Pure and Applied Mathematics Quarterly
Volume 2, Number 2
(*Special Issue: In honor of
John H. Coates, Part 2 of 2*)
395—433, 2006

Obstructions to Uniformity and Arithmetic Patterns in the Primes

Terence Tao

Abstract: In this expository article, we describe the recent approach, motivated by ergodic theory, towards detecting arithmetic patterns in the primes, and in particular establishing in [26] that the primes contain arbitrarily long arithmetic progressions. One of the driving philosophies is to identify precisely what the *obstructions* could be that prevent the primes (or any other set) from behaving "randomly", and then either show that the obstructions do not actually occur, or else convert the obstructions into usable structural information on the primes.

1. Introduction

An important class of problems in additive number theory, many of which are still far from being solved, concerns the existence and distribution of affine-linear arithmetic patterns in the primes and almost primes. Some well-known examples of these problems include:

- (Twin prime conjecture) Does there exist infinitely many numbers n such that $n, n+2$ are both prime?
- (Chen's theorem) [9] There exists infinitely many numbers n such that n is prime, and $n+2$ is the product of at most two primes.
- (Sophie Germain prime conjecture) Does there exist infinitely many numbers n such that $n, 2n+1$ are both prime?
- (Goldbach conjecture) For every sufficiently large even number N, does there exist an n such that n and $N-n$ are both prime?
- (Vinogradov's theorem) [51] For every sufficiently large odd number N, there exists n, m such that n, m, and $N-n-m$ are all prime.

Received May 20, 2005.
1991 *Mathematics Subject Classification.* 11N13, 11B25, 374A5. The author is supported by a grant from the Packard Foundation.

- (Hardy-Littlewood prime tuples conjecture) [31] For any integers $a_1, \ldots, a_k$, which do not fill out all the residue classes of $\mathbb{Z}/p\mathbb{Z}$ for any prime p, there exists infinitely many n such that $n + a_1, \ldots, n + a_k$ are all prime.
- (van der Corput's theorem) [49] There exist infinitely many positive numbers a, r such that $a, a + r, a + 2r$ are all prime.
- (Green-Tao theorem) [27] For any k, there exist infinitely many positive integers a, r such that $a, a + r, \ldots, a + (k - 1)r$ are all prime.

A unifying conjecture that encompasses all of these results is the *generalized Hardy Littlewood prime tuples conjecture*, which we now discuss. As is customary in additive number theory, the most convenient way to count patterns in the primes is to introduce the *von Mangoldt function* $\Lambda : \mathbb{Z} \to \mathbb{R}^+$, defined by setting $\Lambda(n) := \log p$ whenever $n = p^j$ is a power of a prime p for some $j \geqslant 1$, and $\Lambda(n) = 0$ otherwise (in particular Λ vanishes on zero and the negative integers). This function is mostly supported on the primes, and obeys a number of useful properties; for instance, one can encode the unique factorization of the integers via the pleasant identity[1]

$$\log n = \sum_{d|n} \Lambda(d) \tag{1.1}$$

for all $n \in \mathbb{Z}^+$. Also, the prime number theorem can be phrased succinctly as

$$\mathbb{E}(\Lambda(n)|1 \leqslant n \leqslant N) = 1 + o_{N\to\infty}(1) \tag{1.2}$$

where we use $\mathbb{E}(f(n)|n \in A)$ to denote the average $\frac{1}{|A|}\sum_{n\in A} f(n)$, and $o_{N\to\infty}(1)$ denotes a quantity that goes to zero[2] as $N \to \infty$. Thus Λ is essentially normalized to have mean 1. More generally, for any modulus $q \geqslant 1$ and any integer a, we have

$$\mathbb{E}(\Lambda(n)|1 \leqslant n \leqslant N; n = a(\mathrm{mod}\, q)) = \Lambda_{\mathbb{Z}/q\mathbb{Z}}(a) + o_{N\to\infty;q}(1) \tag{1.3}$$

for all sufficiently large N, where $o_{N\to\infty;q}(1)$ is a quantity which goes to zero as $N \to \infty$ for any fixed q, and the "local von Mangoldt function" $\Lambda_{\mathbb{Z}/q\mathbb{Z}}(a)$ is defined as the function which equals $\frac{q}{\phi(q)}$ when a is coprime to q and 0 otherwise, with $\phi(q) = |(\mathbb{Z}/q\mathbb{Z})^\times|$ being the Euler totient function; this result follows by combining the prime number theorem (1.2) with Dirichlet's theorem on the distribution of primes in arithmetic progressions. One can also think of (1.3) as an assertion that the $\Lambda_{\mathbb{Z}/q\mathbb{Z}}$ is essentially the *conditional expectation* of Λ to the σ-algebra generated by the residue classes modulo q.

[1] All sums shall be over the positive integers $\mathbb{Z}^+$ unless otherwise indicated.

[2] Of course, one can make the decay rates much more quantitative, especially if one assumes strong hypotheses such as the Riemann hypothesis. However, our discussion here will be not require any quantitative control of $o(1)$ type error terms.

From the sieve of Eratosthenes, one is led to the heuristic[3]

$$\Lambda(n) \approx 1_{n>0} \prod_{p<R} \Lambda_{\mathbb{Z}/p\mathbb{Z}}(n)$$

where $1 \ll R \ll n$ is an intermediate quantity between 1 and n that we shall be deliberately vague about specifying[4]. The Chinese remainder theorem then suggests that the local factors $\Lambda_{\mathbb{Z}/p\mathbb{Z}}(n)$ in this product should behave "independently". This leads to the following conjecture:

Conjecture 1.1 (Generalized Hardy-Littlewood prime tuples conjecture)**.** *Let m, t be positive integers. For each $1 \leqslant i \leqslant m$, let $\psi_i : \mathbb{Z}^t \to \mathbb{Z}$ be an affine-linear form $\psi_i(x_1, \ldots, x_t) = \sum_{j=1}^t L_{ij} x_j + b_i$ for some integers L_{ij}, b_i, such that the forms ψ_i are all non-constant, and no two are rational multiples of each other. Let N be a large integer, and assume that $b_i = O(N)$ for all $1 \leqslant i \leqslant m$. Then we have*

$$\mathbb{E}(\prod_{i=1}^m \Lambda(\psi_i(x)) | x \in \{1, \ldots, N\}^t) = \alpha_\infty(N) \prod_p \alpha_p + o_{N\to\infty;m,t,L}(1) \tag{1.4}$$

where $L := (L_{ij})_{1\leqslant i\leqslant m, 1\leqslant j\leqslant t}$, $\alpha_\infty(N)$ is the local density at infinity

$$\alpha_\infty(N) := \mathbb{E}(\prod_{i=1}^m 1_{\psi_i(x)>0} | x \in \{1, \ldots, N\}^t)$$

and α_p is the local density at each prime p

$$\alpha_p := \mathbb{E}(\prod_{i=1}^m \Lambda_{\mathbb{Z}/p\mathbb{Z}}(\psi_i(x)) | x \in (\mathbb{Z}/p\mathbb{Z})^t). \tag{1.5}$$

Remark 1.2. *The density $\alpha_\infty(N)$ simply reflects the fact that the primes are positive; this factor is just 1 if all the L_{ij} and b_i are positive. Note we allow the b_i to depend on N, and the error term $o_{N\to\infty;m,t,L}(1)$ is presumed to be independent of the b_i; this is necessary in order for this conjecture to encompass such conjectures as Goldbach's conjecture. One can show that $\alpha_p = 1 + O_{m,t,L}(1/p^2)$ and hence the product $\prod_p \alpha_p$ (also known as the singular series) is always convergent. The conjecture is an assertion that the von Mangoldt function $\Lambda(n)$ behaves "randomly", subject to the structural constraints that it must resemble $1_{n>0}$ "locally at infinity" (e.g. in the sense of* (1.2)*), and must resemble $\Lambda_{\mathbb{Z}/p\mathbb{Z}}$ locally at each prime p (e.g. in the sense of* (1.3)*). One can also extend the conjecture to polynomial ψ_i; this is known as the Bateman-Horn conjecture* [4].

---

[3] If P is a statement, we use 1_P to denote the quantity 1 if P is true and 0 if P is false. Similarly if A is a set, we write $1_A(n)$ for $1_{n\in A}$.

[4] The original sieve of Eratosthenes requires $R = \sqrt{n}$, but this is problematic for a number of reasons, for instance Mertens' theorem shows that a further correction term is required. In practice we shall think of R as being somewhat smaller, for instance a small power of n.

This conjecture, if true, would imply all the conjectures and theorems stated earlier. For instance, it predicts

$$\mathbb{E}(\Lambda(n)\Lambda(n+2)|1 \leqslant n \leqslant N) = \prod_p \alpha_p + o_{N\to\infty}(1) \tag{1.6}$$

where $\alpha_2 := 2$ and $\alpha_p := 1 - \frac{1}{(p-1)^2}$ for all odd primes p. The *twin prime constant*

$$\Pi_2 := \prod_{p \text{ odd}} \alpha_p = 0.66016\ldots > 0$$

is positive, and (1.6) can then easily be seen to imply the twin prime conjecture. Similarly for the other conjectures and theorems stated earlier.

Of course, this conjecture is still hopelessly out of reach in the general case. However, several partial results are known. The bounds (1.2), (1.3) can already handle the $m = 1$ case of this conjecture and more generally they can handle any "non-degenerate" case with $m \leqslant t$. The Hardy-Littlewood circle method, which we discuss below, is roughly speaking able to handle any non-degenerate case with $3 \leqslant m \leqslant t+1$ (thus encompassing Vinogradov's theorem and van der Corput's theorem), as well as a few additional cases[5], but does not seem able to handle the general case. The conjecture is also known to be true if one averages over a suitable subset of the parameters L_{ij}, b_i; see [2]. In the general case, the technique of upper bound sieves in sieve theory can usually yield an upper bound of $C_{m,t}\alpha_\infty(N)\prod_p \alpha_p + o_{N\to\infty;m,t,L}(1)$ for (1.4) for some explicit $C_{m,t}$ (which usually has to be at least 2, thanks to the notorious *parity problem*); see also Section 2 below. Closely related to this are the results of Goldston and Yıldırım, which show that asymptotic formulae such as (1.4) can be recovered (but again with a loss of $C_{m,t}$ on the right-hand side) if one replaces Λ with a slightly larger function ν which is localized to *almost primes* (numbers with no small divisors) rather than primes themselves. The ergodic theory-style transference arguments used in [26], [27] can conversely give *lower bounds* of $c_{m,t}\alpha_\infty(N)\prod_p \alpha_p + o_{N\to\infty;m,t,L}(1)$ for some small $0 < c_{m,t} < 1$, but only for linear forms which are *homogeneous* (no constant term b_i) and which are *translation invariant*, in the sense that they take the form

$$\psi_i(x_1,\ldots,x_t) = x_1 + \tilde{\psi}_i(x_2,\ldots,x_t).$$

In this special case, which covers the case of arithmetic progressions in the primes, there is also some hope of recovering the full asymptotic (1.4); we discuss this below.

[5]For instance, by a clever iteration of the circle method, it was established in [3] that for any k there exist infinitely many k-tuples of distinct primes $p_1,\ldots,p_k$, such that all the midpoints $(p_i+p_j)/2$ are also prime.

In this expository article we shall discuss these techniques, starting with the prime number theorem (but re-interpreted in the perspective of Goldston-Yıldırım majorants), the classical circle method (but re-interpreted in a more "ergodic" perspective), and then turning to long arithmetic progressions in the primes; we also discuss some further recent progress in the case of progressions of length four. In particular we hope to communicate some of the main philosophical ideas underlying the approach in [26], namely:

- Viewing the primes as a dense subset, not of the integers, but instead of a "pseudorandom" set of almost primes (or more precisely, a pseudorandom majorant ν for the von Mangoldt function Λ);
- Attacking problems such as (1.4) by locating the "obstructions to uniformity" which could potentially prevent (1.4) from being true;
- Using tools such as conditional expectation to handle these obstructions to uniformity, or tools such as the circle method to show that they do not occur at all.

This is by no means intended to be an exhaustive survey; see for instance [36] for a more in-depth discussion of many of these issues. We will also not give detailed proofs for most of the assertions in this survey, referring the reader instead to the original papers.

2. The prime number theorem and enveloping sieves

We begin with the classical prime number theorem (1.2). The story of this theorem, and its connection to the zeroes of the Riemann zeta function $\zeta(s) := \sum_n \frac{1}{n^s}$, is of course very well known, but we revisit it to make two points. Firstly, as was observed by Chebyshev, one can obtain upper and lower bounds for (1.2) by elementary means (utilizing the pole of ζ at $s = 1$, but requiring no further knowledge about zeroes or analytic continuation) that are only off by an absolute constant. Secondly, by a refinement of this elementary method one can in fact get asymptotics with $o(1)$ error terms, but at the cost of smoothing out the von Mangoldt function Λ and replacing it by a slightly larger variant, namely an *enveloping sieve* ν for Λ. In fact, it turns out even such results as those in [26], establishing arbitrarily long arithmetic progressions in the primes, can in fact be proven without knowledge of the full prime number theorem (and thus without knowing any non-trivial zero-free region for ζ, or for any other L-function), instead using only[6] these elementary techniques, albeit in conjunction with a deep and powerful theorem of Szemerédi.

[6]Of course, the larger the zero-free region is known for the zeta function, the better the bounds one will obtain on the number of progressions, but if one just wants to obtain the qualitative result that there are infinitely many progressions, no zero-free region beyond the trivial one used here is required.

We begin with the argument of Chebyshev (rephrased here in modern language). If s is any complex number with $\Re(s) > 1$, we may multiply (1.1) by $\frac{1}{n^s}$ and sum in n, and make the change of variables $n = dm$, to obtain

$$\sum_n \frac{\log n}{n^s} = \sum_d \frac{\Lambda(d)}{d^s} \sum_m \frac{1}{m^s} = \sum_d \frac{\Lambda(d)}{d^s} \zeta(s).$$

The right-hand side is $-\zeta'(s)$, and hence we have the standard formula

$$\sum_d \frac{\Lambda(d)}{d^s} = -\frac{\zeta'(s)}{\zeta(s)}. \tag{2.1}$$

From summation by parts we obtain the bounds

$$\zeta(s) = \frac{1}{s-1} + O(1); \quad \zeta'(s) = \frac{1}{(s-1)^2} + O(1) \tag{2.2}$$

when $\Re(s) > 1$ and s is close to 1. In particular, we have a very small zero free region for ζ near $s = 1$. We conclude that

$$\sum_d \frac{\Lambda(d)}{d^s} = \frac{1}{s-1} + O(1) \tag{2.3}$$

whenever $\Re(s) > 1$ and s is close to 1. Subtacting off $\zeta(s)$, and then setting $s = 1 + \frac{1}{\log N} + it$, we obtain

$$\sum_d e^{-it \log d} \frac{\Lambda(d) - 1}{d^{1+1/\log N}} = O(1)$$

for all large N and all small t. By some Fourier analysis, we can then conclude that

$$\sum_d \psi(\log d - \log N) \frac{\Lambda(d) - 1}{d^{1+1/\log N}} = O_\psi(1)$$

for all Schwartz functions ψ whose Fourier transform is supported on a sufficiently small neighbourhood of the origin. This, combined with the trivial observation that Λ is non-negative, is already enough to give the elementary bounds

$$c - o_{N\to\infty}(1) \leqslant \mathbb{E}(\Lambda(n) | 1 \leqslant n \leqslant N) \leqslant C + o_{N\to\infty}(1) \tag{2.4}$$

for some absolute constants $0 < c < 1 < C$.

The estimate (2.4) is not an asymptotic, of course, since $c \neq C$. However, we can recover good asymptotics by smoothing out the von Mangoldt function Λ slightly. We introduce the *Möbius function* $\mu : \mathbb{Z}^+ \to \{-1, 0, +1\}$, defined by $\mu(n) = (-1)^k$ when n is the product of k distinct primes for some $k \geqslant 0$, and $\mu(n) = 0$ otherwise. The significance of this function lies in the inclusion-exclusion formula

$$1_{n=1} = 1_{n>0} \sum_{d|n} \mu(d), \tag{2.5}$$

and hence from (1.1)

$$\begin{aligned}\Lambda(n) &= 1_{n>0} \sum_{m|n} \Lambda(m) 1_{n/m=1} \\ &= 1_{n>0} \sum_{dm|n} \Lambda(m)\mu(d) \\ &= 1_{n>0} \sum_{d|n} \mu(d) \log \frac{n}{d} \\ &= 1_{n>0} \log n \sum_{d|n} \mu(d)(1 - \frac{\log d}{\log n}).\end{aligned} \tag{2.6}$$

Inspired by this, let us define the truncated von Mangoldt functions $\Lambda_{\mathbb{R},\varphi} : \mathbb{Z} \to \mathbb{R}$ by

$$\Lambda_{R,\varphi}(n) := \log R \sum_{d|n} \mu(d) \varphi(\frac{\log d}{\log R}) \tag{2.7}$$

where $R > 1$ is a large parameter, and $\varphi : \mathbb{R} \to \mathbb{R}$ is a function supported on the interval $[-1, 1]$. For instance, the von Mangoldt function itself corresponds to the case when $R = n$ and $\varphi(x) := \max(1 - |x|, 0)$. The case when $R < n$ and $\varphi(x) = \max(1 - |x|, 0)$ was studied by Goldston and Yıldırim; that case is also related to the Selberg upper bound sieve[7], see [27] for further discussion. These functions are more "localized", and hence easier to analyze, than the original von Mangoldt function, in the sense that they only involve divisors d that are less than R[8].

The truncated von Mangoldt functions behave somewhat similarly to the von Mangoldt function, but are concentrated on the *almost primes* rather than the primes themselves. For instance, it is easy to see that $\Lambda_{R,\varphi}(n) = \varphi(0) \log R$ whenever n is a prime larger than R, or more generally if n is the product of primes larger than R. One can also easily establish a fairly elementary "prime number theorem" for these functions, provided that R is not quite as large as N:

[7]The choice $\varphi(x) = \max(1-|x|, 0)$ will give an optimized value of the relative density between Λ and its enveloping sieve, although we will not need such optimization in our arguments. Very recently, however, there has been work of Goldston, Motohashi, Pintz, and Yıldırım, which use precise optimization of higher-dimensional enveloping sieves in order to establish small gaps between primes, thus exploiting enveloping sieves in a rather different way than that discussed here.

[8]This can be viewed as a manifestation of the *uncertainty principle*: localizing a function in the spectral or "frequency" sense (i.e. with respect to the divisors d) must necessarily cause delocalization in physical space (i.e with respect to the variables n).

Proposition 2.1 (Prime number theorem for $\Lambda_{R,\varphi}$)**.** *If $N^\varepsilon \leqslant R \leqslant N^{1-\varepsilon}$ for some $\varepsilon > 0$, and φ is smooth with $\varphi(0) = 1$ and $\varphi'(0) = 0$, then we have*

$$\mathbb{E}(\Lambda_{R,\varphi}(n)|1 \leqslant n \leqslant N) = 1 + o_{N\to\infty;\varepsilon,\varphi}(1). \tag{2.8}$$

Proof. We can expand the left-hand side of (2.8) as

$$\log R \sum_{d\leqslant R} \mu(d)\varphi\left(\frac{\log d}{\log R}\right) \mathbb{E}(1_{d|n}|1 \leqslant n \leqslant N).$$

From the elementary estimate

$$\mathbb{E}(1_{d|n}|1 \leqslant n \leqslant N) = \frac{1}{d} + O(\frac{1}{N})$$

we can thus write the left-hand side of (2.8) as

$$\log R \sum_{d\leqslant R} \frac{\mu(d)}{d}\varphi\left(\frac{\log d}{\log R}\right) + O_\varphi(\log R \sum_{d\leqslant R} \frac{1}{N}).$$

Here the subscripting of $O()$ by φ denotes that the implied constant is allowed to depend on φ. Since φ is supported on $[-1, 1]$, we may remove the restriction $d \leqslant R$. Since we are taking $R \leqslant N^{1-\varepsilon}$, the error term here is $o_{N\to\infty;\varepsilon,\varphi}(1)$. Since we also take $R > N^\varepsilon$, it thus suffices to show that

$$\log R \sum_{d} \frac{\mu(d)}{d}\varphi\left(\frac{\log d}{\log R}\right) = 1 + o_{R\to\infty;\varphi}(1). \tag{2.9}$$

To proceed further we need to split $\varphi(\frac{\log d}{\log R})$ into expressions which are multiplicative in d. This is easiest to establish by Fourier expansion[9]. Since the function $e^x\varphi(x)$ is smooth and compactly supported, we have

$$e^x\varphi(x) = \int_{-\infty}^{\infty} \psi(t)e^{-ixt}\, dt \tag{2.10}$$

for some rapidly decreasing function[10] ψ. We truncate this at $|t| = \log^{1/2} R$ (for instance) to obtain

$$e^x\varphi(x) = \int_{|t|\leqslant\log^{1/2} R} \psi(t)e^{-ixt}\, dt + O_{A,\varphi}(\log^{-A} R)$$

for any $A > 0$. In particular, we have

$$\varphi(\frac{\log d}{\log R}) = \int_{|t|\leqslant\log^{1/2} R} \frac{\psi(t)\, dt}{d^{(1+it)/\log R}} + O_{A,\varphi}(d^{-1/\log R}\log^{-A} R) \tag{2.11}$$

[9]One could also use contour integration methods here instead of Fourier methods; the two approaches are essentially equivalent.

[10]In other words, $\psi(x) = O_{A,\psi}((1 + |x|)^{-A})$ for all $A > 0$ and $x \in \mathbb{R}$.

and hence the left-hand side of (2.9) can be written as

$$\log R \int_{|t|\leqslant \log^{1/2} R} [\sum_d \frac{\mu(d)}{d^{1+(1+it)/\log R}}]\psi(t)\ dt + O_{A,\varphi}(\log R \sum_d \frac{1}{d} d^{-1/\log R} \log^{-A} R).$$

By taking $A = 3$ (say), we see that the error term is $o_{R\to\infty;\varphi}(1)$ and so can be discarded. As for the main term, we first repeat the derivation of (2.1), using (2.5) instead of (1.1), to conclude

$$\sum_d \frac{\mu(d)}{d^s} = \frac{1}{\zeta(s)};$$

by (2.2) we thus have

$$\sum_d \frac{\mu(d)}{d^s} = s - 1 + O(|s-1|^2)$$

when $\Re(s) > 1$ and s is sufficiently close to 1. Setting $s = 1 + \frac{1+it}{\log R}$ for some $|t| \leqslant \log^{1/2} R$ we obtain (for N and hence R sufficiently large)

$$\sum_d \frac{\mu(d)}{d^{1+(1+it)/\log R}} = \frac{1+it}{\log R} + O((1+|t|^2)\log^{-2} R).$$

Inserting this bound into the previous computations, and using the rapid decay of ψ, we can thus write the left-hand side of (2.9) as

$$\int_{|t|\leqslant \log^{1/2} R} (1+it)\psi(t)\ dt + o_{R\to\infty;\varphi}(1).$$

Using the rapid decay of ψ again, we can write this as

$$\int_{-\infty}^{\infty} (1+it)\psi(t)\ dt + o_{R\to\infty;\varphi}(1)$$

which we rewrite in turn as

$$(1 - \frac{d}{dx}) \int_{-\infty}^{\infty} e^{-ixt}\psi(t)\ dt|_{x=0} + o_{R\to\infty;\varphi}(1).$$

Applying (2.10), this becomes

$$\varphi(0) - \varphi'(0) + o_{R\to\infty;\varphi}(1),$$

and the claim follows from the hypotheses on φ. □

One notable drawback of the truncated von Mangoldt functions $\Lambda_{R,\varphi}$ is that, unlike Λ, it is perfectly possible for $\Lambda_{R,\varphi}(n)$ to be negative. This however can be rectified by replacing $\Lambda_{R,\varphi}$ with the variant

$$\nu = \nu_{R,\varphi} := \frac{1}{\log R} \Lambda_{R,\varphi}^2. \tag{2.12}$$

This function is still large on almost primes, indeed $\nu(n) = \Lambda_{R,\varphi}(n) = \varphi(0)^2 \log R$ whenever n is a prime greater than R, or a product of primes greater than R. In particular, if $\log R \sim \log N$ and $\varphi(0) \sim 1$ then we have the pointwise bound

$$0 \leqslant \Lambda(n) \leqslant C\nu(n) \tag{2.13}$$

for all $1 \leqslant n \leqslant N$, where $C := \frac{1}{|\varphi(0)|^2} \frac{\log N}{\log R}$. As observed[11] by Goldston and Yıldırım, we can also modify the above argument to obtain a prime number theorem for ν, although at the cost of reducing the size of R:

Proposition 2.2 (Prime number theorem for ν)**.** *If $N^\varepsilon \leqslant R \leqslant N^{1/2-\varepsilon}$ for some $\varepsilon > 0$, and φ is smooth with $\int_0^1 |\varphi'(x)|^2\, dx = 1$, then we have*

$$\mathbb{E}(\nu(n)|1 \leqslant n \leqslant N) = 1 + o_{N\to\infty;\varepsilon,\varphi}(1). \tag{2.14}$$

Proof. We repeat the proof of Proposition 2.1. We can expand the left-hand side of (2.14) as

$$\log R \sum_{d,d' \leqslant R} \mu(d)\mu(d')\varphi(\frac{\log d}{\log R})\varphi(\frac{\log d'}{\log R})\mathbb{E}(1_{d,d'|n}|1 \leqslant n \leqslant N).$$

From the Chinese remainder theorem we have

$$\mathbb{E}(1_{d,d'|n}|1 \leqslant n \leqslant N) = \frac{1}{[d,d']} + O(\frac{1}{N})$$

where $[d,d']$ is the least common multiple of d and d'. The hypothesis $R \leqslant N^{1/2-\varepsilon}$ allows us to discard the error term as before, leaving us with the task of establishing

$$\log R \sum_{d,d'} \frac{\mu(d)\mu(d')}{[d,d']}\varphi(\frac{\log d}{\log R})\varphi(\frac{\log d'}{\log R}) = 1 + o_{R\to\infty;\varphi}(1).$$

From (2.11) we have

$$\varphi(\tfrac{\log d}{\log R})\varphi(\tfrac{\log d'}{\log R}) = \int_{|t|,|t'| \leqslant \log^{1/2} R} \frac{\psi(t)\psi(t')\ dtdt'}{d^{(1+it)/\log R}(d')^{(1+it')/\log R}} + O_{A,\varphi}((dd')^{-1/\log R}\log^{-A} R).$$

Let us first dispose of the error term. This contribution can be bounded by

$$O_{A,\varphi}(\log R)^{1-A} \sum_{d,d'} \frac{1}{[d,d'](dd')^{-1/\log R}}.$$

[11] Strictly speaking, these authors only consider the case $\varphi(x) = \max(1-|x|,0)$, but the argument extends to general φ without difficulty.

Using unique factorization $\mathbb{Z}^+ = \prod_p p^{\mathbb{Z}^+}$, and the multiplicative nature of the summand, the sum can be expanded as an Euler product

$$\sum_{d,d'} \frac{1}{[d,d'](dd')^{-1/\log R}} = \prod_p \sum_{d,d' \in p^{\mathbb{Z}^+}} \frac{1}{[d,d'](dd')^{-1/\log R}}.$$

One can compute

$$\sum_{d,d' \in p^{\mathbb{Z}^+}} \frac{1}{[d,d'](dd')^{-1/\log R}} = 1 + O(1/p^{1+1/\log R}) \leqslant (1 - 1/p^{1+1/\log R})^{-O(1)}.$$

On the other hand, from (2.2) and the Euler product

$$\zeta(s) = \prod_p \sum_{n \in p^{\mathbb{Z}^+}} \frac{1}{n^s} = \prod_p (1 - 1/p^s)^{-1}$$

we have

$$\prod_p (1 - 1/p^s)^{-1} = \frac{1}{s-1} + O(1) \tag{2.15}$$

for $\Re(s) > 1$ and s close to 1. From this we see that the total contribution of the error term is $O_{A,\varphi}(\log^{O(1)-A} R)$, which is acceptable since A can be chosen to be large.

It remains to control the main term, which is

$$\log R \int_{|t|,|t'| \leqslant \log^{1/2} R} \left(\sum_{d,d'} \frac{\mu(d)\mu(d')}{[d,d']d^{(1+it)/\log R}(d')^{(1+it')/\log R}} \right) \psi(t)\psi(t')\, dtdt'. \tag{2.16}$$

The expression inside the parentheses can be expanded as an Euler product

$$\prod_p \sum_{d,d' \in p^{\mathbb{Z}}+} \frac{\mu(d)\mu(d')}{[d,d']d^{(1+it)/\log R}(d')^{(1+it')/\log R}}$$

which one can compute as

$$\prod_p (1 - \frac{1}{p^{1+(1+it)/\log R}} - \frac{1}{p^{1+(1+it')/\log R}} + \frac{1}{p^{1+(2+it+it')/\log R}}).$$

After some Taylor expansion, we can write this as

$$\prod_p \frac{(1 - \frac{1}{p^{1+(1+it)/\log R}})(1 - \frac{1}{p^{1+(1+it')/\log R}})}{1 - \frac{1}{p^{1+(2+it+it')/\log R}}} (1 + O(\frac{(1+|t|+|t'|)\log p}{p^2 \log R})). \tag{2.17}$$

Since $\sum_p \frac{\log p}{p^2}$ is convergent, and $|t|, |t'| \leqslant \log^{1/2} R = o_{R\to\infty}(\log R)$, we have

$$\prod_p (1 + O(\frac{(1+|t|+|t'|)\log p}{p^2 \log R})) = 1 + o_{R\to\infty}(1)$$

Applying (2.15), we can thus write (2.17) as

$$(1 + o_{R\to\infty}(1)) \log^{-1} R \frac{(1+it)(1+it')}{2+it+it'}.$$

The contribution of the error term to (2.16) is $o_{R\to\infty,\varphi}(1)$, thanks to the rapid decrease of ψ. Hence we are left with the expression

$$\int_{|t|,|t'|\leqslant \log^{1/2} R} \frac{(1+it)(1+it')}{2+it+it'} \psi(t)\psi(t')\ dtdt'$$

and by using the rapid decay of ψ again, we see that we will be done as soon as we establish the identity

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{(1+it)(1+it')}{2+it+it'} \psi(t)\psi(t')\ dtdt' = 1.$$

Since

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

the left-hand side can be written as

$$\int_0^\infty (\int_{-\infty}^\infty \psi(t)(1+it)e^{-(1+it)x}\ dx)^2\ dt.$$

But by dividing (2.10) by e^x and then differentiating in x, we obtain

$$\varphi'(x) = -\int_{-\infty}^\infty \psi(t)(1+it)e^{-ixt}\ dt$$

and the claim follows. □

It turns out that the above elementary argument is quite flexible, and can also give more sophisticated estimates for ν, similar to (1.4). Indeed we have

Theorem 2.3 (Generalized Hardy-Littlewood prime tuples conjecture for ν)**.** *Let m, t be positive integers. For each $1 \leqslant i \leqslant m$, let $\psi_i : \mathbb{Z}^t \to \mathbb{Z}$ be an affine-linear form $\psi_i(x_1,\ldots,x_t) = \sum_{j=1}^t L_{ij}x_j + b_i$ for some integers L_{ij}, b_i, such that the forms ψ_i are all non-constant, and no two are rational multiples of each other. Let N be a large integer, and assume that $b_i = O(N)$ for all $1 \leqslant i \leqslant m$. If $N^\varepsilon \leqslant R \leqslant N^{1/2m-\varepsilon}$ for some $\varepsilon > 0$, and φ is smooth with $\int_0^1 |\varphi'(x)|^2\ dx = 1$, then we have*

$$\mathbb{E}(\prod_{i=1}^m \nu(\psi_i(x))|x \in \{1,\ldots,N\}^t) = \prod_p \alpha_p + o_{N\to\infty;m,t,L,\varepsilon,\varphi}(1) \tag{2.18}$$

where α_p was defined in (1.5).

We will not prove this result here, but remark that the proof is a routine extension of that used to prove Proposition 2.2, and very similar results were proven in [17], [18], [19], [26], [45]. One can also obtain moment bounds for ν

in terms of various multilinear integrals involving ψ; see [17], [18], [19] for some computations of this sort. The density at infinity, α_∞, is missing, because ν extends to the negative integers as well as the positive ones. Note that as the order m of the correlation increases, the range of available R decreases, so if we set R equal to a fixed power of N, we only obtain correlations to finitely high order.

In the language of [37], [38], the function $C\nu$ appearing in (2.13) is an *enveloping sieve* for the von Mangoldt function Λ. Results such as Theorem 2.3 establish correlation estimates for this sieve, which in turn automatically imply *upper* bounds for expressions such as (1.4) which are off by a constant $C_{m,t} > 1$; thus the enveloping sieve can be used as an upper bound sieve, though it has many other uses also, thanks in large part to correlation estimates such as[12] Theorem 2.3. More advanced methods in sieve theory can of course be used to reduce this loss $C_{m,t}$, although the parity problem prevents one from removing this constant entirely by sieve-theoretic methods.

We have asserted earlier that ν is concentrated on the almost primes, which are coprime to all numbers less than R. Let us provide some further evidence of this claim. From (2.10) we have

$$\varphi(\frac{\log d}{\log R}) = \int_{-\infty}^{\infty} \psi(t) d^{-(1+it)/\log R}\ dt$$

and hence by (2.7)

$$\Lambda_{R,\varphi}(n) = \log R \int_{-\infty}^{\infty} \psi(t) \sum_{d|n} \mu(d) d^{-(1+it)/\log R}\ dt.$$

We can factorize the sum as an Euler product

$$\sum_{d|n} \mu(d) d^{-(1+it)/\log R}\ dt = \sum_{p|n} (1 - p^{-(1+it)/\log R})$$

and conclude

$$\Lambda_{R,\varphi}(n) = \log R \int_{-\infty}^{\infty} \psi(t) \prod_{p|n} (1 - p^{-(1+it)/\log R})\ dt$$

and similarly by (2.12)

$$\nu(n) = \log R \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \psi(t)\psi(t') \prod_{p|n} (1 - p^{-(1+it)/\log R})(1 - p^{-(1+it')/\log R})\ dtdt'.$$

[12] By modifying the enveloping sieve slightly, one can also get some useful estimates on the Fourier coefficients of ν, see [27]. Of course, similar estimates are also known for the Fourier coefficients of Λ itself, though the estimates for ν are simpler and do not require the theory of Siegel zeroes. In particular, the estimates are effective without requiring strong hypotheses such as GRH.

Since $\psi(t)$ is rapidly decreasing, the integral effectively localizes t to be close to 1. The factor $(1 - p^{-(1+it)/\log R})$ is then close to 0 when $p \ll R$ and oscillates around 1 when $p \gg R$. Thus we expect $\Lambda_{R,\varphi}(n)$ and $\nu(n)$ to be small when n has one or more prime factors $\ll R$, and these quantities should be close to $\log R$ when n is a product of primes $\gg R$, though in some exceptional cases (when the phases of $p^{-(1+it)/\log R}$ align in an unfavourable way) one may expect $\Lambda_{R,\varphi}(n)$ to be somewhat larger than this[13]. Thus we have the rough heuristics

$$\Lambda(n) \approx (\log N)1_P; \quad \nu(n) \approx (\log R)1_{AP} \tag{2.19}$$

for $n \sim N$, where P denotes the primes up to N and AP denotes the almost primes at level R up to N (i.e. the products of primes larger than R). Observe that $\Lambda(n)$ and $\nu(n)$ both have average $1 + o_{N\to\infty}(1)$, which thus suggests that P has density about $\frac{\log R}{\log N}$ inside AP; one can obtain more precise estimates here using Buchstab's formula. On the other hand, Theorem 2.3 combined with the heuristic (2.19) suggests that the set AP is very nicely distributed if $R = N^\varepsilon$ for some suitably small ε. Thus, in summary, the primes P form a set of positive density ($\approx \varepsilon$) inside the almost primes at level $R = N^\varepsilon$, and the latter set has a well-controlled distribution. This turns out to be a very useful perspective for a number of problems, as it bypasses the difficulty that the primes have only a density of $\frac{1}{\log N} = o_{N\to\infty}(1)$ with respect to the integers $\{1, \dots, N\}$. Thus the almost primes AP (or more precisely the enveloping sieve ν) forms a better majorant for the primes P (or more precisely the von Mangoldt function Λ) than the integers (or the constant 1).

3. The W-trick

As we have seen, any correlation estimate involving Λ or ν will involve a number of local densities α_p; these densities ultimately arise from the fact that the projections $\Lambda_{\mathbb{Z}/p\mathbb{Z}}$ of Λ to the residue classes modulo p are not constant. Note that due to the rapid convergence of the product $\prod_p \alpha_p$, it is only the small divisors p for which this non-uniformity is significant. However, if one does not care much about the exact order of decay in the $o(1)$ errors, then there is a cheap trick, which we call the "W-trick", available to essentially eliminate the role of these local factors, so that one only has to deal with functions which are uniform with respect to small divisors.

This trick works as follows. We introduce a new parameter $1 \ll w \ll N$; this will eventually be set to a very slowly growing function of N, such as $\log\log N$, although for the purposes of getting qualitative $o(1)$ bounds it is not particularly

[13] On the other hand, (2.7) shows that $\Lambda_{R,\varphi}(n)$ can be crudely bounded by $O_\varphi(\tau(n)\log R)$, where $\tau(n) = \sum_{d|n} 1$ is the divisor function. As is well known, the divisor function has size $O(\log n)$ on the average, though it can get significantly larger than this for very smooth n. However, it is always $O_\varepsilon(n^\varepsilon)$ for any $\varepsilon > 0$, and hence $\Lambda_{R,\varphi}$ and ν also have this type of bound.

important what w is. We let $W := \prod_{p<w} p$ be the product of all the primes less than w. The prime numbers larger than w will then be distributed in the residue classes $\{Wn+b : n \in \mathbb{Z}\}$, where b is one of the $\phi(W)$ numbers in $\{1, \dots, W\}$ which are coprime to W. For each of these numbers b, we introduce the renormalized von Mangoldt function

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

and similarly the renormalized truncated von Mangoldt functions

$$\Lambda_{R,\varphi,b(\mathrm{mod}\, W)}(n) := \frac{W}{\phi(W)}\Lambda_{R,\varphi}(Wn+b)$$

and the renormalized enveloping sieve

$$\nu_{b(\mathrm{mod}\, W)}(n) := \frac{W}{\phi(W)}\nu(Wn+b).$$

Then the functions $\Lambda_{b(\mathrm{mod}\, W)}(n)$ behave like Λ except that the projections modulo q are now extremely close to 1 for small q. Indeed from (1.3) and the Chinese remainder theorem, one easily verifies

$$\mathbb{E}(\Lambda_{b(\mathrm{mod}\, W)}(n)|1 \leqslant n \leqslant N; n = a(\mathrm{mod}\, q)) = 1 + o_{N\to\infty;w}(1)$$

for $1 \leqslant q \leqslant w$. The analogue of Conjecture 1.1 is then the assertion that

$$\mathbb{E}(\prod_{i=1}^{m} \Lambda_{b_i}(\psi_i(x))|x \in \{1, \dots, N\}^t) = \alpha_\infty(N) \prod_{p>w} \alpha_p + o_{N\to\infty;m,t,L,w}(1) \tag{3.1}$$

whenever $b_1, \dots, b_m \in \{1, \dots, W\}$ are coprime to W; thus the local factors corresponding to primes less than or equal to w in (1.4) are eliminated, at the cost of letting the $o(1)$ term depend on w. Actually it is not hard to see that (1.4) is in fact equivalent to (3.1). In many cases, the remaining local factor $\prod_{p>w} \alpha_p$ is in fact $1 + o_{w\to\infty;m,t,L}(1)$; for instance, this is the case if no two of the linear parts $(L_{ij})_{1\leqslant j\leqslant t}$ of the affine forms $\psi_1, \dots, \psi_m$ are not rational multiples of each other. However, there are some important cases where the remaining local factors are significant. For instance, for $1 \leqslant a \leqslant N$, the prime tuples conjecture predicts

$$\mathbb{E}(\Lambda_b(x)\Lambda_b(x+a)|1 \leqslant x \leqslant N) = \prod_{p>w:p|a} (1+\frac{1}{p})(1 + o_{w\to\infty}(1) + o_{N\to\infty;w}(1)).$$

The expression $\tau(a) := \prod_{p>w:p|a}(1+\frac{1}{p})$ is small for most a, for instance one can establish the moment estimates

$$\mathbb{E}(\tau^q(a)|1 \leqslant a \leqslant N) = O_q(1) \tag{3.2}$$

for all $1 \leqslant q < \infty$ (indeed one can refine the right-hand side to $1 + o_{w\to\infty;q}(1)$). However it is not bounded, as can be seen by taking n to be the product of a large number of primes, each of which is slightly larger than w. Nevertheless it is

a good heuristic to view $\prod_{p>w} \alpha_p$ as being close to 1 for "most" choices of forms ψ_i.

Similar considerations apply to the enveloping sieve ν. For instance, one can establish that

$$\mathbb{E}(\prod_{i=1}^{m} \nu_{b_i}(\psi_i(x))|x \in \{1,\dots,N\}^t) = 1 + o_{N\to\infty;m,t,L,\varepsilon,\varphi,w}(1) \tag{3.3}$$

whenever no two linear parts of the affine forms $\psi_1,\dots,\psi_m$; this is essentially[14] the *linear forms condition* verified in [26, Proposition 9.8]. Similarly, one can show[15]

$$\mathbb{E}(\prod_{i=1}^{m} \nu_{b_i}(x+a_i)|x \in \{1,\dots,N\}^t) \leqslant \sum_{1\leqslant i<j\leqslant m} \tilde{\tau}(a_i - a_j) \tag{3.4}$$

where $\tilde{\tau} : \mathbb{Z} \to \mathbb{R}^+$ is a slight variant of τ which is even and obeys the moment conditions (3.2); this is essentially the *correlation condition* verified in [26, Proposition 9.10]. Morally, one should think of the right-hand side of (3.4) as being bounded, with only a few exceptions such as when $a_i - a_j$ is zero or very smooth (contains a large number of prime factors larger than w).

The linear forms condition (3.3) is an assertion that the ν_b are distributed *pseudorandomly* throughout $\{1,\dots,N\}$; more informally, the almost primes AP when restricted to a coset $\{Wn + b : n \in \mathbb{Z}\}$ with b coprime to W, behave pseudorandomly inside each such coset. This is consistent with the heuristics used to support the Hardy-Littlewood prime tuples conjecture, such as Cramer's probabilistic model for the primes. In this context, a useful probabilistic model for $\nu_b(n)$ would be a function which equalled $\frac{W}{\phi(W)} \log R$ with probability $\frac{W}{\phi(W)\log R}$ independently for each n, and equalled 0 otherwise. The prime tuples conjecture then asserts that the Λ_b also behave in a similarly pseudorandom manner (but with $\log R$ essentially replaced by $\log N$).

The linear forms condition (3.3) shows that the correlations of ν_b are very close to the correlations of the constant function 1, thus ν_b is close to 1 in a "weak" sense. One of the philosophies underlying the work in [26] is a *transference principle* which asserts, informally, that many results which are true for functions bounded by constant function 1, are likely to extend to functions bounded by pseudorandom functions such as ν_b, or variants such as $\nu_b + 1$.

Any counting problem concerning the von Mangoldt function Λ can of course be subdivided into a counting problem involving the Λ_b. For instance, suppose

[14] The conditions verified in [26] actually refer to a version of ν_b adapted to $\mathbb{Z}/N\mathbb{Z}$ rather than $\{1,\dots,N\}$, but the distinction between the two is rather minor.

[15] The diagonal cases $a_i = a_j$ can be treated using the crude bound $\nu(n) = O_\varepsilon(N^\varepsilon)$ for any $\varepsilon > 0$ and $n = O(N)$.

one wanted to establish a bound such as

$$\mathbb{E}(\Lambda(a)\dots\Lambda(a+(k-1)r)|1\leqslant a,r\leqslant N)\geqslant c_k - o_{N\to\infty;k}(1)$$

for all $k\geqslant 1$ and $N\geqslant 1$, and $c_k>0$; this bound is in fact obtained in [26], and implies that the primes contain arbitrarily long arithmetic progressions. In order to achieve this bound, it suffices to show that for all w there exist $b\in\{1,\dots,W\}$ coprime to W such that

$$\mathbb{E}(\Lambda_b(a)\dots\Lambda_b(a+(k-1)r)|1\leqslant a,r\leqslant N)\geqslant c'_k - o_{w\to\infty;k}(1) - o_{N\to\infty;k,w}(1) \tag{3.5}$$

for some other $c'_k>0$. Indeed,if such a bound were true, it would imply that

$$\mathbb{E}(\Lambda_b(a)\dots\Lambda_b(a+(k-1)r)|1\leqslant a,r\leqslant N)\geqslant c'_k/2$$

(say) whenever w was sufficiently large depending on k, and N was sufficiently large depending on w and k. But since Λ_b is a renormalized component of Λ using the affine-linear transformation $n\mapsto Wn+b$ (which preserves arithmetic progressions), we then observe that

$$\mathbb{E}(\Lambda(a)\dots\Lambda(a+(k-1)r)|1\leqslant a,r\leqslant N)\geqslant c_{k,w}$$

for some $c_{k,w}>0$. Fixing w to be a suitably large constant depending only on k, we obtain the claim.

This reduction from Λ to Λ_b is used in [26]. Indeed, (3.5) is established for all $1\leqslant b<W$ which are coprime to W. In the proof, the only facts needed are the bounds $0\leqslant\Lambda_b\leqslant C\nu_b$ (which is inherited from (2.13)) and $\mathbb{E}(\Lambda_b(n)|1\leqslant n\leqslant N)>c-o_{N\to\infty;w}(1)$ (which comes from (1.3)). In fact, since we only need to establish (3.5) for a *single* b, it is possible to avoid using Dirichlet's theorem altogether, and simply use the pigeonhole principle to locate a b for which $\Lambda_b(n)$ has large mean. This observation has the interesting application that it allows one to extend the result in [26] to obtain arbitrarily long progressions, not just in the primes, but in fact in any subset of the primes (or almost primes) of positive relative density.

In summary, the W-trick allows one to easily eliminate the influence of small divisors, resulting in functions Λ_b, ν_b which are much more uniformly distributed than their non-renormalized counterparts Λ, ν. Of course, the price one pays for doing so is that the $o(1)$ error terms, as well as the c_k bounds employed above, deteriorate rather substantially; however if one is only interested in qualitative results then this trick is essentially cost-free.

4. Fourier obstructions to uniformity

We now discuss the problem of counting the progressions of length three in the primes. This can of course be done by the circle method, and this is essentially what we do here, but we shall adopt the philosophy of counting progressions by first establishing what the obstructions are to uniformity, and then dealing with

these obstructions in some manner. The W-trick is already one way to eliminate one obstruction to uniformity, namely irregular distribution when localized to small primes, which in the language of the circle method allows one to ignore the contribution of the major arcs (except the major arc near 1). We will see other ways to deal with obstructions to uniformity later in this article.

The standard way to count progressions of length three in the primes is to try to obtain asymptotics, or at least bounds, for the average

$$\mathbb{E}(\Lambda(a)\Lambda(a+r)\Lambda(a+2r)|1 \leqslant a, r \leqslant N). \tag{4.1}$$

Indeed Conjecture 1.1 already predicts an explicit asymptotic for this quantity, and Theorem 2.3 gives an upper bound which is only off by an absolute constant. One would then use the Fourier transform right away, to convert this expression to an integral involving an exponential sum such as $\mathbb{E}(\Lambda(n)e(-n\alpha)|1 \leqslant n \leqslant N)$, where α is a real number and $e(x) := e^{2\pi ix}$. This sum would then be estimated in two different ways, one when α is major arc (close to a rational with small denominator) and one when α is minor arc. The minor arc computation is reasonably elementary (ultimately relying on variants of the identity (2.6), the Cauchy-Schwarz inequality, and some bilinear cancellation in the expression $e(-nm\alpha)$) but the major arc computation is somewhat deeper, relying among other things on the Siegel-Walfisz theorem.

It turns out that one can proceed in a more elementary fashion if one is not seeking an asymptotic, but only a non-zero lower bound on the quantity (4.1) (which will certainly be enough to imply the qualitative result that there are infinitely many progressions of length three in the primes). Instead of needing to control the exponential sums of Λ, one only needs to control the exponential sums of a majorant ν or ν_b, which is much simpler. However, one does need one additional ingredient, namely *Roth's theorem* [41]. Roth's original formulation of this theorem asserts that any subset of the integers with positive upper density, necessarily contains infinitely many progressions of length three. Varnavides [50] showed that this qualitative version is in fact equivalent to the following more quantitative statement:

Theorem 4.1 (Quantitative Roth theorem)**.** [41],[50] *Let $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ be a function such that $0 \leqslant f(n) \leqslant 1$ for all $n \in \mathbb{Z}/N\mathbb{Z}$, and such that $\mathbb{E}(f(n)|n \in \mathbb{Z}/N\mathbb{Z}) \geqslant \delta$ for some $0 < \delta < 1$. Then we have*

$$\mathbb{E}(f(a)f(a+r)f(a+2r)|a, r \in \mathbb{Z}/N\mathbb{Z}) \geqslant c(\delta)$$

for some $c(\delta) > 0$.

The best value of $c(\delta)$ currently known is $c(\delta) \gg \delta^{C/\delta^2}$ for some absolute constant C, see [8]. However for the qualitative arguments we give below, we do not need to know the exact value of $c(\delta)$. We also do not need to know the proof of Theorem 4.1; we may treat it as a "black box". We do remark however that the

known proofs of this theorem, involving either Fourier analysis, ergodic theory, or graph theory, are extremely instructive and are very consistent with the philosophy outlined here of detecting obstructions to uniformity and then somehow dealing with each of the obstructions which occur. For us, the power of Roth's theorem lies in the fact that very little structural information is demanded of f (in particular, no arithmetic structure or Fourier-analytic structure is required), besides the important constraint that f is bounded[16].

At present, Roth's theorem does not directly allow us to obtain any non-trivial lower bound on (4.1) for two reasons. The first (rather trivial) reason is that we have stated Roth's theorem in $\mathbb{Z}/N\mathbb{Z}$ rather than on $\{1,\dots,N\}$, but there are some easy truncation arguments (which we omit) to pass back and forth between these two settings, possibly after modifying N by a factor of 2 or so. The more serious difficulty is that Λ is not bounded, and if we do normalize Λ to be bounded (e.g. by dividing by $\log N$) then δ becomes too small for Roth's theorem to be of any use. However, as it turns out it is relatively easy to *decompose* Λ into a bounded function (for which Roth's theorem is applicable) and a "uniform" error (which has a negligible impact on (4.1)).

Before we do this, we need to understand exactly what type of functions will give a negligible impact to expressions such as (4.1). To phrase things a little more concretely, let us work in the cyclic group $\mathbb{Z}/N\mathbb{Z}$ instead of the progression $\{1,\dots,N\}$, taking N to be odd, and consider an expression such as

$$\mathbb{E}(f(a)g(a+r)h(a+2r)|a,r\in\mathbb{Z}/N\mathbb{Z}) \tag{4.2}$$

for some functions $f,g,h:\mathbb{Z}/N\mathbb{Z}\to\mathbb{C}$. To begin the discussion let us take f,g,h to be bounded in magnitude by 1, although for applications to the primes we will eventually need to discard this hypothesis.

Since f,g,h are bounded by 1, it is clear that (4.2) is also bounded in magnitude by 1. However, in many cases, (4.2) will be much smaller than 1. For instance, if one of f,g,h is small in some averaged sense, say if the L^1 norm $\mathbb{E}(|f(n)||n\in\mathbb{Z}/N\mathbb{Z})$ is small, then (4.2) will be small also. Also, if one of f,g,h fluctuates randomly, for instance if $f(n)=\pm1$ for each n, with each $f(n)$ attaining $+1$ or -1 independently with equal probability, then it is easy to see that (4.2) will

[16]Indeed, our entire philosophy here is in some sense the polar opposite of the more conventional approach, in which one builds up as much information about the primes (or any other number-theoretic object) as possible, for instance using deep estimates on Dirichlet L-functions, and then uses all this information to then attack quantities such as (4.1). In contrast, we adopt a minimalist approach (in the spirit of sieve theory) in which we treat the primes as nothing more than a generic subset of the almost primes with positive relative density, ignoring all the rich arithmetic structure. That this approach works at all, is entirely due to the existence of such theorems as Roth's theorem, which apply to all sets of positive density (or bounded functions with large mean). However, as we shall see later it is possible to blend the two approaches and use deeper facts about the primes to obtain sharper results.

be quite small with high probability. Let us informally call a function *linearly uniform*[17] if the expression (4.2) is necessarily small as soon as at least one of f, g, h is set equal to this function. Thus for instance functions with small L^1 norm, or randomly fluctuating functions, will be linearly uniform. Since (4.2) is linear in f, g, and h separately, we thus see that we can modify f, g, or h by a linearly uniform function without significantly affecting (4.2), and so linearly uniform functions are "negligible" for the purposes of counting progressions of length three. On the other hand, from the identity

$$\mathbb{E}(e(\alpha a)e(-2\alpha(a+r))e(\alpha(a+2r))|a,r \in \mathbb{Z}/N\mathbb{Z}) = 1$$

for any $\alpha \in \frac{1}{N}\mathbb{Z}$, we see that the function $n \mapsto e(\alpha n)$ is not linearly uniform. More generally, since

$$\mathbb{E}(f(a)e(-2\alpha(a+r))e(\alpha(a+2r))|a,r \in \mathbb{Z}/N\mathbb{Z}) = \mathbb{E}(f(n)e(-\alpha n)|n \in \mathbb{Z}/N\mathbb{Z})$$

we see that any function f which has a large correlation (inner product) with a linear phase function $e(\alpha n)$, will not be linearly uniform. Thus linear phase functions are *obstructions* to linear uniformity; this may help explain the "linear" in the terminology "linear uniformity".

The effectiveness of the circle method, at least for the task of counting progressions of length three, ultimately lies in the fact that linear phase functions are the *only* obstructions to linear uniformity, at least when everything is bounded; thus if a bounded function has small correlation with every linear phase function, then it is linearly uniform. More precisely:

Lemma 4.2. *Let $f, g, h : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ be functions bounded by 1, and suppose that*

$$|\mathbb{E}(f(n)e(-\xi n/N)|n \in \mathbb{Z}/N\mathbb{Z})| \leqslant \varepsilon$$

for some $\varepsilon > 0$ and all $\xi \in \mathbb{Z}/N\mathbb{Z}$. Then we have

$$|\mathbb{E}(f(a)g(a+r)h(a+2r)|a,r \in \mathbb{Z}/N\mathbb{Z})| \leqslant \varepsilon.$$

Not co-incidentally, Lemma 4.2 is also the first step used in the Fourier-analytic proof of Roth's theorem; however, we will not discuss this connection here.

[17] The notation here is due to Gowers [20]. The term "uniform" arises because linearly uniform functions behave like a signed probabilistic point process with the uniform distribution; another possible terminology is "linearly unbiased". Somewhat confusingly, this usage of the word "uniform" is completely different from, and in fact in opposition to, the notion of "uniformly bounded"; indeed, we will later need to rely crucially on the fact that linearly uniform functions can be very far from being uniformly bounded.

Proof. Writing $\hat{f}(\xi) := \mathbb{E}(f(n)e(-\xi n/N)|n \in \mathbb{Z}/N\mathbb{Z})$ for all $\alpha \in \mathbb{Z}/N\mathbb{Z}$, and similarly for $\hat{g}$ and $\hat{h}$, we have the Fourier inversion formulae

$$f(a) = \sum_{\xi \in \mathbb{Z}/N\mathbb{Z}} \hat{f}(\xi)e(\xi a/N);$$
$$g(a+r) = \sum_{\lambda \in \mathbb{Z}/N\mathbb{Z}} \hat{g}(\lambda)e(\lambda(a+r)/N);$$
$$h(a+2r) = \sum_{\eta \in \mathbb{Z}/N\mathbb{Z}} \hat{h}(\eta)e(\eta(a+2r)/N).$$

Substituting these formulae and simplifying, we eventually obtain the identity

$$\mathbb{E}(f(a)g(a+r)h(a+2r)|a,r \in \mathbb{Z}/N\mathbb{Z}) = \sum_{\xi \in \mathbb{Z}/N\mathbb{Z}} \hat{f}(\xi)\hat{g}(-2\xi)\hat{h}(\xi). \tag{4.3}$$

On the other hand, from Plancherel's identity and the boundedness of g and h we have

$$\sum_{\xi \in \mathbb{Z}/N\mathbb{Z}} |\hat{g}(-2\xi)|^2 \leqslant 1; \quad \sum_{\xi \in \mathbb{Z}/N\mathbb{Z}} |\hat{h}(\xi)|^2 \leqslant 1$$

while from the hypothesis on f we have $|\hat{f}(\xi)| \leqslant \varepsilon$ for all ξ. The claim then follows from Hölder's inequality. □

Now we return to the task of estimating (4.1). Applying the W-trick to make Λ more uniformly distributed, it will suffice to obtain an estimate of the form

$$\mathbb{E}(\Lambda_b(a)\Lambda_b(a+r)\Lambda_b(a+2r)|1 \leqslant a,r \leqslant N) \geqslant c - o_{W\to\infty}(1) - o_{N\to\infty;W}(1)$$

for some absolute constant $c > 0$. Let us cheat a little bit by identifying $\{1,\dots,N\}$ with $\mathbb{Z}/N\mathbb{Z}$ (ignoring issues of truncation and wraparound, which are actually not difficult to deal with), so that we are now faced with establishing a lower bound for

$$\mathbb{E}(\Lambda_b(a)\Lambda_b(a+r)\Lambda_b(a+2r)|a,r \in \mathbb{Z}/N\mathbb{Z}). \tag{4.4}$$

We would like to use Lemma 4.2 to strip away the linearly uniform components of Λ_b. However, we are faced with the difficulty that Λ_b is not uniformly bounded. Fortunately, we can use the fact that Λ_b is majorized by an enveloping sieve ν_b. Actually we will not quite use the enveloping sieve ν_b constructed in the previous section, but use a slight variant $\tilde{\nu}_b$ which is closely related to the Selberg sieve. The enveloping sieve ν_b can be written down explicitly, but it is a little messy; see [27] for a definition, together with a full analysis and comparison of these two enveloping sieves. For this expository paper, suffice it to say that we still have the basic majorization

$$0 \leqslant \Lambda_b \leqslant C\tilde{\nu}_b \tag{4.5}$$

and that the Fourier coefficients of the Selberg enveloping sieve $\tilde{\nu}_b$ can be computed very explicitly; for instance one can show that

$$\widehat{\tilde{\nu}_b}(\xi) = o_{W\to\infty}(1) + o_{N\to\infty;W}(1) \tag{4.6}$$

for all $\xi \in \mathbb{Z}/N\mathbb{Z}\backslash\{0\}$. Using this and other bounds, together with orthogonality arguments such as those used in the large sieve (or of Tomas-Stein restriction theory), it is possible to obtain a weighted form of the Plancherel theorem, namely that

$$\|\hat{f}\|_{l^p(\mathbb{Z}/N\mathbb{Z})} \ll_p 1 \tag{4.7}$$

whenever $p > 2$ and $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ is bounded pointwise by $\tilde{\nu}_b + 1$; see [27] (and also [23]). The key point in these estimates is that no factor of $\log N$ appears on the right-hand side, despite the fact that all the L^q moments of Λ and ν (except the L^1 moment) contains such a logarithmic factor. Using this estimate we can obtain a weighted variant of Lemma 4.2:

Lemma 4.3. [27] *Let $f, g, h : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ be functions bounded in magnitude by $\tilde{\nu}_b + 1$, and suppose that*

$$|\mathbb{E}(f(n)e(-\xi n/N)|n \in \mathbb{Z}/N\mathbb{Z})| \leqslant \varepsilon$$

for some $\varepsilon > 0$ and all $\xi \in \mathbb{Z}/N\mathbb{Z}$. Then we have

$$|\mathbb{E}(f(a)g(a+r)h(a+2r)|a, r \in \mathbb{Z}/N\mathbb{Z})| \ll \varepsilon^{1/2}.$$

Proof. From (4.3) and Hölder's inequality we have

$$|\mathbb{E}(f(a)g(a+r)h(a+2r)|a, r \in \mathbb{Z}/N\mathbb{Z})| \leqslant \\ \leqslant \|\hat{f}\|^{1/2}_{l^\infty(\mathbb{Z}/N\mathbb{Z})}\|\hat{f}\|^{1/2}_{l^{5/2}(\mathbb{Z}/N\mathbb{Z})}\|\hat{g}\|_{l^{5/2}(\mathbb{Z}/N\mathbb{Z})}\|\hat{h}\|_{l^{5/2}(\mathbb{Z}/N\mathbb{Z})}$$

(for instance). From hypothesis we have $\|\hat{f}\|_{l^\infty(\mathbb{Z}/N\mathbb{Z})} \leqslant \varepsilon$. The claim now follows from (4.7). □

Thus, even when considering functions that are merely bounded by $\tilde{\nu}_b + 1$ instead of bounded by 1, it is still the case that linear phase functions are the only obstruction to orthogonality. One can view this as a weak version of Plancherel's theorem, transferred to the enveloping sieve $\tilde{\nu}_b + 1$.

At this point one could try to show that Λ_b, or more precisely the normalized function $\Lambda_b - 1$, has small correlation with all linear phase functions,

$$\mathbb{E}((\Lambda_b(n) - 1)e(-\xi n/N)|n \in \mathbb{Z}/N\mathbb{Z}) = o_{W\to\infty}(1) + o_{N\to\infty;W}(1).$$

This, together with Lemma 4.3, would imply that Λ_b can be replaced with 1 with negligible error in (4.4) and we would conclude that

$$\mathbb{E}(\Lambda_b(a)\Lambda_b(a+r)\Lambda_b(a+2r)|a, r \in \mathbb{Z}/N\mathbb{Z}) = 1 + o_{W\to\infty}(1) + o_{N\to\infty;W}(1),$$

which would of course be consistent with the Hardy-Littlewood prime tuples conjecture. This strategy can indeed be carried out, though it requires a Vinogradov-type analysis of exponential sums; it also gives the correct asymptotic for (4.1). Indeed, this is essentially the approach taken by van der Corput when establishing infinitely many progressions of length three in the primes. However, there is a more "low-tech" approach that will give the same qualitative result (but not the asymptotic). Roughly speaking[18], the idea is as follows. We allow for the possibility that exponential sums $\mathbb{E}(\Lambda_b(n)e(-\alpha n)|n \in \mathbb{Z}/N\mathbb{Z})$ could be large, thus providing some additional obstructions to uniformity. However, the estimate (4.7) limits the total number of obstructions that could exist. More precisely, if we introduce a threshold $0 < \varepsilon < 1$ and let $S \subset \mathbb{Z}/N\mathbb{Z}$ denote the exceptional frequencies ξ which obstruct linear uniformity, in the sense that

$$|\mathbb{E}(\Lambda_b(n)e(-\xi n/N)|n \in \mathbb{Z}/N\mathbb{Z})| \geqslant \varepsilon,$$

then (4.7) shows that $|S| \ll_\varepsilon 1$. The Vinogradov exponential sum technique will eventually show that S consists only of the zero frequency 0 for W, N large enough, but we will avoid using this fact, instead treating S as a set for which the only information known is the cardinality bound. This approach has the advantage of being more flexible, for instance we will also be able to recover the result of Green [23] that any subset of the primes with positive relative density contains infinitely many progressions of length three.

The set S represents all the obstructions to uniformity. We can remove these obstructions by the device of *conditional expectation*, which is a slightly different way than the W-trick of removing non-uniformities, though certainly in the same philosophical spirit. One considers the *Bohr set* $B(S, \rho) \subset \mathbb{Z}/N\mathbb{Z}$ for some small radius $0 < \rho < 1$ defined by

$$B(S, \rho) := \{n \in \mathbb{Z}/N\mathbb{Z} : \|n\xi\|_{\mathbb{R}/\mathbb{Z}} < \rho \text{ for all } \xi \in S\},$$

where $\|x\|_{\mathbb{R}/\mathbb{Z}}$ denotes the distance from x to the nearest integer. One should think of this Bohr set as being roughly analogous to the subgroup $W\mathbb{Z}$ of $\mathbb{Z}$, thus translates $x + B(S, \rho)$ are the analogues of residue classes modulo W. When executing the W-trick, we passed to a single residue class; here, however, we shall proceed in a more "ergodic" fashion, averaging out the effect of each translate $x + B(S, \rho)$. More precisely we split

$$\Lambda_b = \Lambda_{b,U^\perp} + \Lambda_{b,U}$$

where $\Lambda_{b,U^\perp}$ is the "anti-linearly-uniform" component

$$\Lambda_{b,U^\perp}(x) := \Lambda_{b,U^\perp} * \frac{N}{|B(S,\rho)|}1_{B(S,\rho)} * \frac{N}{|B(S,\rho)|}1_{B(S,\rho)}(x)$$

[18]For the detailed rigourous argument, see [27].

where the convolution $f * g$ on $\mathbb{Z}_N$ is defined by

$$f * g(x) := \mathbb{E}(f(n)g(x-n)|n \in \mathbb{Z}/N\mathbb{Z}),$$

and $\Lambda_{b,U}(x)$ is the "linearly uniform component"

$$\Lambda_{b,U} := \Lambda_b - \Lambda_{b,U^\perp}.$$

The function $\Lambda_{b,U^\perp}$ encapsulates all the obstructions to linear uniformity encountered by Λ_b; the convolution kernel

$$K := \frac{N}{|B(S,\rho)|}1_{B(S,\rho)} * \frac{N}{|B(S,\rho)|}1_{B(S,\rho)}$$

can be thought of as a sort of "Fejér kernel" adapted to $B(S,\rho)$. A key observation is that unlike Λ_b, the function $\Lambda_{b,U^\perp}$ is *bounded*. Indeed, from the majorization (4.5) we have

$$0 \leqslant \Lambda_{b,U^\perp}(x) \ll \tilde{\nu}_b * K(x)$$

and then by using Fourier expansion of $1_{B(S,\rho)}$ and (4.6) one can show

$$\tilde{\nu}_b * K(x) \ll 1 + o_{W\to\infty;|S|,\rho}(1) + o_{N\to\infty;W,|S|,\rho}(1).$$

Since $|S| \ll_\varepsilon 1$, we thus have the uniform boundedness

$$0 \leqslant \Lambda_{b,U^\perp}(x) \ll 1 + o_{W\to\infty;\varepsilon,\rho}(1) + o_{N\to\infty;W,\varepsilon,\rho}(1). \tag{4.8}$$

In particular we see that $\Lambda_{b,U}$ is pointwise bounded by a constant multiple of $\tilde{\nu}_b + 1$. Also, since the kernel K is normalized to have mean 1, we have

$$\mathbb{E}(\Lambda_{b,U^\perp}(x)|x \in \mathbb{Z}/N\mathbb{Z}) = \mathbb{E}(\Lambda_b(x)|x \in \mathbb{Z}/N\mathbb{Z}) = 1 + o_{W\to\infty}(1) + o_{N\to\infty;W}(1).$$

Thus $\Lambda_{b,U^\perp}$ is bounded, non-negative and has large mean, and so Roth's theorem can be applied (after a renormalization by a bounded scalar) to conclude

$$\mathbb{E}(\Lambda_{b,U^\perp}(a)\Lambda_{b,U^\perp}(a+r)\Lambda_{b,U^\perp}(a+2r)|a,r \in \mathbb{Z}/N\mathbb{Z}) \geqslant c - o_{W\to\infty;\varepsilon,\rho}(1) - o_{N\to\infty;W,\varepsilon,\rho}(1) \tag{4.9}$$

for some absolute constant $c > 0$.

The function $\Lambda_{b,U}$ can be regarded as the portion of Λ_b remaining after all the obstructions to uniformity have been removed. By the definition of S, one can easily show that $\Lambda_{b,U^\perp}$ has small correlation with all linear phase functions:

$$|\mathbb{E}(\Lambda_{b,U}(n)e(-\xi n/N)|n \in \mathbb{Z}/N\mathbb{Z})| \ll \varepsilon + \rho \text{ for all } \xi \in \mathbb{Z}/N\mathbb{Z},$$

and thus by several applications of Lemma 4.3 we can replace Λ_b by $\Lambda_{b,U^\perp}$ with a small error:

$$\begin{aligned}\mathbb{E}(\Lambda_b(a)\Lambda_b(a+r)\Lambda_b(a+2r)|a,r \in \mathbb{Z}/N\mathbb{Z}) \\ = \mathbb{E}(\Lambda_{b,U^\perp}(a)\Lambda_{b,U^\perp}(a+r)\Lambda_{b,U^\perp}(a+2r)|a,r \in \mathbb{Z}/N\mathbb{Z}) + O(\varepsilon+\rho).\end{aligned}$$

Applying (4.9) we conclude that

$$\mathbb{E}(\Lambda_b(a)\Lambda_b(a+r)\Lambda_b(a+2r)|a,r \in \mathbb{Z}/N\mathbb{Z}) \geqslant c/2$$

if ε, ρ are sufficiently small, W is sufficiently large depending on ε, ρ, and N is sufficiently large depending on ε, ρ, W. This is enough to establish infinitely arithmetic progressions of length threein the primes, and more generally‘in any subset of the primes with positive relative density. Similar arguments work for other sets that are fairly large and which can be dominated by a suitable enveloping sieve. For instance, in [27] it was shown that there were infinitely many arithmetic progressions p_1, p_2, p_3 in the primes, where the numbers p_1+2, p_2+2, p_3+2 are either prime or the product of two primes; this is achieved by combining the arguments above with (a quantitative version of) the famous result of Chen [9] that there are infinitely many primes p such that $p+2$ is the product of at most two primes.

5. Quadratic obstructions to uniformity

Let us now consider the task of counting progressions of length four in the primes, or more precisely of obtaining an asymptotic for

$$\mathbb{E}(\Lambda(a)\Lambda(a+r)\Lambda(a+2r)\Lambda(a+3r)|1 \leqslant a, r \leqslant N).$$

The Hardy-Littlewood prime tuples conjecture predicts that this quantity is equal to $\prod_p \alpha_p + o_{N\to\infty}(1)$, where α_p is the local density

$$\alpha_p := \mathbb{E}(\Lambda_{\mathbb{Z}/p\mathbb{Z}}(a)\Lambda_{\mathbb{Z}/p\mathbb{Z}}(a+r)\Lambda_{\mathbb{Z}/p\mathbb{Z}}(a+2r)\Lambda_{\mathbb{Z}/p\mathbb{Z}}(a+3r)|a, r \in \mathbb{Z}/p\mathbb{Z}).$$

To put it another way, the number of progressions $a, a+r, a+2r, a+3r$ of primes with $1 \leqslant a, r \leqslant N$ is predicted to be $\frac{N^2}{\log^4 N}(\prod_p \alpha_p + o_{N\to\infty}(1))$. The result of [26] establishes a lower bound

$$\mathbb{E}(\Lambda(a)\Lambda(a+r)\Lambda(a+2r)\Lambda(a+3r)|1 \leqslant a, r \leqslant N) \geqslant c - o_{N\to\infty}(1)$$

for some absolute constant $c > 0$, which is enough to establish infinitely many progressions of length four in the primes, but does not give the asymptotic. In this section we describe a more recent (though significantly more complicated) approach in [28], [29], [30] which will give the correct asymptotic:

Theorem 5.1. [28], [29], [30] *We have*

$$\mathbb{E}(\Lambda(a)\Lambda(a+r)\Lambda(a+2r)\Lambda(a+3r)|1 \leqslant a, r \leqslant N) = \prod_p \alpha_p + o_{N\to\infty}(1).$$

We now sketch the main ideas of proof of this theorem. Firstly, by the W-trick, it will suffice to show that

$$\mathbb{E}(\Lambda_{b_0}(a)\Lambda_{b_1}(a+r)\Lambda_{b_2}(a+2r)\Lambda_{b_3}(a+3r)|1 \leqslant a, r \leqslant N) = 1 + o_{W\to\infty}(1) + o_{N\to\infty;W}(1)$$

for all $b_0, \ldots, b_3$ coprime to W. Let us again cheat a little bit by identifying $\{1, \ldots, N\}$ with $\mathbb{Z}/N\mathbb{Z}$ (ignoring some minor truncation issues), so that we now

wish to prove that

(5.1)

$$\mathbb{E}(\Lambda_{b_0}(a)\Lambda_{b_1}(a+r)\Lambda_{b_2}(a+2r)\Lambda_{b_3}(a+3r)|a,r\in\mathbb{Z}/N\mathbb{Z}) = 1+o_{W\to\infty}(1)+o_{N\to\infty;W}(1).$$

It is convenient to take N to be a prime. We are thus faced with the problem of understanding quartilinear expressions such as

$$\mathbb{E}(f(a)g(a+r)h(a+2r)j(a+3r)|a,r\in\mathbb{Z}/N\mathbb{Z}); \tag{5.2}$$

to begin the discussion let us suppose that f,g,h,j are bounded in magnitude by 1. Let us informally call a function *quadratically uniform* if the above expression is automatically small whenever one of f,g,h,j is replaced with that function. As in the preceding section, it is easy to see that linear phase functions obstruct quadratic uniformity; however, a new difficulty arises in that *quadratic* phase functions such as $e(\alpha n^2)$ also obstruct quadratic uniformity. This can be seen for instance by the identity

$$\begin{aligned}\mathbb{E}(f(a)e(-3\alpha(a+r)^2)e(3\alpha(a+2r)^2)e(-\alpha(a+3r)^2)|a,r\in\mathbb{Z}/N\mathbb{Z})\\ = \mathbb{E}(f(n)e(-\alpha n^2)|n\in\mathbb{Z}/N\mathbb{Z}).\end{aligned}$$

More generally, one can show that any *quadratic nilsequence* of the form $F(g^n x)$, where $g\in G$ lives in a 2-step nilpotent Lie group G, x lives in a compact quotient[19] G/Γ of G by a closed subgroup Γ, and $F: G/\Gamma\to\mathbb{C}$ is a continuous function, will similarly be an obstruction to quadratic uniformity; see [28]. The quadratic phases $e(\alpha n^2)$ are good examples of quadratic nilsequence; another example is the generalized quadratic phase $e(\lfloor\alpha n\rfloor\lfloor\beta n\rfloor\gamma)$ for some real numbers α,β,γ, though strictly speakign one needs to smooth out the greatest integer function $\lfloor x\rfloor$ in order to genuinely obtain a quadratic nilsequence.

The appearance of these quadratic phases shows that the circle method is now insufficient to establish quadratic uniformity; functions such as $e(\alpha n^2)$ can give significant contributions to (5.2) despite having very small Fourier coefficients. However, quadratic uniformity can still be captured by the very useful *Gowers uniformity norms*[20] $U^d(\mathbb{Z}/N\mathbb{Z})$, defined recursively for $d=0,1,\ldots$ as

$$\begin{aligned}\|f\|_{U^0(\mathbb{Z}/N\mathbb{Z})} &:= \mathbb{E}(f(x)|x\in\mathbb{Z}/N\mathbb{Z});\\ \|f\|_{U^{d+1}(\mathbb{Z}/N\mathbb{Z})} &= \mathbb{E}(\|T^h f\overline{f}\|^{2^d}_{U^d(\mathbb{Z}/N\mathbb{Z})}|h\in\mathbb{Z}/N\mathbb{Z})^{1/2^{d+1}}\end{aligned}$$

[19] There is an intriguing superficial similarity between the emergence of the 2-step nilmanifolds G/Γ which arise in the analysis of progressions of length 4, and the cusp manifolds $SL_2(\mathbb{R})/\Gamma$ which appear for instance in Kloosterman's refinement of the Hardy-Littlewood circle method (which of course corresponds to the unit circle $\mathbb{R}/\mathbb{Z}$). However, we do not know of a concrete connection between these two different extensions of the circle method.

[20] These are genuine norms for $d\geqslant 2$; see [21], [26], [25], [46].

where T^h is the shift operator $T^h f(x) := f(x+h)$, thus for instance

$$\begin{aligned}
\|f\|_{U^1(\mathbb{Z}/N\mathbb{Z})} &= |\mathbb{E}(\overline{f(n)}f(n+h)|n,h \in \mathbb{Z}/N\mathbb{Z})|^{1/2} \\
&= |\mathbb{E}(f)| \\
\|f\|_{U^2(\mathbb{Z}/N\mathbb{Z})} &= |\mathbb{E}(f(n)\overline{f(n+h_1)f(n+h_2)}f(n+h_1+h_2)|n,h_1,h_2 \in \mathbb{Z}/N\mathbb{Z})| \\
&= (\sum_{\xi \in \mathbb{Z}/N\mathbb{Z}} |\hat{f}(\xi)|^4)^{1/4} \\
\|f\|_{U^3(\mathbb{Z}/N\mathbb{Z})} &= |\mathbb{E}(\overline{f(n)}f(n+h_1)f(n+h_2)f(n+h_3) \\
&\quad \overline{f(n+h_1+h_2)f(n+h_1+h_3)f(n+h_2+h_3)}f(n+h_1+h_2+h_3) \\
&\quad |n,h_1,h_2,h_3 \in \mathbb{Z}/N\mathbb{Z})|.
\end{aligned}$$

The relationship between Gowers uniformity norms, and quadratic (or higher order) uniformity, is given by

Lemma 5.2 (Generalized von Neumann theorem)**.** *Let $k \geqslant 3$, and let $N \geqslant k-1$ be prime. If $f_0, \ldots, f_{k-1} : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ are bounded in magnitude by 1, then*

$$|\mathbb{E}(f_0(a)f_1(a+r)\ldots f_{k-1}(a+(k-1)r)|a,r \in \mathbb{Z}/N\mathbb{Z})| \leqslant \inf_{0 \leqslant j \leqslant k} \|f_j\|_{U^{k-1}(\mathbb{Z}/N\mathbb{Z})}.$$

In particular we have

$$|\mathbb{E}(f_0(a)f_1(a+r)f_2(a+2r)f_3(a+3r)|a,r \in \mathbb{Z}/N\mathbb{Z})| \leqslant \inf_{0 \leqslant j \leqslant 3} \|f_j\|_{U^3(\mathbb{Z}/N\mathbb{Z})}.$$

This lemma can be deduced from $k-1$ applications of the Cauchy-Schwarz inequality, interspersed with $k-1$ applications of the van der Corput identity

$$|\mathbb{E}(f(n)|n \in \mathbb{Z}/N\mathbb{Z})|^2 = \mathbb{E}(T^h f(n)\overline{f}(n)|n,h \in \mathbb{Z}/N\mathbb{Z});$$

we leave the details to the reader (or see [20], [21], [34], [26], [44], [25], [46]).

The above lemma shows that functions with small $U^3(\mathbb{Z}/N\mathbb{Z})$ norm are quadratically uniform. As before, this lemma is not directly applicable to the problem of finding progressions in primes, since functions such as Λ_b are not bounded. However, because Λ_b can be bounded by an enveloping sieve ν_b which obeys the good correlation estimates in (3.3), we can use the following extension of the generalized von Neumann theorem:

Lemma 5.3 (Relative generalized von Neumann theorem)**.** [26] *Let $k \geqslant 3$, and let $N > k-1$ be prime. If $f_0, \ldots, f_{k-1} : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ are such that f_j is bounded by $\nu_{b_j} + 1$ for some b_j coprime to W, then (if $R = N^{c_k}$ for some sufficiently small c_k)*

$$\begin{aligned}
|\mathbb{E}(f_0(a)\ldots f_{k-1}(a+(k-1)r)|a,r \in \mathbb{Z}/N\mathbb{Z})| \ll \\
\ll_k \inf_{0 \leqslant j \leqslant k} \|f_j\|_{U^{k-1}(\mathbb{Z}/N\mathbb{Z})} + o_{N\to\infty;W,k}(1) + o_{W\to\infty;k}(1).
\end{aligned}$$

This lemma is more complicated to prove than Lemma 5.2 but is still primarily an application of the Cauchy-Schwarz inequality; see[21] [26], with a heavy reliance on the linear forms estimates (3.3). Note that this generalization of Lemma 5.2 is consistent with the transference principle mentioned earlier.

In light of this lemma, we see that in order to establish the asymptotic (5.1), it will suffice to show that $\Lambda_b - 1$ is quadratically uniform, or more precisely that

$$\|\Lambda_b - 1\|_{U^3(\mathbb{Z}/N\mathbb{Z})} = o_{N\to\infty;W}(1) + o_{W\to\infty}(1) \tag{5.3}$$

for all b coprime to W. This is not easy to do directly, since the quantity $\|\Lambda_b - 1\|_{U^3(\mathbb{Z}/N\mathbb{Z})}$ is basically the same type of expression that appears in the Hardy-Littlewood prime tuples conjecture, and is beyond the reach of the circle method. Nevertheless, one can proceed by locating all the obstructions to quadratic uniformity, and then checking that the function $\Lambda_b - 1$ is orthogonal to all of these.

We have already observed that the quadratic nilsequences $F(g^n x)$ are obstructions to quadratic uniformity. Recent developments [34], [5] in ergodic theory strongly suggest[22] that these are in fact the only obstructions to quadratic uniformity. By building on the pioneering combinatorial and analytical technology of Gowers [20], a quantitative version of this assertion was made in [28]. More precisely:

Theorem 5.4 (Inverse theorem for $U^3(\mathbb{Z}/N\mathbb{Z})$)**.** [28] *Let $0 < \eta < 1$. Then there exists a collection $\mathcal{N}$ of $O_\eta(1)$ triples (G, Γ, F), where G is a 2-step nilpotent Lie group, Γ is a closed co-compact subgroup of G, and $F : G/\Gamma \to \mathbb{C}$ is a smooth function, with the following property: if N is an odd prime and $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ is bounded by 1 and is such that $\|f\|_{U^3(\mathbb{Z}/N\mathbb{Z})}$, then there exists a triple (G, Γ, F) from this collection, a group element $g \in G$, a point $x \in G/\Gamma$, and a shift $h \in \mathbb{Z}/N\mathbb{Z}$ such that*

$$|\mathbb{E}(T^h f(n)\overline{F(g^n x)}| - N/2 < n < N/2)| \gg_\eta 1.$$

One can explicitly describe the collection $\mathcal{N}$, and give quantitative bounds on the dimension of G/Γ and the smoothness of F, as well as the dependence of the implied constant on η; see [28].

The proof of Theorem 5.4 is quite lengthy, using many tools of Gowers in additive combinatorics and Fourier analysis. On the other hand, it may well

[21] The argument in [26] treats the case when all the b_j are equal, but one can easily modify it to treat the case of distinct b_j.

[22] Roughly speaking, the ergodic theory setting corresponds to considering averages such as $\mathbb{E}(f(a)f(a+r)f(a+2r)f(a+3r)|1 \leqslant a \leqslant N, 1 \leqslant r \leqslant H)$ where the shift range H goes to infinity much more slowly than N does. As such, there does not appear to be a direct "correspondence principle" between the results in [34], [5] and the type of results considered here, but there is certainly a very strong analogy between the two. See [35] for more on the ergodic theory perspective to these problems.

be that a "softer" proof, without the quantitative bounds, is available by the ergodic-theory methods in [34], [5]. In [30], the results from [26] (and more precisely, Theorem 6.2 below) were used to extend Theorem 5.4 to the case when f is merely bounded by $\nu_b + 1$ rather than by 1; again, this is consistent with the transference principle. By applying this extended version of Theorem 5.4, we see that one can prove (5.3) as soon as one demonstrates the asymptotic orthogonality estimate

(5.4)

$$\mathbb{E}((T^h\Lambda_b(n)-1)\overline{F(g^nx)}|-N/2<n<N/2) = o_{N\to\infty;W,F,G,\Gamma}(1)+o_{W\to\infty;F,G,\Gamma}(1)$$

for all quadratic nilsequences $F(g^nx)$.

This type of result is essentially an exponential sum estimate on Λ, and can thus be attacked by the standard Vinogradov-type methods. A model case is the estimate

$$\mathbb{E}((\Lambda_b(n)-1)e(-\alpha n^2)|1\leqslant n\leqslant N) = o_{N\to\infty}(1)$$

for all $\alpha\in\mathbb{R}$, which was essentially obtained in [16]. The general case of quadratic nilsequences is treated in [29], [30]. In those papers it is convenient to first prove the preliminary estimate

$$\mathbb{E}(\mu(n)\overline{F(g^nx)}|1\leqslant n\leqslant N) \ll_{A,F,G,\Gamma} \log^{-A} N$$

for all $A>0$ whenever F is smooth; see [30]. This can be considered a generalization of Davenport's estimate[10]

$$\mathbb{E}(\mu(n)e(-\alpha n)|1\leqslant n\leqslant N) \ll_A \log^{-A} N$$

and is proven by broadly similar, though significantly more technical, methods (in particular, Vaughan's identity, a division into major and minor arcs, and Cauchy-Schwarz type arguments to deal with the minor arcs). It is however simpler to deal with the Möbius function $\mu(n)$ than the modified von Mangoldt function $\Lambda_b(n)-1$, as μ is bounded, and also obeys a somewhat more pleasant Vaughan identity than Λ. Using this estimate and some elementary arguments, it is already possible to establish

$$\mathbb{E}((T^h\Lambda_b(n)-T^h\Lambda_{R,\varphi,b}(n))\overline{F(g^nx)}|-N/2<n<N/2) = \\ = o_{N\to\infty;W,F,G,\Gamma}(1)+o_{W\to\infty;F,G,\Gamma}(1)$$

where $\Lambda_{R,\varphi,b}(n) := \frac{W}{\phi(W)}\Lambda_{R,\varphi}(Wn+b)$ and $\Lambda_{R,\varphi}$ was defined[23] in (2.7); as usual we set R to be a small power of N and φ to be a suitable cutoff function. By the triangle inequality, it thus remains to verify that

$$\mathbb{E}((T^h\Lambda_{R,\varphi,b}(n)-1)\overline{F(g^nx)}|-N/2<n<N/2) = o_{N\to\infty;W,F,G,\Gamma}(1)+o_{W\to\infty;F,G,\Gamma}(1).$$

It turns out that the simplest way to do this is to apply the Cauchy-Schwarz inequality (in the spirit of Lemma 5.2 and Lemma 5.3, and in particular on the

[23] Actually, any reasonable truncated divisor sum approximation to Λ could be used in place of $\Lambda_{R,\varphi}$ here.

Gowers-Cauchy-Schwarz inequality introduced in [21], and also playing a key role in [26]), to reduce matters to the U^3 estimate

$$\|\Lambda_{R,\varphi,b}(n) - 1\|_{U^3(\mathbb{Z}/N\mathbb{Z})} = o_{N\to\infty;W}(1) + o_{W\to\infty}(1),$$

which in turn can be established by a Goldston-Yıldırım correlation estimate, similar in spirit to (3.3). See [30].

It is entirely possible that the techniques discussed in this section extend to give an asymptotic for longer progressions in the primes, though there are serious new difficulties that appear (similar to the new difficulties that appear in [21] when compared against [20]). We (in joint work with Ben Green) hope to report on this problem in a future paper.

6. Ergodic obstructions to uniformity

In the previous section, we outlined a rather complicated approach that yielded an asymptotic for the number of progressions of length four in the primes. As we already saw though in the length three case, it can often be significantly easier to establish the weaker result of a non-trivial lower bound for the number of such progressions, using tools such as Roth's theorem. This was achieved in [26], in particular establishing that the primes contain arbitrarily long arithmetic progressions. The argument can be seen as a variant of the above arguments, but in which the "hard" obstructions of nilsequences are replaced by much "softer" obstructions coming from ergodic averages. These soft obstructions are insufficiently explicit to easily allow for establishing asymptotic orthogonality results such as (5.4), but they are still controllable to the extent that one can modify the arguments of Section 4, using the soft obstructions to build generalized Bohr sets with which to split Λ_b into a uniform component, which is negligible, and an anti-uniform component, which can be treated by a theorem of Szemerédi.

We turn to the details. The famous theorem of Szemerédi [43] asserts that every subset of integers of positive density contains arbitrarily long arithmetic progressions. A quantitative version of this theorem, which generalizes Theorem 4.1, is as follows:

Theorem 6.1 (Quantitative Szemerédi theorem)**.** *Let $k \geqslant 1$, and let $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ be a function such that $0 \leqslant f(n) \leqslant 1$ for all $n \in \mathbb{Z}/N\mathbb{Z}$, and such that $\mathbb{E}(f(n)|n \in \mathbb{Z}/N\mathbb{Z}) \geqslant \delta$ for some $0 < \delta < 1$. Then we have*

$$\mathbb{E}(f(a)f(a+r)\dots f(a+(k-1)r)|a, r \in \mathbb{Z}/N\mathbb{Z}) \geqslant c(k,\delta)$$

for some $c(k,\delta) > 0$.

This theorem can be deduced from Szemerédi's original theorem from the averaging argument of Varnavides [50]; see also [44] for a direct proof.

As in Section 4, the task (after applying the W-trick) is to obtain a non-trivial lower bound for

$$\mathbb{E}(\Lambda_b(a)\dots\Lambda_b(a+(k-1)r)|a,r\in\mathbb{Z}/N\mathbb{Z}), \tag{6.1}$$

where we once again gloss over the distinction between $\mathbb{Z}/N\mathbb{Z}$ and $\{1,\dots,N\}$ to simplify the discussion. Again, we cannot apply Theorem 6.1 directly because of the unboundedness of Λ_b. However, we can proceed by establishing the following structure theorem, that decomposes any non-negative function bounded by the enveloping sieve ν_b into a Gowers uniform component (with small Gowers uniformity norm), a non-negative bounded component, and a small error.

Theorem 6.2 (Structure theorem)**.** [26] *Let $k\geqslant 1$, and let $R=N^{c_k}$ for some sufficiently small $c_k>0$. Let $f:\mathbb{Z}/N\mathbb{Z}\to\mathbb{R}$ be such that $0\leqslant f(n)\leqslant\nu_b(n)$. Let $0<\varepsilon<1$. Then functions $f_U, f_{U^\perp}:\mathbb{Z}/N\mathbb{Z}\to\mathbb{C}$ such that*

$$\|f_U\|_{U^{k-1}(\mathbb{Z}/N\mathbb{Z})}=o_{\varepsilon\to 0;k}(1) \tag{6.2}$$

and

$$0\leqslant f_{U^\perp}(n)\leqslant 1+o_{\varepsilon\to 0;k}(1)+o_{N\to\infty;\varepsilon,k}(1) \tag{6.3}$$

and

$$0\leqslant f_U(n)+f_{U^\perp}(n)\leqslant f(n)$$

for all $n\in\mathbb{Z}/N\mathbb{Z}$. Furthermore, we have

$$\mathbb{E}(|f(n)-f_{U^\perp}(n)-f_U(n)||n\in\mathbb{Z}/N\mathbb{Z})=o_{\varepsilon\to 0;k}(1). \tag{6.4}$$

and

$$\mathbb{E}(f_{U^\perp}(n)|n\in\mathbb{Z}/N\mathbb{Z})=\mathbb{E}(f(n)|n\in\mathbb{Z}/N\mathbb{Z})+o_{\varepsilon\to 0;k}(1). \tag{6.5}$$

Assuming this theorem, a lower bound for (6.1) can be easily accomplished. By (4.5) we can apply Theorem 6.2 with $f:=c\Lambda_b$ for some absolute constant $c>0$, to obtain a majorization

$$0\leqslant f_U+f_{U^\perp}\leqslant c\Lambda_b.$$

It then suffices to obtain a lower bound for

$$\mathbb{E}((f_U+f_{U^\perp})(a)\dots(f_U+f_{U^\perp})(a+(k-1)r)|a,r\in\mathbb{Z}/N\mathbb{Z}).$$

All the terms involving at least one factor of f_U are $o_{\varepsilon\to 0;k}(1)+o_{N\to\infty;\varepsilon,k}(1)$, thanks mainly to (6.2) and Lemma 5.3. The remaining term involving $f_{U^\perp}$ is at least $c_k-o_{\varepsilon\to 0;k}(1)-o_{N\to\infty;\varepsilon,k}(1)$, thanks to Theorem 6.1 and (6.5). Setting ε suitably small, and then N sufficiently large, we obtain a non-trivial lower bound for (6.1).

Thus Theorem 6.2 allows one to transfer Theorem 6.1 to a relative setting, adapted to the enveloping sieve ν_b. A similar argument also allows one to use Theorem 6.2 to transfer Theorem 5.3 to the relative setting; see [30].

It remains to prove Theorem 6.2. Let us fix f. The first guess is to take $f_{U^\perp}$ to be the mean of f, $f_{U^\perp} := \mathbb{E}(f)$, and then set $f_U := f - f_{U^\perp}$. It is clear that $f_{U^\perp}$ is non-negative, and also

$$f_{U^\perp} = \mathbb{E}(f) \leqslant \mathbb{E}(\nu_b) = 1 + o_{W\to\infty;k}(1) + o_{N\to\infty;W,k}(1).$$

Also we trivially have (6.5) and (6.4). The only difficulty is that we do not necessarily have (6.2); there is no reason why f_U needs to be Gowers uniform (i.e. have small $U^{k-1}(\mathbb{Z}/N\mathbb{Z})$ norm). However, if this is the case, it turns out to be possible to locate a precise obstruction which is preventing f_U from being uniform, and transfer this obstruction from f_U to $f_{U^\perp}$. This may not remove all the non-uniformity from f_U, but it will increase the energy ($L^2(\mathbb{Z}/N\mathbb{Z})$ norm) of $f_{U^\perp}$ by a significant amount, and so after iterating this process a finite number of times we will eventually end up with a Gowers uniform f_U.

The above type of argument has also been used before in ergodic theory (most notably in Furstenberg's structure theorem [14]), and also in the proof of the Szemerédi regularity lemma [43]; not co-incidentally, both of those cited papers concerned Szemerédi's theorem (Theorem 6.1). The argument in Section 4 involving convolution with a Bohr set generated by all the Fourier obstructions to uniformity is also an argument of this type (although in that case one transferred all the obstructions from f_U to $f_{U^\perp}$ at once, rather than one at a time). The main difficulty in executing the above idea is to maintain (6.3) throughout this procedure, i.e. to keep $f_{U^\perp}$ non-negative and bounded by 1 (plus negligible errors). To achieve the non-negativity, the simplest way is to use the machinery of conditional expectation (as is done in Furstenberg's structure theorem, and implicitly in the Szemerédi regularity lemma). To achieve the boundedness, one needs some control on the obstructions to uniformity that one is transferring to $f_{U^\perp}$. In the Fourier-analytic argument, these obstructions are linear phase functions $e(\alpha n)$, and one can use Fourier-analytic control in the enveloping sieve (see (4.6)) to keep $f_{U^\perp}$ bounded. To adopt a similar argument in the general case, one might imagine one would need a similarly explicit description of these obstructions, for instance using the nilsequences of the preceding section. However, it turns out that one can get by using a much less explicit obstruction to uniformity, first introduced in ergodic theory[24].

In order to make the above strategy rigourous, we need two basic concepts, that of a *dual function* and that of *conditional expectation*. The dual function $\mathcal{D}_d f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ of a function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ is defined recursively for $d =$

[24] More precisely, the key observation for ergodic theory is that the obstructions to weak mixing (which roughly corresponds to Gowers uniformity) are given by almost periodic functions, and more specifically given any function f which fails to be weakly mixing (so that $\langle T^h f, f\rangle$ does not converge on average to zero), one can construct the non-trivial almost periodic function $F := \lim_{H\to\infty} \mathbb{E}(\langle T^h f, f\rangle T^h f | -H \leqslant h \leqslant H)$, which has a positive correlation with f. See for instance [14]; for the connection with the Gowers uniformity norms see [34], [35].

$0, 1, 2, \ldots$ by the formula

$$\mathcal{D}_0 f = 1; \quad \mathcal{D}_{d+1} f = \mathbb{E}(\mathcal{D}_d(f\overline{T^h f})T^h f | h \in \mathbb{Z}/N\mathbb{Z});$$

thus for instance

$$\begin{aligned}
\mathcal{D}_1 f(n) &= \mathbb{E}(f) \\
\mathcal{D}_2 f(n) &= \mathbb{E}(f(n+h_1)f(n+h_2)\overline{f(n+h_1+h_2)}|h_1, h_2 \in \mathbb{Z}/N\mathbb{Z}) \\
&= \mathbb{E}(\langle f, T^h f\rangle T^h f(n)|h \in \mathbb{Z}/N\mathbb{Z}) \\
&= \sum_{\xi \in \mathbb{Z}/N\mathbb{Z}} |\hat{f}(\xi)|^2 f(\xi) e(n\xi/N) \\
\mathcal{D}_3 f(n) &= \mathbb{E}(f(n+h_1)f(n+h_2)f(n+h_3) \\
&\quad \overline{f(n+h_1+h_2)f(n+h_1+h_3)f(n+h_2+h_3)} \\
&\quad f(n+h_1+h_2+h_3)|h_1, h_2, h_3 \in \mathbb{Z}/N\mathbb{Z})
\end{aligned}$$

where $\langle , \rangle$ denotes the usual inner product $\langle f, g\rangle = \mathbb{E}(f\overline{g})$. One can easily use induction to verify that

$$\langle f, \mathcal{D}_{k-1} f\rangle = \|f\|^{2^{k-1}}_{U^{k-1}(\mathbb{Z}/N\mathbb{Z})}. \tag{6.6}$$

Thus if f fails to be Gowers uniform of order $k-1$, it correlates with a dual function $\mathcal{D}_{k-1} f$. These dual functions will serve as our obstructions to Gowers uniformity; they are simple to describe but are not very explicit, as they involve a function f for which we have only limited control. Nevertheless, there is a large amount of averaging contained in the non-linear operator $\mathcal{D}_{k-1}$, which will allow us to obtain satisfactory control on these dual functions.

To proceed further, we need to understand the properties of dual functions better. The first important (and easy) property is that dual functions are always bounded: more precisely, we have $|\mathcal{D}_{k-1} f| \ll_k 1$ whenever f is pointwise bounded by $\nu_b + 1$. Indeed, in such a case we have

$$|\mathcal{D}_{k-1} f| \leqslant \mathcal{D}_{k-1}(\nu_b + 1),$$

and several applications of (3.3) gives the bound $\mathcal{D}_{k-1}(\nu_b + 1)$ (see [26]).

The second important (but significantly deeper) property is that a dual function, and more generally any polynomial combination of dual functions, is highly "Gowers anti-uniform" in the sense that it is essentially orthogonal to all Gowers uniform functions, and in particular to the function $\nu_b - 1$ (which can easily be shown to be Gowers uniform, thanks to several applications of (3.3)). Indeed, it turns out that we have

$$\langle \nu_b - 1, P(\mathcal{D}_{k-1}(f_1), \ldots \mathcal{D}_{k-1}(f_m))\rangle = o_{N\to\infty;m,P,W}(1) + o_{W\to\infty;m,P}(1) \tag{6.7}$$

for any polynomial $P(x_1, \ldots, x_m)$ of m variables, and any functions $f_1, \ldots, f_m : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ bounded in magnitude by $\nu_b + 1$. This fact is elementary to prove,

but not entirely trivial; it is obtained by a large number of applications of the Cauchy-Schwarz and Hölder inequalities, combined with the correlation condition (3.4). See [26].

One should compare the above facts with the situation in the Fourier-analytic argument. In that argument, the role of dual functions was played by the linear phase functions $e(\alpha n)$, which are certainly bounded. A polynomial combination of linear phase functions is nothing more than a trigonometric polynomial, and (4.6) then shows that $\nu - 1$ is indeed mostly orthogonal to such polynomial combinations.

To exploit these facts about dual functions, we need to introduce the machinery of σ-algebras and conditional expectation.

Definition 6.3. A *σ-algebra* is a collection $\mathcal{B}$ of subsets of $\mathbb{Z}/N\mathbb{Z}$ which contains $\emptyset$ and $\mathbb{Z}/N\mathbb{Z}$ and is closed under union, intersection, and complementation. A function $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ is *$\mathcal{B}$-measurable* if all its level sets lie in $\mathcal{B}$. If $\mathcal{B}$ is a σ-algebra and $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$, we define the *conditional expectation* $\mathbb{E}(f|\mathcal{B}) : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ of f with respect to $\mathcal{B}$ to be the function

$$\mathbb{E}(f|\mathcal{B})(x) := \mathbb{E}(f|\mathcal{B}(x)) = \frac{1}{\mathcal{B}(x)} \sum_{n \in \mathcal{B}(x)} f(n)$$

for all $x \in \mathbb{Z}/N\mathbb{Z}$ where $\mathcal{B}(x)$ is the smallest set in $\mathcal{B}$ which contains x. If $\mathcal{B}_1, \mathcal{B}_2$ are two σ-algebras, we use $\mathcal{B}_1 \vee \mathcal{B}_2$ to denote the smallest σ-algebra which contains both $\mathcal{B}_1$ and $\mathcal{B}_2$.

A basic fact in measure theory is that any algebra of functions generates a σ-algebra. The estimate (6.7) asserts, morally speaking, that $\nu_b - 1$ is asymptotically orthogonal to the algebra generated by dual functions, and thus should also be orthogonal to the σ-algebra generated by dual functions. Indeed, we can make this precise as follows. Given any dual function $\mathcal{D}_{k-1}(f)$ and any cutoff $\varepsilon > 0$, we can generate a σ-algebra $\mathcal{B}_\varepsilon(\mathcal{D}_{k-1}(f))$, by partitioning the complex plane $\mathbb{C}$ into squares of length ε, and using the inverse images of these squares under $\mathcal{D}_{k-1}(f)$ as the atoms of the σ-algebra. There is some choice in how to choose this partition; a random translation of the standard partition will work here. A key result in [26] is then that for any $m \geqslant 1$ and any functions $f_1, \ldots, f_m$ bounded in magnitude by $\nu_b + 1$, we have the uniform distribution property

$$\mathbb{E}(\nu_b - 1|\mathcal{B}_\varepsilon(\mathcal{D}_{k-1}(f_1)) \vee \ldots \vee \mathcal{B}_\varepsilon(\mathcal{D}_{k-1}(f_m)) = \\ = o_{\varepsilon \to 0;m,k}(1) + o_{W \to \infty;m,k,\varepsilon}(1) + o_{N \to \infty;m,k,\varepsilon,N}(1) \tag{6.8}$$

except on an exceptional set Ω which is small in the sense that

$$\mathbb{E}((\nu_b + 1)1_\Omega) = o_{\varepsilon \to 0;m,k}(1) + o_{W \to \infty;m,k,\varepsilon}(1) + o_{N \to \infty;m,k,\varepsilon,N}(1).$$

This claim can be derived fairly quickly from (6.7) and the Weierstrass approximation theorem[25]; see [26].

We can now sketch the proof of Theorem 6.2. As mentioned earlier, the idea is to detect any obstructions to uniformity in f_U (in the guise of dual functions $\mathcal{D}_{k-1}(f_1), \ldots, \mathcal{D}_{k-1}(f_m)$, where $f_1, \ldots, f_m$ are bounded in magnitude by $\nu_b + 1$) and transfer them to $f_{U^\perp}$ one at a time. Oversimplifying somewhat (in particular, glossing over the role of the exceptional set Ω), the algorithm for doing so is as follows:

- Step 0. Set $m = 0$.
- Step 1. Set $f_{U^\perp} := \mathbb{E}(f|\mathcal{B}_\varepsilon(\mathcal{D}_{k-1}(f_1)) \vee \ldots \vee \mathcal{B}_\varepsilon(\mathcal{D}_{k-1}(f_m))$ (so initially we would just have $f_{U^\perp} = \mathbb{E}(f)$), and then set $f_U := f - f_{U^\perp}$. Clearly $f_{U^\perp}$ is non-negative and has the same mean as f; from (6.8) we ensure that $f_{U^\perp}$ is bounded.
- Step 2. If f_U is Gowers uniform, in the sense that $\|f_U\|_{U^{k-1}(\mathbb{Z}/N\mathbb{Z})} \leqslant \varepsilon^{1/2}$, then we are done. Otherwise, we set $f_{m+1} := f_U$, increment m by 1, and return to Step 1.

It turns out that every time we return from Step 2 to Step 1, the energy $\mathbb{E}(|f_{U^\perp}|^2)$ of $f_{U^\perp}$ increases by at least $c_{\varepsilon,k}$ (plus some negligible $o(1)$ errors), where $c_{\varepsilon,k} > 0$ is an explicit positive quantity depending only on ε and k; see [26]. Intuitively, the reason for this is as follows. If f_U is not Gowers uniform, then by (6.6) f_U as a large correlation with $\mathcal{D}_{k-1}(f_U) = \mathcal{D}_{k-1}(f_{m+1})$. But f_U, by construction, is orthogonal to all the functions which are measurable with respect to the σ-algebra $\mathcal{B}_\varepsilon(\mathcal{D}_{k-1}(f_1)) \vee \ldots \vee \mathcal{B}_\varepsilon(\mathcal{D}_{k-1}(f_m)$, while $\mathcal{D}_{k-1}(f_{m+1})$ lies (modulo negligible errors) in the larger σ-algebra $\mathcal{B}_\varepsilon(\mathcal{D}_{k-1}(f_1)) \vee \ldots \vee \mathcal{B}_\varepsilon(\mathcal{D}_{k-1}(f_{m+1})$. The energy increment then follows (morally, at least) from the following simple lemma:

Lemma 6.4 (Correlation implies energy increment)**.** *Let $\mathcal{B} \subseteq \mathcal{B}'$ be σ-algebras, and let f, g be functions such that f is orthogonal to all $\mathcal{B}$-measurable functions, while g is $\mathcal{B}'$-measurable and bounded in magnitude by* 1*. Then we have the energy increment*

$$\mathbb{E}(|\mathbb{E}(f|\mathcal{B}')|^2) \geqslant \mathbb{E}(|\mathbb{E}(f|\mathcal{B})|^2) + |\langle f, g\rangle|^2.$$

Proof. From the $\mathcal{B}'$-measurability of g we have

$$\langle f, g\rangle = \langle \mathbb{E}(f|\mathcal{B}'), g\rangle.$$

Also, since f is orthogonal to all $\mathcal{B}$-measurable functions, we have $\mathbb{E}(f|\mathcal{B}) = 0$. Thus

$$\langle f, g\rangle = \langle \mathbb{E}(f|\mathcal{B}') - \mathbb{E}(f|\mathcal{B}), g\rangle.$$

[25] As our functions here are complex valued, we have to consider polynomials which involve the conjugates of the dual functions $\mathcal{D}_{k-1}(f_j)$ as well as the dual functions themselves, but this does not cause any additional difficulty

Applying Cauchy-Schwarz and the boundedness of g we conclude

$$\mathbb{E}(|\mathbb{E}(f|\mathcal{B}') - \mathbb{E}(f|\mathcal{B})|^2) \geqslant |\langle f, g\rangle|^2$$

and the claim then follows from Pythagoras' theorem. □

In practice, we cannot quite use this simple lemma because of the presence of the exceptional sets Ω, but it is still possible to obtain the energy increment by carefully modifying the above argument; see [26].

Observe that the energy $\mathbb{E}(|f_{U^\perp}|^2)$ increments by a fixed factor at each stage of the iteration, but remains bounded independently of the number of steps of the iteration (ignoring some negligible $o(1)$ type errors). Thus the algorithm can only run for a bounded number of steps, which keeps all the $o(1)$ errors under control. After doing all the book-keeping, one eventually arrives at a proof of Theorem 6.2; see [26] for the full details. As discussed earlier, this is enough to establish that the primes contain arbitrarily long arithmetic progressions; the same argument also shows that any subset of the primes of positive relative density contain arbitrarily long arithmetic progressions. One can also follow through the argument carefully to eventually yield a lower bound

$$\mathbb{E}(\Lambda(a)\dots\Lambda(a+(k-1)r)|1 \leqslant a, r \leqslant N) \geqslant c(k) - o_{N\to\infty;k}(1)$$

for some explicitly computable $c(k) > 0$; the exact value is rather poor, depending on both the quantitative error bounds in the correlation estimates (3.3), (3.4), as well the constant in Theorem 6.1.

7. Further directions

The transference methods here should be applicable to some other situations. For instance, a variant of the above argument was used recently in [48] to show that the Gaussian primes in $\mathbb{Z}[i]$ contain infinitely many constellations of any prescribed shape and orientation; one needs to replace Szemerédi's theorem by the somewhat stronger "hypergraph removal lemma" of Gowers [22] and Rödl-Skokan [39], [40] (see also [47]), and the presence of the conjugation operation $z \mapsto \overline{z}$ in the Galois group $Gal(\mathbb{Q}[i]/\mathbb{Q})$ causes some technical difficulties, but otherwise the strategy is almost identical. We refer the reader to [47] and [48] for further details. Similar results should also hold for other number fields that enjoy unique factorization. For instance, one should be able to show that given any finite field F, the monic irreducible polynomials of one variable in $F[x]$ should contain affine subspaces over F of arbitrarily high dimension.

A more challenging extension would be to obtain a multidimensional relative Szemerédi theorem, which would assert that given any dimension $d \geqslant 1$, and given the set of primes $P = \{2, 3, 5, \dots\}$, that any subset of P^d of positive relative density should contain infinitely many constellations of any prescribed shape and orientation. For P^d replaced by $\mathbb{R}^d$, this result was proven in [13], and

also follows from the hypergraph removal lemma mentioned briefly earlier. A major new difficulty here is that the natural enveloping sieve for P^d is not very pseudorandom, even after applying the higher-dimensional analogue of the W-trick; the lack of pseudorandomness, even for P^2, can be seen by the observation that if the two acute corners of a right-angled triangle (with sides parallel to the axes) lie in P^2, then the third corner also automatically lies in P^2, despite P^2 being quite sparse. We do not know how to resolve this problem.

It should also be possible to establish arbitrarily long progressions $a, a + r, \ldots, a+(k-1)r$ in the primes (or any positive relative density subset thereof), in which the spacing r is significantly smaller than the base point a, obtaining for instance progressions such that $r = O_{\varepsilon,k}(a^\varepsilon)$ for any given ε. This is likely to follow by localizing the above theory to intervals of length $O(N^\varepsilon)$ in $\{N+1, \ldots, 2N\}$.

A more difficult result would be to obtain a polynomial Szemerédi theorem for the primes. More precisely, if $P_1, \ldots, P_k : \mathbb{Z} \to \mathbb{Z}$ were any polynomials mapping the integers to the integers with $P_1(0) = \ldots = P_k(0) = 0$, then there should be infinitely many k-tuplets $a + P_1(r), \ldots, a + P_k(r)$ with $r \neq 0$, such that all the $a + P_j(r)$ are prime. If the primes were replaced by a positive density subset of $\mathbb{Z}$, then this result was obtained by Bergelson and Leibman [6]. If one wished to localize this problem to $\mathbb{Z}/N\mathbb{Z}$, it would be necessary to restrict r to be at most a small power of N, and so one may first have to understand the previous problem concerning progressions with small spacing before tackling this problem. The hypothesis $P_1(0) = \ldots = P_k(0) = 0$ seems to unfortunately be rather crucial to the method (for instance, one can easily construct counterexamples to the Bergelson-Leibman theorem without this hypothesis), which is a pity as one would otherwise have a route to prove such conjectures as the twin primes conjecture or more generally the Hardy-Littlewood prime tuple conjecture.

Another problem (communicated by Vitaly Bergelson) which might now be feasible is to establish that the set $P-1 = \{1, 2, 4, 6, 10, \ldots\}$ formed by decrementing one from each prime, is an IP set, or more precisely given any k there exist distinct $a_1, \ldots, a_k$ such that the finite sums $\{\sum_{j \in J} a_j : J \subseteq \{1, \ldots, k\}; J \neq \emptyset\}$ are contained in $P - 1$. The case $k = 2$ can be handled by the circle method, but the higher k remain open. Such a result would then lead to a number of combinatorial consequences, see for instance [7] for further discussion.

References

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

[2] A. Balog, *The prime k-tuplets conjecture on average,* Analytic Number Theory (Allerton Parl, IL. 1989), 47–75, Progr. Math. **85**. Birkhäuser Boston, 1990.

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

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

[5] V. Bergelson, B. Host and B. Kra, *Multiple recurrence and nilsequences*, preprint.

[6] V. Bergelson and A. Leibman, *Polynomial extensions of van der Waerden's and Szemerédi's theorems,* J. Amer. Math. Soc. **9** (1996), 725–753.
[7] V. Bergelson, I. Ruzsa, *Squarefree numbers, IP sets and ergodic theory*, "Paul Erdos and his Mathematics I", Bolyai Society Mathematical Studies, 11, Budapest (2002), 147-160.
[8] J. Bourgain, *On triples in arithmetic progression*, GAFA **9** (1999), 968–984.
[9] J.-R. Chen, *On the representation of a large even integer as the sum of a prime and a product of at most two primes,* Sci. Sinica **16** (1973), 157–176.
[10] H. Davenport, *On some infinite series involving arithmetical functions. II*, Quart. J. Math. Oxf. **8** (1937), 313–320
[11] P. Erdős, P. Turán, *On some sequences of integers*, J. London Math. Soc. **11** (1936), 261–264.
[12] H. Furstenberg, *Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions*, J. Analyse Math. **31** (1977), 204–256.
[13] H. Furstenberg, Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*. J. Analyse Math. **34** (1978), 275–291.
[14] H. Furstenberg, Y. Katznelson and D. Ornstein, *The ergodic-theoretical proof of Szemerédi's theorem,* Bull. Amer. Math. Soc. **7** (1982), 527–552.
[15] 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.
[16] A. Ghosh, *The distribution of* αp^2 *modulo* 1, Proc. London Math. Soc. (3) 42 (1981), no. 2, 252–269.
[17] D. Goldston and C.Y. Yıldırım, *Higher correlations of divisor sums related to primes, I: Triple correlations,* Integers **3** (2003) A5, 66pp.
[18] D. Goldston and C.Y. Yıldırım, *Higher correlations of divisor sums related to primes, III: k-correlations,* preprint (available at AIM preprints)
[19] D. Goldston and C.Y. Yıldırım, *Small gaps between primes, I,* preprint.
[20] T. Gowers, *A new proof of Szemerédi's theorem for arithmetic progressions of length four*, GAFA **8** (1998), 529–551.
[21] T. Gowers, *A new proof of Szemerédi's theorem*, GAFA **11** (2001), 465-588.
[22] T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, preprint.
[23] B.J. Green, *Roth's theorem in the primes,* preprint.
[24] B.J. Green, *A Szemerédi-type regularity lemma in abelian groups,* preprint.
[25] B.J. Green, *Finite field models in arithmetic combinatorics*, preprint.
[26] B.J. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions,* preprint.
[27] B.J. Green and T. Tao, *Restriction theory of Selberg's sieve, with applications,* preprint.
[28] B.J. Green and T. Tao, *An inverse theorem for the Gowers* U^3 *norm*, preprint.
[29] B.J. Green and T. Tao, *Quadratic uniformity of the Möbius function*, preprint.
[30] B.J. Green and T. Tao, *Two linear equations in four prime unknowns*, preprint.
[31] 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
[32] D.R. Heath-Brown, *Three primes and an almost prime in arithmetic progression,* J. London Math. Soc. (2) **23** (1981), 396–414.
[33] D.R. Heath-Brown, *Linear relations amongst sums of two squares,* Number theory and algebraic geometry — to Peter Swinnerton-Dyer on his 75th birthday, CUP (2003).
[34] B. Host, B. Kra, *Non-conventional ergodic averages and nilmanifolds,* to appear in Ann. Math.
[35] B. Kra, *The Green-Tao Theorem on arithmetic progressions in the primes: an ergodic point of view*, preprint.

[36] A. Kumchev, D. Tolev, *An invitation to additive prime number theory*, Serdica Math. J. **31** (2005), 1–74.
[37] O. Ramaré, *On Snirel'man's constant,* Ann. Scu. Norm. Pisa **21** (1995), 645–706.
[38] O. Ramaré and I.Z. Ruzsa, *Additive properties of dense subsets of sifted sequences,* J. Th. Nombres de Bordeaux **13** (2001) 559–581.
[39] V. Rödl, J. Skokan, *Regularity lemma for k-uniform hypergraphs*, to appear, Random Structures and Algorithms.
[40] V. Rödl, J. Skokan, *Applications of the regularity lemma for uniform hypergraphs*, preprint.
[41] K.F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 245-252.
[42] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Acta Math. Acad. Sci. Hungar. **20** (1969), 89–104.
[43] ———, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 299–345.
[44] T. Tao, *A quantitative ergodic theory proof of Szemerédi's theorem*, preprint.
[45] T. Tao, *A remark on Goldston-Yildirim correlation estimates*, unpublished.
[46] T. Tao, *Arithmetic progressions in the primes*, El Escorial conference proceedings.
[47] T. Tao, *A variant of the hypergraph removal lemma*, preprint.
[48] T. Tao, *The Gaussian primes contain arbitrarily shaped constellations*, preprint.
[49] J.G. van der Corput, *Über Summen von Primzahlen und Primzahlquadraten,* Math. Ann. **116** (1939), 1–50.
[50] P. Varnavides, *On certain sets of positive density,* J. London Math. Soc. **34** (1959) 358–360.
[51] I.M. Vinogradov, *Representation of an Odd Number as a Sum of Three Primes*, Comptes rendus (Doklady) de l'Académie des Sciences de l'U.R.S.S. **15** (1937a), 169–172.
[52] T. Ziegler, *Universal characteristic factors and Furstenberg averages*, preprint.
[53] ———, *A non-conventional ergodic theorem for a nilsystem*, preprint.

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