta(1-z)z^4},$$

which is one of the terms appearing in our formula for I.

The residue at $z' = 0$ is

$$\int_{\Gamma_0} f(z,0)\frac{R^z}{z^2}dz,$$

which is $O(e^{-\delta\sqrt{\log R}})$ for some $\delta > 0$ by (11.2). The value of I_2 is the sum of these two quantities and the integral over the new contour Γ_0, which is

$$\int_{\Gamma_0}\int_{\Gamma_0} f(z,z')\frac{\zeta(1+z+z')R^{z+z'}}{\zeta(1+z)\zeta(1+z')z^2z'^2}dzdz'. \qquad (11.4)$$

In this integrand we have $|f| = \exp(O_m(\log^{1/3} R))$, and also the portion involving the three ζs is $O(1)\log^2(|\Im z| + 2)\log^2(|\Im z'| + 2)$ by Lemma 11.1. Using (11.2) twice it follows that

$$(11.4) = O_m(e^{-\delta\sqrt{\log R}})$$

for some $\delta > 0$.

Thus we now have estimates for I_1 and I_2 up to errors of $O_m(e^{-\delta\sqrt{\log R}})$. Putting all of this together completes the proof of the lemma. □

Proof of Lemma 10.4. Let $G = G(z, z')$ be an analytic function of $2m$ complex variables on the domain $\mathcal{D}_\sigma^m$ obeying the derivative bounds (10.10). We will allow all our implicit constants in the $O()$ notation to depend on m, β, σ. We are interested in the integral

$$I(G, m) := \frac{1}{(2\pi i)^{2m}} \int_{\Gamma_1} \cdots \int_{\Gamma_1} G(z_j, z_j') \prod_{j=1}^{s} \frac{\zeta(1 + z_j + z_j')}{\zeta(1 + z_j)\zeta(1 + z_j')} \frac{R^{z_j + z_j'}}{z_j^2 z_j'^2} \, dz_j dz_j',$$

and wish to prove the estimate

$$I(G, m) := G(0, \ldots, 0)(\log R)^m + \sum_{j=1}^{m} O(\|G\|_{C^j(\mathcal{D}_\sigma^s)}(\log R)^{m-j}) + O(e^{-\delta\sqrt{\log R}}).$$

The proof is by induction on m. The case $m = 1$ is a swift deduction from Lemma 11.3, the only issue being an estimation of the term

$$\frac{1}{2\pi i} \int_{\Gamma_0} G(z_1, -z_1) \frac{dz_1}{\zeta(1 + z_1)\zeta(1 - z_1) z_1^4}.$$

It is not hard to check (using Lemma 11.1) that

$$\int_{\Gamma_0} \left| \frac{dz_1}{\zeta(1 + z_1)\zeta(1 - z_1) z_1^4} \right| = O(1), \tag{11.5}$$

and so this term is $O(\sup_{z \in \mathcal{D}_\sigma^1} |G(z)|) = O(\|G\|_{C^1(\mathcal{D}_\sigma^1)})$.

Suppose then that we have established the result for $m \geqslant 1$ and wish to deduce it for $m + 1$. Applying Lemma 11.3 in the variables z_{m+1}, z_{m+1}', we get $I(G, m + 1) =$

$$\begin{aligned}
&\frac{\log R}{(2\pi i)^{2m}} \int_{\Gamma_1} \cdots \int_{\Gamma_1} G(z_1, \ldots, z_m, 0, z_1', \ldots, z_m', 0) \prod_{j=1}^{m} \frac{\zeta(1 + z_j + z_j')}{\zeta(1 + z_j)\zeta(1 + z_j')} \frac{R^{z_j + z_j'}}{z_j^2 z_j'^2} dz_j dz_j' \\
&+ \frac{1}{(2\pi i)^{2m}} \int_{\Gamma_1} \cdots \int_{\Gamma_1} H(z_1, \ldots, z_m, z_1', \ldots, z_m') \prod_{j=1}^{m} \frac{\zeta(1 + z_j + z_j')}{\zeta(1 + z_j)\zeta(1 + z_j')} \frac{R^{z_j + z_j'}}{z_j^2 z_j'^2} dz_j dz_j' \\
&+ O(e^{-\delta\sqrt{\log R}}) \\
=&I(G(z_1, \ldots, z_m, 0, z_1', \ldots, z_m', 0), m) \log R + I(H, m) + O(e^{-\delta\sqrt{\log R}})
\end{aligned}$$

where $\delta > 0$ and $H : \mathcal{D}_\sigma^m \to \mathbb{C}$ is the function

$$\begin{aligned}
H(z_1, \ldots, z_m, z_1', \ldots, z_m') :=& \frac{\partial G}{\partial z_{m+1}'}(z_1, \ldots, z_m, 0, z_1', \ldots, z_m', 0) \\
&+ \frac{1}{2\pi i} \int_{\Gamma_0} G(z_1, \ldots, z_m, z_{m+1}, z_1', \ldots, z_m', -z_{m+1}) \frac{dz_{m+1}}{\zeta(1 + z_{m+1})\zeta(1 - z_{m+1}) z_{m+1}^4}.
\end{aligned}$$

The error term $O(e^{-\delta\sqrt{\log R}})$ which we claim here arises by applying (10.10) and several applications of (11.3).

Now both of the functions $G(z_1, \ldots, z_m, 0, z_1', \ldots, z_m', 0)$ and $H(z_1, \ldots, z_m, z_1', \ldots, z_m')$ are analytic on $\mathcal{D}_\sigma^m$ and (appealing to (11.5)) we have $\|H\|_{C^j(\mathcal{D}_\sigma^m)} = O_m(\|G\|_{C^{j+1}(\mathcal{D}_\sigma^{m+1})})$ for $0 \leqslant j \leqslant m$. Using the inductive hypothesis, we therefore obtain $I(G, m+1) =$

$$\begin{aligned}
& G(0,\ldots,0)(\log R)^{m+1} + \sum_{j=1}^{m} O_m(\|G(\cdot,0,\cdot,0)\|_{C^j(\mathcal{D}_\sigma^m)}(\log R)^{m+1-j}) \\
& + H(0,\ldots,0)(\log R)^m + \sum_{j=1}^{m} O_m(\|H\|_{C^j(\mathcal{D}_\sigma^m)}(\log R)^{m-j}) + O(e^{-\delta\sqrt{\log R}}) \\
= \; & G(0,\ldots,0)(\log R)^{m+1} + \sum_{j=1}^{m} O_m(\|G\|_{C^j(\mathcal{D}_\sigma^{m+1})}(\log R)^{m+1-j}) \\
& + H(0,\ldots,0)(\log R)^m + \sum_{j=1}^{m} O_m(\|G\|_{C^{j+1}(\mathcal{D}_\sigma^{m+1})}(\log R)^{m-j}) + O(e^{-\delta\sqrt{\log R}}) \\
= \; & G(0,\ldots,0)(\log R)^{m+1} + \sum_{j=1}^{m+1} O_m(\|G\|_{C^j(\mathcal{D}_\sigma^{m+1})}(\log R)^{m+1-j}) + O(e^{-\delta\sqrt{\log R}}),
\end{aligned}$$

which is what we wanted to prove. □

References

[1] I. Assani, *Pointwise convergence of ergodic averages along cubes*, preprint.

[2] A. Balog, *Linear equations in primes,* Mathematika **39** (1992) 367–378.

[3] ———, *Six primes and an almost prime in four linear equations,* Can. J. Math. **50** (1998), 465–486.

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

[5] J. Bourgain *A Szemerédi-type theorem for sets of positive density in $\mathbb{R}^k$,* Israel J. Math **54** (1986), no. 3, 307–316.

[6] ———, *On triples in arithmetic progression*, GAFA **9** (1999), 968–984.

[7] S. Chowla, *There exists an infinity of 3—combinations of primes in A. P.*, Proc. Lahore Philos. Soc. **6**, (1944). no. 2, 15–16.

[8] P. Erdős, P. Turán, *On some sequences of integers*, J. London Math. Soc. **11** (1936), 261–264.

[9] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.

[10] H. Furstenberg, Y. Katznelson and D. Ornstein, *The ergodic-theoretical proof of Szemerédi's theorem,* Bull. Amer. Math. Soc. **7** (1982), 527–552.

[11] H. Furstenberg, B. Weiss, *A mean ergodic theorem for* $1/N \sum_{n=1}^{N} f(T^n x) g(T^{n^2} x)$, Convergence in ergodic theory and probability (Columbus OH 1993), 193–227, Ohio State Univ. Math. Res. Inst. Publ., 5. de Gruyter, Berlin, 1996.

[12] D. Goldston and C.Y. Yıldırım *Higher correlations of divisor sums related to primes, I: Triple correlations,* Integers **3** (2003) A5, 66pp.

[13] ———, *Higher correlations of divisor sums related to primes, III: k-correlations,* preprint (available at AIM preprints)

[14] ———, *Small gaps between primes, I,* preprint.

[15] T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, GAFA **8** (1998), 529–551.

[16] ———, *A new proof of Szemerédi's theorem*, GAFA **11** (2001), 465-588.

[17] ———, *Hypergraph regularity and the multidimensional Szemerédi theorem,* preprint.

[18] B.J. Green, *Roth's theorem in the primes,* to appear in Annals of Math.

[19] ———, *A Szemerédi-type regularity lemma in abelian groups,* preprint.

[20] B.J. Green and T. Tao, *Restriction theory of Selberg's sieve, with applications,* preprint.
[21] 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
[22] D.R. Heath-Brown, *Three primes and an almost prime in arithmetic progression,* J. London Math. Soc. (2) **23** (1981), 396–414.
[23] ———, *Linear relations amongst sums of two squares,* Number theory and algebraic geometry — to Peter Swinnerton-Dyer on his 75th birthday, CUP (2003).
[24] B. Host and B. Kra, *Convergence of Conze-Lesigne averages,* Ergodic Theory and Dynamical Systems **21** (2001), no. 2, 493–509.
[25] ———, *Non-conventional ergodic averages and nilmanifolds,* to appear in Ann. Math.
[26] ———, *Convergence of polynomial ergodic averages,* to appear in Israel. Jour. Math.
[27] I. Łaba and M. Lacey, *On sets of integers not containing long arithmetic progressions,* unpublished. Available at `http://www.arxiv.org/pdf/math.CO/0108155`.
[28] A. Moran, P. Pritchard and A. Thyssen, *Twenty-two primes in arithmetic progression,* Math. Comp. **64** (1995), no. 211, 1337–1339.
[29] O. Ramaré, *On Snirel'man's constant,* Ann. Scu. Norm. Pisa **21** (1995), 645–706.
[30] O. Ramaré and I.Z. Ruzsa, *Additive properties of dense subsets of sifted sequences,* J. Th. Nombres de Bordeaux **13** (2001) 559–581.
[31] R. Rankin, *Sets of integers containing not more than a given number of terms in arithmetical progression.* Proc. Roy. Soc. Edinburgh Sect. A, **65** 1960/1961 332–344 (1960/61).
[32] K.F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 245-252.
[33] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104.
[34] ———, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299–345.
[35] ———, *Regular partitions of graphs,* in "Proc. Colloque Inter. CNRS" (J.-C. Bermond, J.-C. Fournier, M. Las Vergnas, D. Sotteau, eds.) (1978), 399–401.
[36] E.C. Titchmarsh, *The theory of the Riemann zeta function,* Oxford University Press, 2nd ed, 1986.
[37] J.G. van der Corput, *Über Summen von Primzahlen und Primzahlquadraten,* Math. Ann. **116** (1939), 1–50.
[38] P. Varnavides, *On certain sets of positive density,* J. London Math. Soc. **34** (1959) 358–360.
[39] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, preprint.
[40] ———, *A non-conventional ergodic theorem for a nilsystem*, preprint.

Pacific Institute for the Mathematical Sciences, Room 205, 1933 West Mall, University of British Columbia, Vancouver BC, Canada,
E-mail address: `bjg23@hermes.cam.ac.uk`

Department of Mathematics, University of California at Los Angeles, Los Angeles CA 90095
E-mail address: `tao@math.ucla.edu`