Astérisque

JULIA WOLF

Arithmetic and polynomial progressions in the primes [after Gowers, Green, Tao and Ziegler]

Astérisque, tome 352 (2013), Séminaire Bourbaki,
exp. n° 1054, p. 389-427

<http://www.numdam.org/item?id=AST_2013__352__389_0>

Séminaire BOURBAKI Mars 2012
64e année, 2011-2012, n° 1054, p. 389 à 427

ARITHMETIC AND POLYNOMIAL PROGRESSIONS IN THE PRIMES
[after Gowers, Green, Tao and Ziegler]

by **Julia WOLF**

1. INTRODUCTION

In 2004 Green and Tao **[25]** proved the following groundbreaking result.

THEOREM 1 (Green-Tao theorem). — *The primes contain arbitrarily long arithmetic progressions. Moreover, the same is true of any subset of the primes of positive relative density.*[(1)(2)]

Theorem 1 vastly generalizes van der Corput's result **[11]** that there are infinitely many 3-term arithmetic progressions in the primes, as well as a significant strengthening due to Green **[22]**, which established the existence of 3-term progressions in any subset of the primes of positive relative density. It also represents a special case of a conjecture by Erdős and Turán **[12]**, dating back to 1936.

CONJECTURE 2 (Erdős-Turán conjecture). — *Any subset* $X \subseteq \mathbb{N}$ *satisfying*

$$\sum_{x \in X} \frac{1}{x} = +\infty$$

contains arbitrarily long arithmetic progressions.

What is truly remarkable about Theorem 1 is the diversity of methods which are brought together in its proof: it combines tools from arithmetic combinatorics (especially so-called higher-order Fourier analysis) with traditional analytic number

(1) It is easy to see that the primes cannot contain an infinite arithmetic progression. Suppose that $P(j) = a + jd$ for some $a, d \in \mathbb{N}$ takes prime values for $j = 0, 1, \ldots, k$. Then $P(a) \equiv 0 \bmod a$ and $P(a) > a$. But if $P(a) = ma$ for some integer $m > 1$, it is no longer prime and hence $k < a$.

(2) The longest currently known arithmetic progression in the primes, $43\ 142\ 746\ 595\ 714\ 191 + 23\ 681\ 770 \times 223\ 092\ 870 \times j$, where $j = 0, 1, \ldots 25$, was found by Périchon with software by Wróblewski and Reynolds in 2010.

theory, all while taking inspiration from ergodic theory. The Erdős-Turán conjecture suggests that the existence of arithmetic structure in the primes is in fact merely a consequence of their density, and so it is perhaps not surprising that analytic and combinatorial (rather than classical number-theoretic) methods should play a major role. Indeed, in 1975 Szemerédi [**47**] showed in a purely combinatorial fashion that any subset of the integers of positive upper density contains arbitrarily long arithmetic progressions.[3] Denoting by $[N]$ the set of integers $\{1, 2, \ldots, N\}$, we have the following finitary version of this statement.

THEOREM 3 (Szemerédi's theorem). — *Suppose that $A \subseteq [N]$ is a subset of density α which contains no k-term arithmetic progressions. Then*

$$\alpha = o_k(1),$$

where $o_k(1)$ is a quantity that tends to zero as N tends to infinity.

However, our current understanding of the decay rate of α does not allow us to immediately deduce Theorem 1. Indeed, the best known bound on the density of a subset of $[N]$ that contains no 3-term progressions, due to recent work of Sanders [**43**], is of the form $(\log N)^{-(1-o(1))}$, falling just short of the density of the primes. For longer progressions, the discrepancy is much more alarming. For length 4, the best known bound on α is of the form $\exp(-c\sqrt{\log \log N})$ [**26**], while for longer progressions it is $(\log \log N)^{-c}$ for some small positive constant c depending on k [**17**]. On the other hand, the best known example of a 3-term progression free set has density $\exp(-c\sqrt{\log N})$ [**3**], which is believed by many to be closer to the truth. However, proving upper bounds of this shape seems very much out of reach of currently available techniques (but see a recent result of Schoen and Shkredov [**46**]).

Many excellent expository articles have been written on the proof of the Green-Tao theorem [**23, 36, 48**]; in particular, it was covered in the Séminaire Bourbaki in 2005 by Bernard Host [**33**]. In contrast to Host's ergodic theoretic perspective we adopt a more analytic viewpoint in the present exposition. Moreover, our main focus will be on the developments that have taken place since the original proof of the Green-Tao theorem, which have brought new understanding and a number of additional exciting results to the subject.

Shortly after the proof of Theorem 1, Green and Tao [**31**] extended their result from arithmetic progressions to solutions of more general systems of linear equations in the primes. This work covered essentially all systems of linear equations for which the conclusion is neither trivially false nor known to be extremely difficult (such as those systems related to Goldbach's conjecture or the twin primes problem), but was

[3] A qualitative proof was given by Furstenberg [**14**] in 1977 and initiated the long-standing and fruitful interaction between combinatorics and ergodic theory.

conditional on two conjectures: the *inverse conjecture for the uniformity norms*, and the *Möbius nilsequences conjecture*. The latter was established by Green and Tao **[28]** shortly afterwards, and the former very recently by the same authors in joint work with Ziegler **[30]**. With the completion of this very substantial research programme the authors are able to assert not only the existence of general linear patterns in the primes, but give precise asymptotics for their frequency in the interval $[N]$.

In a further step towards generalization, Tao and Ziegler **[50]** proved in 2008 that the primes contain arbitrarily long polynomial progressions.

THEOREM 4 (Tao-Ziegler theorem). — *Given polynomials $P_1, \dots, P_k \in \mathbb{Z}[m]$ such that $P_1(0) = \cdots = P_k(0) = 0$, there exist infinitely many integers x, m such that $x + P_1(m), \dots, x + P_k(m)$ are simultaneously prime. Moreover, the same is true of any subset of the primes of positive relative density.*

The first non-trivial example of such a polynomial pattern is a configuration consisting of two elements that differ by a square, which corresponds to $P_1(m) = 0$, $P_2(m) = m^2$. In dense subsets of the integers the existence of such a configuration is guaranteed by a theorem of Sárközy **[45]**, which is obtained using a sophisticated application of the circle method. In fact, and in contrast with the situation for 3-term arithmetic progressions described above, the best known bound in Sárközy's theorem is strong enough to directly imply the existence of square differences in any positive-density subset of the primes. In the case of more general polynomial configurations, however, the results from arithmetic combinatorics are very far from implying a statement resembling that of Theorem 4. Worse, there is currently *no quantitative theorem at all* in the literature asserting that if a subset $A \subseteq [N]$ is dense enough, then it contains a polynomial configuration of the above type.[4] [5] What we do have is a *qualitative* polynomial Szemerédi theorem due to Bergelson and Leibman **[6]** proved by ergodic theoretic methods, whose statement is fundamental to the proof of Theorem 4. Moreover, Tao and Ziegler rely heavily on an induction technique which allowed Bergelson and Leibman to linearize a system of polynomials in successive stages, known as *PET induction* **[4]**.

The general strategy of proof for Theorem 4 is largely the same as in the case of (linear) arithmetic progressions. There are two main novelties here: first, a *transference principle* is explicitly formulated for the first time, which was implicit in and absolutely fundamental to the proof of the Green-Tao theorem. Roughly speaking, the problem is that the von Mangoldt function (a weighted indicator function of the

[4] A paper by Green **[21]**, which proves the existence of a 3-term progression whose common difference is a sum of two squares in any dense subset of the integers, may be regarded as an exception.
[5] There are, however, colouring results of this type, see the combinatorial proof of the polynomial van der Waerden theorem by Walters **[54]**.

primes) is unbounded, while the quantitative techniques from arithmetic combinatorics only apply to bounded functions. However, it turns out that the von Mangoldt function can be majorized by a so-called *pseudorandom measure* which is quite well-behaved. In particular, it can be shown that many statements involving truly bounded functions hold also, by transference, for functions that are bounded by this pseudorandom measure. The transference principle has received significant simplifications and a new conceptual context through work done by Gowers [**18**], and it is this more recent viewpoint that we shall adopt in our exposition. A similar approach was independently discovered by Reingold, Trevisan, Tulsiani and Vadhan [**40**] in theoretical computer science, where the transference principle is known as the *dense model theorem* and has found several applications in the context of complexity theory.

The second novelty concerns a new family of norms. In his work on Szemerédi's theorem, Gowers [**17**] introduced the U^k *norms*, often called *uniformity norms* or *Gowers norms*, and showed that the $(k+1)$-term progression count of a function is approximately invariant under small perturbations in the U^k norm – in other words, the uniformity norms control long arithmetic progressions. This raises the question of what can be said about functions that are large in the U^k norm, which is answered by the so-called *inverse theorem*. It is one of the central results in arithmetic combinatorics (although for $k > 3$, its strong form was only a conjecture until very recently) and states, roughly speaking, that if the U^k norm of a function is large, then the function correlates with a polynomial structure of degree $k-1$.[(6)] The inverse theorem, together with the above-mentioned approximate invariance under small perturbations in the U^k norm, is essentially sufficient to prove Szemerédi's theorem.

It is not too difficult to see that the uniformity norms are not sufficient for controlling polynomial configurations. To put it very simply, the reason is that in a linear configuration such as $x, x+d, x+2d$, which defines a 3-term progression, the range of both variables x and d is essentially linear in N. In contrast, in a configuration such as $x, x+m^2$, which represents a square difference, the range of m has to be restricted to $\sqrt{N}$. Dealing with smaller parameter ranges required Tao and Ziegler to introduce new *local uniformity norms*, and study some of their properties.

To conclude this section we give a brief overview of the structure of this paper. In Section 2 we define the uniformity norms and develop some of the fundamental notions of higher-order Fourier analysis, following Gowers's harmonic analysis approach to Szemerédi's theorem. In Section 3 we show how one uses the existence of a pseudorandom measure and the transference principle together with Szemerédi's theorem to obtain the Green-Tao theorem on arithmetic progressions in the primes. Finally, in

(6) The uniformity norms have also appeared in the context of ergodic theory. A deep result of Host and Kra [**34**] on the structure of characteristic factors for certain multiple ergodic averages is in some sense analogous to the above-mentioned inverse theorem.

Section 4, we give more detail on the additional ingredients that are needed to treat polynomial configurations and obtain a proof of the Tao-Ziegler theorem.

We shall give computations mostly in model cases. The reader wishing to skip the proofs should be able to do so at only moderate cost.

2. HIGHER-ORDER FOURIER ANALYSIS AND SZEMERÉDI'S THEOREM

The first non-trivial case of Szemerédi's theorem, namely the existence of 3-term progressions in sufficiently dense subsets of the integers, was established by Roth [41] in 1953. Here and in the sequel the symbol $\ll$ stands for "is bounded above by a constant times".

THEOREM 5 (Roth's theorem). — *Suppose that $A \subseteq [N]$ is a subset of density α which contains no 3-term arithmetic progressions. Then*

$$\alpha \ll (\log\log N)^{-1}.$$

The proof proceeds via a dichotomy between randomness and structure, which relies heavily on the Fourier transform. Given a function $f : \mathbb{Z}_N \to \mathbb{C}$, we define its Fourier transform $\widehat{f} : \mathbb{Z}_N \to \mathbb{C}$, for each $t \in \mathbb{Z}_N$, by

$$\widehat{f}(t) = \mathbb{E}_x f(x) \exp(2\pi i x t/N),$$

where, as is standard in the field, we use the expectation operator $\mathbb{E}_{x\in\mathbb{Z}_N}$ to denote the normalized sum $\frac{1}{N}\sum_{x\in\mathbb{Z}_N}$.[7] Now given $A \subseteq [N]$, if all non-trivial Fourier coefficients of the characteristic function 1_A are "small" (the trivial one, with the above normalization, being equal to the density α), then we can count the number of 3-term progressions in A precisely. This is because a set whose Fourier coefficients are small is distributed somewhat *uniformly* in the interval $[N]$, and, for the purpose of counting 3-term progressions, behaves like a random set whose elements are chosen independently with probability α. Specifically, the number of 3-term progressions in such a uniform set of density α is roughly $\alpha^3 N^2$, which is the number expected in the random case. On the other hand, if A is *non-uniform*, then there must exist a large Fourier coefficient. By the definition of the Fourier transform, this means that the characteristic function of A exhibits a bias in some preferred "direction", i.e. it has increased density on the (approximate) level set of at least one non-trivial character. This level

[7] One often prefers to work with the discrete Fourier transform on the finite abelian group $\mathbb{Z}_N$ instead of $[N] \subseteq \mathbb{Z}$. It is not difficult to embed $[N]$ into the slightly larger group $\mathbb{Z}_{N'}$ for some prime N' while sacrificing a factor of at most a constant. In the sequel we shall therefore make no distinction between $[N]$ and $\mathbb{Z}_N$, and switch between the two settings whenever convenient.

set can be seen to contain a very long arithmetic progression (of length tending to infinity with N), and so by focusing on A restricted to this progression and rescaling, the argument can be iterated. The process stops when the density of A exceeds 1, which is clearly absurd. The trade-off between the density increase obtained at each step, and the size of the arithmetic progression on which this increase is obtained, gives rise to the bound on α stated above.[8]

Unfortunately, the first part of this dichotomy breaks down when one tries to count progressions of length at least 4: it is *not* true that a set which is uniform in the Fourier sense defined above always contains the expected number of 4-term progressions.

EXAMPLE 1. — *The set* $A \subseteq \mathbb{Z}_N$ *defined by*

$$A = \{x \in \mathbb{Z}_N : x^2 \in [-\alpha N, \alpha N)\}$$

is uniform in the sense that $\sup_{t \neq 0} |\widehat{1}_A(t)|$ *is "small", but it contains "too many" 4-term progressions.*

Indeed, it is straightforward to show, using Gauss sums, that the Fourier transform on the non-trivial frequencies is very small (of the order of $N^{-(1/2-o(1))}$). It is perhaps less immediate to confirm that this set contains many more than the expected number of 4-term progressions, which is easily seen to be approximately $\alpha^4 N^2$. The reason is the elementary quadratic identity

$$x^2 - 3(x+d)^2 + 3(x+2d)^2 - (x+3d)^2 = 0,$$

which holds for all $x, d \in \mathbb{Z}_N$. Indeed, since $(x+3d)^2$ is a small linear combination of x^2, $(x+d)^2$ and $(x+2d)^2$, the event that $x+3d$ lies in A is *not* independent of $x, x+d$ and $x+2d$ being in A, and so the number of 4-term progressions in A is more like $c\alpha^3 N^2$ for some absolute constant c. [9]

2.1. The uniformity norms

If the Fourier transform is unable to help us reliably count 4-term progressions, we require an alternative analytic tool that does so. Developing such a tool was one of several profound innovations that were introduced by Gowers [**17**] in his harmonic analysis approach to Szemerédi's theorem.

[8] Various important refinements have been made to this basic Fourier iteration method, notably by Bourgain [**7**] and Sanders [**43**].

[9] It turns out that quadratic identities of this type are the *only* obstructions to uniform sets containing the expected number of solutions to a given system of linear equations, see [**19**].

DEFINITION 1 (Uniformity norms). — *Let $k \geq 2$ be an integer. For any function $f : \mathbb{Z}_N \to \mathbb{C}$, define the U^k norm via the formula*

$$\|f\|_{U^k}^{2^k} = \mathbb{E}_{x \in \mathbb{Z}_N, y \in \mathbb{Z}_N^k} \prod_{\omega \in \{0,1\}^k} \mathcal{C}^{|\omega|} f(x + \omega \cdot y),$$

where for $y = (y_1, \ldots, y_k) \in \mathbb{Z}_N^k$ and $\omega = (\omega_1, \ldots, \omega_k) \in \{0,1\}^k$ we have written $\omega \cdot y = \omega_1 y_1 + \cdots + \omega_k y_k$, as well as $\mathcal{C}^{|\omega|} f = f$ if $|\omega| = \omega_1 + \cdots + \omega_k$ is even and $\overline{f}$ otherwise.

First, it is not hard to show (but neither is it obvious) that this expression defines a norm for all $k \geq 2$. The main technical device for this purpose is the so-called *Gowers-Cauchy-Schwarz inequality*, which is proved using several applications of the ordinary Cauchy-Schwarz inequality. It states that for any family of functions $g_\omega : \mathbb{Z}_N \to \mathbb{C}$, $\omega \in \{0,1\}^k$, we have the bound

$$|\mathbb{E}_{x \in \mathbb{Z}_N, y \in \mathbb{Z}_N^k} \prod_{\omega \in \{0,1\}^k} \mathcal{C}^{|\omega|} g_\omega(x + \omega \cdot y)| \leq \prod_{\omega \in \{0,1\}^k} \|g_\omega\|_{U^k}.$$

With the help of this inequality it is straightforward to verify that the uniformity norms form a nested sequence

$$\|f\|_{U^2} \leq \|f\|_{U^3} \leq \cdots \leq \|f\|_{U^k} \leq \ldots \|f\|_\infty.$$

We also record the fact that the U^k norms can be defined inductively via the formula

$$\|f\|_{U^k}^{2^k} = \mathbb{E}_{h \in \mathbb{Z}_N} \|\Delta_h f\|_{U^{k-1}}^{2^{k-1}},$$

where $\Delta_h f(x) = f(x)\overline{f(x+h)}$ should be viewed as a discrete derivative of f. To see this, consider the case where f is a phase function of the form $f(x) = \exp(2\pi i g(x)/N)$, so that $\Delta_h f(x) = \exp(2\pi i (g(x) - g(x+h))/N)$. Finally, one checks by simple calculation that

$$\|f\|_{U^2}^4 = \mathbb{E}_{x,a,b \in \mathbb{Z}_N} f(x)\overline{f(x+a)f(x+b)}f(x+a+b) = \sum_t |\widehat{f}(t)|^4 = \|\widehat{f}\|_4^4,$$

so that the U^2 norm is equivalent, for bounded functions, to the ℓ^∞ norm of the Fourier transform. [10] This immediately implies that the proof of Roth's theorem based on the randomness-structure dichotomy sketched above can be rephrased replacing the Fourier transform by the U^2 norm. And proofs involving the U^2 norm turn out to be much more amenable to generalization than those involving the Fourier transform itself (although we shall recover some sort of "generalized Fourier transform" later).

[10] Geometrically, we can picture the U^2 norm as averaging the value of f over all 2-dimensional parallelograms with vertices $x, x+a, x+b$ and $x+a+b$, and analogously its kth-order sibling as averaging over k-dimensional parallelepipeds. These parallelepiped structures also play an important role in ergodic theory, where corresponding semi-norms were defined by Host and Kra **[34]**.

2.2. The generalized von Neumann theorem

It was shown by Gowers that the U^k norm controls the count of $(k+1)$-term progressions in the following sense.

PROPOSITION 6 (U^k norm controls $(k+1)$-APs). — *Let $k \geq 2$ be an integer, and let $f_0, f_1, \ldots, f_k : \mathbb{Z}_N \to \mathbb{C}$ be functions satisfying $\|f_j\|_\infty \leq 1$ for $j = 0, 1, \ldots, k$. Then*

$$|\mathbb{E}_{x,d \in \mathbb{Z}_N} f_0(x) f_1(x+d) f_2(x+2d) \ldots f_k(x+kd)| \leq \min_{0 \leq j \leq k} \|f_j\|_{U^k}.$$

Having defined the U^k norms, and more importantly realized their potential for controlling progressions, the proof of this statement is not very difficult (if a little tedious). It involves a number of applications of the Cauchy-Schwarz inequality, combined with a suitable reparametrization of the progression itself. To give the reader a taste, we shall show in the next few lines how to control 3-term progressions by the U^2 norm when all functions f_j are equal to f.

PROOF FOR $k = 2$: We first set $u = x + d$ and write

$$|\mathbb{E}_{x,d} f(x) f(x+d) f(x+2d)|^2 = |\mathbb{E}_{x,d} f(x) f(u) f(2u-x)|^2,$$

which by Cauchy-Schwarz and the boundedness assumption on f is bounded above by

$$\mathbb{E}_u |\mathbb{E}_x f(x) f(2u-x)|^2 = \mathbb{E}_u \mathbb{E}_{x,x'} f(x) f(2u-x) \overline{f(x') f(2u-x')}.$$

Reparameterizing once more by setting $x' = x + a$ and $2u - x' = x + b$ gives

$$\mathbb{E}_{x,a,b} f(x) \overline{f(x+a) f(x+b)} f(x+a+b) = \|f\|_{U^2}^4,$$

which concludes the proof.[11] □

This argument can of course be generalized, not only to longer progressions but to solutions of (almost) any system of linear equations with integer coefficients. This generalization was carried out by Green and Tao **[31]** in their paper on linear equations in the primes, resulting in the following proposition, dubbed the *generalized von Neumann theorem.*[12] In the statement a notion of *complexity* of a system of linear forms appears, of which we shall not give the exact definition since it is rather cumbersome and will not play an important role in the remainder of the article.[13]

---

[11] The attentive reader will have noticed that we have in fact shown that $|\mathbb{E}_{x,d} f(x) f(x+d) f(x+2d)| \leq \|f\|_{U^2}^2$, which is stronger than the statement originally claimed. For three distinct functions f_0, f_1, f_2, however, Proposition 6 is best possible.

[12] Our terminology is non-standard here. Green and Tao gave this name to Proposition 13 below, which we shall refer to as the *relative* generalized von Neumann theorem.

[13] The definition of the complexity of a linear system is based on how many times one needs to apply the Cauchy-Schwarz inequality in the proof of Proposition 7. For a precise statement see Definition 1.5 in **[31]**.

PROPOSITION 7 (Generalized von Neumann theorem). — *For $i = 1, \dots, m$, let $f_i : \mathbb{Z}_N \to \mathbb{C}$ be a family of functions satisfying $\|f_i\|_\infty \leq 1$ for all $i = 1, \dots, m$. Suppose that $(L_i)_{i=1}^m$ is a system of linear forms of complexity $k-1$ in d variables with integer coefficients. Then we have*

$$|\mathbb{E}_{x \in \mathbb{Z}_N^d} \prod_{i=1}^m f_i(L_i(x))| \leq \min_{i=1,\dots,m} \|f_i\|_{U^k}.$$

The reader may wish to bear in mind a $(k+1)$-term arithmetic progression as an example of a system of complexity $k-1$, in which case we recover Proposition 6. What matters for the purpose of counting linear configurations in dense sets (and in the primes) is that the average on the left-hand side can be controlled by *some* U^k norm. (14)

2.3. Inverse and decomposition theorems

This section is not strictly necessary for a first attempt at understanding the Green-Tao and the Tao-Ziegler theorem, and may thus be omitted on first reading. But it does play a crucial role for some of the subtler points we wish to make later on in the context of both linear and polynomial patterns.

Proposition 6 above says that the count of $(k+1)$-term arithmetic progressions is stable under small perturbations in the U^k norm. In particular, if we could write the indicator function 1_A of a set A as $1_A = g + h$, where $\|h\|_{U^k}$ is sufficiently small and g is bounded, then we would have

$$\mathbb{E}_{x,d \in \mathbb{Z}_N} \prod_{j=0}^k 1_A(x + jd) \approx \mathbb{E}_{x,d \in \mathbb{Z}_N} \prod_{j=0}^k g(x + jd).$$

Now if the part $g = 1_A - h$ had some helpful structure, in particular one that might allow us to actually compute the average on the right-hand side, then we would be able to give a good (and hopefully strictly positive) estimate of the quantity on the left-hand side. In particular, we might be able to show that if the density α of $A \subseteq [N]$ is fixed and N is sufficiently large, then

$$\mathbb{E}_{x,d \in \mathbb{Z}_N} \prod_{j=0}^k 1_A(x + jd) \geq c(\alpha)$$

for some constant $c(\alpha)$ only depending on α. It is not hard to prove that this statement is equivalent to Szemerédi's theorem.

A function whose U^k norm is small is said to be *uniform of degree* $k-1$. A natural question therefore arises: if h represents the uniform part of a decomposition of the characteristic function 1_A, what does the *non-uniform* part g look like? Equivalently,

(14) What the smallest such k is turns out to be an interesting question, see [19].

what is the structure of a function whose U^k norm is large? Let us first answer this question for $k = 2$.

THEOREM 8 (Inverse theorem for U^2). — *Let $\delta > 0$. Suppose that $f : \mathbb{Z}_N \to \mathbb{C}$ satisfies $\|f\|_\infty \leq 1$ and $\|f\|_{U^2} \geq \delta$. Then there exists a linear phase function $\phi : \mathbb{Z}_N \to \mathbb{C}$ such that*

$$|\mathbb{E}_x f(x)\phi(x)| \geq \delta^2.$$

This function ϕ is of the form $\phi(x) = \exp(2\pi ixt/N)$ for some $t \in \mathbb{Z}_N$.

Proof. — We have already seen that $\|f\|_{U^2} = \|\widehat{f}\|_4$, which can be bounded above by $\|\widehat{f}\|_\infty^2 \|f\|_\infty^2$. Thus for a bounded function, the condition $\|f\|_{U^2} \geq \delta$ implies that $\|\widehat{f}\|_\infty \geq \delta^2$. Therefore, by definition of the Fourier transform, there exists a linear phase function $\phi(x) = \exp(2\pi ixt/N)$ for some $t \in \mathbb{Z}_N$ such that $|\mathbb{E}_x f(x)\phi(x)| \geq \delta^2$. □

In other words, if f has large U^2 norm, then it correlates with a linear phase. For higher values of k, we can a priori only make a rather trivial statement, for which we need the following definition.

DEFINITION 2 (Dual function). — *For any $f : \mathbb{Z}_N \to \mathbb{C}$, define the* dual function $\mathcal{D}f$ *of order k by the formula*

$$\mathcal{D}f(x) = \mathbb{E}_{y \in \mathbb{Z}_N^k} \prod_{\omega \in \{0,1\}^k, \omega \neq 0^k} \mathcal{C}^{|\omega|} f(x + \omega \cdot y).$$

It is immediate from this definition and that of the U^k norm that if $\|f\|_{U^k}$ is large, then so is the inner product $\langle f, \mathcal{D}f \rangle := \mathbb{E}_x f(x)\mathcal{D}f(x)$, where $\mathcal{D}f$ is the kth order dual function of f. Therefore, a function whose U^k norm is large correlates with its own kth order dual function. This is not a very useful statement, however, since we have little tangible information about the structure of the dual function of an arbitrary function f. However, we shall encounter these *soft* obstructions to uniformity again later on.

In fact, a much stronger and more explicit statement is true, which is called the *inverse theorem* for the U^k norm. For simplicity, we shall first state the case $k = 3$. We shall informally call a function a *generalized quadratic phase* if it "behaves quadratically" on an "approximate subgroup" of $\mathbb{Z}_N$, and give a rigorous definition in the more general case below.[15]

[15] For example, ϕ can be expressed as $1_B(x)\exp(2\pi iq(x)/N)$, where $B(K, \rho) = \{x \in \mathbb{Z}_N : \|xt/N\| \leq \rho$ for all $t \in K\}$ is a Bohr set whose width ρ and dimension $|K|$ are bounded in terms of δ, and $q : B \to \mathbb{Z}_N$ is a function that satisfies the quadratic identity $q(x) - q(x+a) - q(x+b) - q(x+c) + q(x+a+b) + q(x+a+c) + q(x+b+c) - q(x+a+b+c) = 0$, whenever all these terms are defined.

THEOREM 9 (Inverse theorem for U^3). — *Let $\delta > 0$. Suppose that $f : \mathbb{Z}_N \to \mathbb{C}$ satisfies $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq \delta$. Then there exists a generalized quadratic phase ϕ such that*

$$|\mathbb{E}_x f(x)\phi(x)| \geq c(\delta),$$

for some constant $c(\delta)$ that depends on δ.

Note that it is easy to see by direct calculation that when $f(x) = \exp(2\pi i q(x)/N)$ for some quadratic polynomial q, then $\|f\|_{U^3} = 1$, so the inverse theorem gives a weak converse to this statement: if the U^3 norm is bounded away from 0, then f correlates with a generalized quadratic phase.

The proof of this inverse theorem is ingenious and combines a number of heavyweight tools from additive combinatorics. It was largely contained in Gowers's proof of Szemerédi's theorem [**17**] culminating in a slightly weaker statement than Theorem 9; the strengthened version stated above is due to Green and Tao [**24**] who added one additional ingredient to the proof. The very rough idea of the proof of Theorem 9 is as follows: if $\|f\|_{U^3}$ is large, then by the inductive definition of the U^3 norm we know that for many values of h, $\Delta_h f$ has large U^2 norm. This means that, for many values of h, the derivative $\Delta_h f$ of f correlates with a linear phase function. The main difficulty lies in collecting these weak linear structures together in such a way that one is able to integrate the statement and conclude that f itself correlates with a quadratic structure. This crucial step is achieved with the help of a purely combinatorial result known as *Freiman's theorem* [**13, 42**]. [(16)] The bound in Theorem 9, i.e. the dependence of $c(\delta)$ on δ, which is currently quasi-polynomial in nature as a consequence of a recent result of Sanders [**44**], has been shown to be equivalent to the bound in Freiman's theorem, and improving it remains an important focus of research in the area.

It is only very recently that Green, Tao and Ziegler [**30**], in a major breakthrough, were able to prove an inverse theorem for higher values of k. Both statement and proof draw inspiration from ergodic theory, and in particular the deep structure theory of characteristic factors induced by the analogue of the U^k norms in the ergodic context, developed in a seminal article by Host and Kra [**34**].[(17)] We shall need the following definition, which made its first appearance in [**5**].

DEFINITION 3 (Nilsequence). — *Let G be a k-step nilpotent group, i.e. a connected, simply connected Lie group with central series $G = G_1 \supseteq \cdots \supseteq G_{k+1} = \{1\}$.*

[(16)] Freiman's theorem states that a set $B \subseteq \mathbb{Z}$ whose sumset $B + B = \{b + b' : b, b' \in B\}$ is small, i.e. $|B + B| \leq K|B|$ for some constant K, is efficiently contained in a somewhat rigid algebraic substructure (a generalized arithmetic progression) whose size and dimension depend only on K.

[(17)] For related work on the analysis of multiple ergodic averages see also Conze and Lesigne [**10**], Furstenberg and Weiss [**15**] and Ziegler [**55**].

Let $\Gamma \subseteq G$ be a discrete co-compact subgroup. Then the quotient G/Γ is a k-step nilmanifold, *and a sequence of the form $(F(g^n x))_{n\in\mathbb{N}}$ with $g \in G, x \in G/\Gamma$ and continuous $F : G/\Gamma \to \mathbb{R}$ is called a* k-step nilsequence.

Note that $G/\Gamma = \mathbb{R}/\mathbb{Z}$ is an example of a 1-step nilmanifold, and that any function of the form $n \mapsto F(x + n\alpha)$ for $\alpha \in \mathbb{R}$ and continuous F is a 1-step nilsequence. In particular, the linear characters $n \mapsto \exp(2\pi i n\alpha)$ are examples of basic 1-step nilsequences. For higher values of k, including $k = 2$, a k-step nilsequence displays an undeniably non-commutative character.(18) A simple example of a 2-step nilmanifold is given by the quotient of the matrix group G and its discrete subgroup Γ defined by

$$G = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}, \quad \Gamma = \begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix},$$

which can be identified topologically with the 2-torus. Now let $g \in G$ be given by

$$g = \begin{pmatrix} 1 & m & \beta \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix},$$

where $m \in \mathbb{Z}$ and $\alpha, \beta \in \mathbb{R}$. Then a shift of $(x, y) \in \mathbb{T}^2$ by g is given by $(x, y) \mapsto (x + \alpha, y + \beta + mx)$, and the nilsequence $F(g^n(x, y))$ for $n \in \mathbb{N}$ is given by $F(x + n\alpha, y + n\beta + \frac{1}{2}mn(n + 1)\alpha)$, clearly exhibiting the claimed quadratic behaviour. Observe in particular that a quadratic phase function such as $n \mapsto \exp(\pi i n(n + 1)\alpha)$ belongs to the family of basic 2-step nilsequences.(19)

We are now able to state the inverse theorem for the U^k norm, which asserts that a function whose U^k norm is large correlates with a $(k - 1)$-step nilsequence.

THEOREM 10 (Inverse theorem for U^k). — *Let $0 < \delta \leq 1$ and $k \geq 1$ be an integer. Then there exists a finite collection $\mathcal{M}_{k,\delta}$ of k-step nilmanifolds G/Γ, each equipped with some smooth Riemannian metric $d_{G/\Gamma}$ as well as constants $C(k, \delta)$, $c(k, \delta) > 0$ with the following property. Whenever $N \geq 1$ and $f : [N] \to \mathbb{C}$ with $\|f\|_\infty \leq 1$ is a function such that $\|f\|_{U^{k+1}[N]} \geq \delta$, then there exist a nilmanifold $G/\Gamma \in \mathcal{M}_{k,\delta}$, some $g \in G$ and a function $F : G/\Gamma \to \mathbb{C}$ with Lipschitz constant at most $C(k, \delta)$ with respect to the metric $d_{G/\Gamma}$, such that*

$$|\mathbb{E}_{n\in[N]} f(n)\overline{F(g^n x)}| \geq c(k, \delta).$$

(18) This partly explains our difficulty in defining a "generalized quadratic phase" above.
(19) Further examples of 2-step nilsequences, such as the Heisenberg nilflow, are given in full detail in Section 12 of **[24]**.

The passage from weak linearity of the derivative to the quadratic nature of the function itself, which we briefly alluded to earlier, amounts in the more general case to a certain "cohomological" task, in which one has to show that a certain "cocyle" is essentially a "coboundary". This work is very recent and the current proof is almost certainly not the right one, so we shall say no more about it here but instead refer the interested reader to the carefully crafted announcement [**29**].

Given an inverse theorem, it is possible to deduce a *decomposition theorem* that allows us to write a bounded function f as a sum $f = g + h$, where h is uniform of degree $k-1$ and g has polynomial structure of degree $k-1$. Note that in the case $k = 2$, such a decomposition simply consists of a partition of the usual Fourier expansion into small and large coefficients, which is straightforward to write down since the linear characters form an orthonormal basis. For $k > 2$, no such canonical basis exists, and a number of other methods have been employed, such as an energy increment strategy on factors [**27**], or the so-called boosting method from theoretical computer science [**51**]. The connection between inverse theorems and decomposition theorems has recently been formalized by Gowers [**18**] (see also the first application in [**20**]), who showed that the Hahn-Banach theorem from functional analysis can be used to deduce decomposition theorems from inverse theorems for a large class of norms, including the uniformity norms. In the same paper, he showed that the Hahn-Banach theorem gives an alternative proof of the transference principle, which we shall discuss in Section 3.2. Moreover, we shall see in Section 3.4 that it is precisely the existence of a higher-order inverse and corresponding strong decomposition theorem which provides asymptotics for linear configurations in the primes.

2.4. Polynomial generalizations

It is natural to ask for polynomial generalizations of Szemerédi-type theorems: is it true that any sufficiently dense subset of the first N integers contains a given polynomial configuration? For example, is it true that any sufficiently dense subset of $[N]$ contains a configuration of the form $x, x+n^2$ for $x, n \in \mathbb{N}$? The latter question was answered in the affirmative by Sárközy [**45**], whose theorem (with the best known bound due to Pintz, Steiger and Szemerédi [**38**]) we state below.[(20)]

THEOREM 11 (Sárközy's theorem). — *Suppose that $A \subseteq [N]$ is a subset of density α which contains no two distinct elements whose difference is a perfect square. Then*

$$\alpha \ll (\log N)^{-\frac{1}{4}\log\log\log\log N}.$$

[(20)] An analogous statement is easily seen to be false for a difference of the form n^2+1. Since there are no squares congruent to 2 mod 3, we can take the set of multiples of 3 as a (very dense) counterexample.

By the prime number theorem, the exceptionally strong quantitative information in Theorem 11 immediately implies the existence of square differences in any positive density subset of the primes, for density reasons alone. The proof of Sárközy's theorem proceeds via classical Fourier analysis on $\mathbb{Z}$, following the circle method approach of Hardy and Littlewood, and can be extended to configurations of the form $x, x + P(n)$, where P is an "intersective" polynomial.[21] Unfortunately, no such quantitative results are known for more general polynomial systems such as $x, x + n^2, x + 2n^2$ or $x, x + n, x + n^2$. What we do have is a result of Bergelson and Leibman [**6**], proved entirely within the realm of ergodic theory, which states in a purely qualitative fashion that any subset of the integers of positive upper density contains the translate of a simultaneous image of a system of polynomials with zero constant coefficient.[22]

THEOREM 12 (Bergelson-Leibman theorem). — *Let $A \subseteq \mathbb{Z}$ be a set of positive upper density, and let $P_1, \dots, P_k$ be polynomials with rational coefficients satisfying $P_i(0) = 0$ for $i = 1, \dots, k$. Then there exist $x, n \in \mathbb{Z}$ such that $x + P_i(n) \in A$ for $i = 1, \dots, k$.*

Theorem 12 follows from an abstract result about the convergence of multiple ergodic averages in a measure-preserving dynamical system, via Furstenberg's correspondence principle [**14**]. The latter gives an explicit way of constructing a dynamical system in such a way that the recurrence results obtained therein can be transferred to corresponding statements in the integers. A combinatorial proof of the Bergelson-Leibman theorem is yet to be found. We shall discuss this theorem further in the context of the primes in Section 4.1.

3. LINEAR CONFIGURATIONS IN THE PRIMES

Having explored what is known about arithmetic structure in dense subsets of the integers in some detail in the preceding section, we now turn our attention to the primes. It is convenient to weight the indicator function of the primes so that their density is roughly constant throughout an interval. The tool traditionally used for this purpose is the von Mangoldt function, denoted by Λ and defined by

$$\Lambda(n) = \begin{cases} \log(p) & \text{if } n = p^m \text{ for some } m \in \mathbb{N}, \\ 0 & \text{otherwise.} \end{cases}$$

[21] An intersective polynomial has a root modulo q for every $q \in \mathbb{N}$.
[22] The most general version of Theorem 12 is in fact multidimensional and holds for any system of intersective polynomials.

The prime number theorem tells us that $\mathbb{E}_{n\in[N]}\Lambda(n) = 1 + o(1)$ (and is in fact equivalent to this statement). The first obstacle that needs to be overcome when trying to prove Theorem 1 is that the primes are not equidistributed amongst all residue classes. As a trivial example, note that very few primes (indeed) are even. However, it is easy to check that the odd primes no longer show any bias towards either residue class modulo 2 after applying the map $x \mapsto (x-1)/2$. In a similar manner, one can remove the bias with respect to two of the residue classes modulo 3, and similarly for all small primes p. This is called the *W-trick* by Green and Tao, and leads to the definition of the modified von Mangoldt function $\widetilde{\Lambda} : \mathbb{N} \to \mathbb{R}^+$ via the formula[23]

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

where ϕ is the Euler totient function and $W = \prod_{p\leq w} p$ is the product of all primes not exceeding a threshold w.[24] Clearly the normalization is chosen so that $\mathbb{E}_{n\in[N]}\widetilde{\Lambda}(n) = 1 + o(1)$, by the Siegel-Walfisz theorem on primes in arithmetic progressions.[25]

A second and much more serious obstacle is that the function $\widetilde{\Lambda}$ is unbounded, and therefore none of the results from the preceding section are directly applicable. Overcoming this latter impediment is where the main achievement of Green and Tao lies: they developed a way of "transferring" known results and techniques for dense subsets to a non-dense setting. The main idea is to embed the primes into a well-behaved set with respect to which they *are* dense. A natural candidate for such a set would be the set of almost primes. However, it turns out to be more convenient to use not a set but a majorizing *pseudorandom measure.*[26]

DEFINITION 4 (Pseudorandom measure). — *We say that $\nu : \mathbb{Z}_N \to \mathbb{R}^+$ is a k-pseudorandom measure that majorizes the primes if it satisfies the following properties.*[27]

- *ν is normalized:*

$$\mathbb{E}_{x\in\mathbb{Z}_N}\nu(x) = 1 + o(1);$$

[23] In order to establish the existence of arbitrarily long progressions in a positive-density subset A of the primes, one needs to replace the residue class $n \equiv 1 \bmod W$ by $n \equiv b \bmod W$, where $(b, W) = 1$ and the residue class b is chosen according to the pigeonhole principle so as to coincide with a positive fraction of A.

[24] The parameter w can be thought of as roughly $\log\log N$, although Green and Tao remark that it is possible to set it equal to a very large constant.

[25] Note that $\widetilde{\Lambda}(n)$ differs from $\frac{\phi(W)}{W}\Lambda(Wn+1)$ only on the negligible set of prime powers.

[26] The pseudorandom measure should more accurately be called a weight function, or probability density. It is also sometimes referred to as an *enveloping sieve*, see for example **[39]**.

[27] The exact values of the constants depending on k which appear in the definition are rather unimportant, and in fact it is quite probable that these conditions can be somewhat relaxed in general.

- ν *majorizes* $\tilde{\Lambda}$*: for all* $\epsilon_k N \leq n \leq 2\epsilon_k N$*, where* $\epsilon_k = (2^k(k+4)!)^{-1}$*, we have*

$$\nu(n) \geq k^{-1}2^{-k-5}\tilde{\Lambda}(n);$$

- ν *satisfies the* linear forms condition*: for any* $t \leq 3k-4$*,* $m \leq k \cdot 2^{k-1}$*,*

$$\mathbb{E}_{x \in \mathbb{Z}_N^t} \nu(\psi_1(x)) \dots \nu(\psi_m(x)) = 1 + o_k(1),$$

where the rational coefficients of the affine linear forms $\psi_i : \mathbb{Z}_N^t \to \mathbb{Z}_N$ *have denominator at most* k*, and the* ψ_i *satisfy a mild non-degeneracy condition;*

- ν *satisfies the* correlation condition*: for any* $h_1, \dots, h_m \in \mathbb{Z}_N$*,* $m \leq 2^{k-1}$*,*

$$\mathbb{E}_{x \in \mathbb{Z}_N} \nu(x+h_1) \dots \nu(x+h_m) \leq \sum_{1 \leq i < j \leq m} \tau(h_i - h_j),$$

where $\tau : \mathbb{Z}_N \to \mathbb{R}^+$ *satisfies* $\mathbb{E}_{x \in \mathbb{Z}_N} \tau^q(x) \leq C(m,q)$ *for all* $1 \leq q < \infty$*.*

In Section 3.1, we extend Proposition 6 to functions that are not necessarily bounded by a constant, but rather by a pseudorandom measure. Subsequently we give an outline of the key step of the proof of the Green-Tao theorem, namely the transference principle, which allows us to apply Szemerédi's theorem to functions that are majorized by a pseudorandom measure. It is only in Section 3.3 that we concern ourselves with the concrete existence of a pseudorandom majorant for the primes. Finally, in Section 3.4, we briefly discuss more general linear configurations in the primes, and sketch how recent developments in higher-order Fourier analysis yield asymptotics for such configurations.

While we do not give complete proofs of the results stated, we intend to point out where the individual properties of the measure ν come into play. For reasons of brevity we shall be rather cavalier about the use of $o(1)$ and $O(1)$ notation. In practice, however, one needs to check carefully that the quantities this notation hides do not grow too large upon multiplication with each other.

3.1. The generalized von Neumann theorem relative to a pseudorandom measure

Our first task is to extend Proposition 6 to the case where each function f_i is bounded by a pseudorandom measure ν. The argument is very similar to the one we gave earlier for $k = 2$, except that one now needs the linear forms condition to take care of the various averages of ν that arise from multiple applications of the Cauchy-Schwarz inequality.

PROPOSITION 13 (Relative generalized von Neumann). — *Suppose that* ν *is a* $(k+1)$*-pseudorandom measure, meaning in particular that it satisfies the corresponding linear forms condition. Suppose that* $f_0, f_1, \dots, f_k : \mathbb{Z}_N \to \mathbb{C}$ *are functions*

satisfying $|f_j(x)| \leq \nu(x)$ *for all* $j = 0, 1, \ldots, k$. *Then there exists a constant* $C(k)$ *such that*

$$|\mathbb{E}_{x,d\in\mathbb{Z}_N} \prod_{j=0}^{k} f_j(x+jd)| \leq C(k) \min_{0\leq j\leq k} \|f_j\|_{U^k}.$$

We shall only give the proof in the case where $k = 2$ and the minimum on the right-hand side is attained for $j = 0$, which contains all the ingredients of the general case but is less notation-intensive.

PROOF FOR $k = 3$: Reparametrizing the system

$$\mathbb{E}_{x,d} f_0(x) f_1(x+d) f_2(x+2d) = \mathbb{E}_{y_1,y_2} f_0(y_1+y_2) f_1(y_2/2) f_2(-y_1)$$

to ensure that the variables y_1, y_2 are separated, the triangle inequality implies that

$$|\mathbb{E}_{x,d} f_0(x) f_1(x+d) f_2(x+2d)| \leq \mathbb{E}_{y_1} \nu(-y_1) |\mathbb{E}_{y_2} f_0(y_1+y_2) f_1(y_2/2)|.$$

A first application of the Cauchy-Schwarz inequality, using the fact that $\mathbb{E}\nu = 1+o(1)$, gives an upper bound of

$$(1+o(1)) \left(\mathbb{E}_{y_1} \nu(-y_1) |\mathbb{E}_{y_2} f_0(y_1+y_2) f_1(y_2/2)|^2\right)^{1/2},$$

and expanding out the square yields

$$(1+o(1)) \left(\mathbb{E}_{y_2,y_2'} f_1(y_2/2) f_1(y_2'/2) \mathbb{E}_{y_1} \nu(-y_1) f_0(y_1+y_2) f_0(y_1+y_2')\right)^{1/2}.$$

A second application of Cauchy-Schwarz gives an upper bound of $(1+o(1))$ times

$$\left(\mathbb{E}_{y_1,y_1',y_2,y_2'} \nu(-y_1)\nu(-y_1')\nu(y_2/2)\nu(y_2'/2) f_0(y_1+y_2) f_0(y_1+y_2') f_0(y_1'+y_2) f_0(y_1'+y_2')\right)^{1/4},$$

which can be rewritten as

$$(1+o(1)) \left(\mathbb{E}_{x,h_1,h_2} f_0(x) f_0(x+h_1) f_0(x+h_2) f_0(x+h_1+h_2) W(x,h_1,h_2)\right)^{1/4},$$

where $W(x,h_1,h_2) = \mathbb{E}_y \nu(-y)\nu(-y-h_1)\nu((x-y)/2)\nu((x-y+h_2)/2)$. It remains to show that at not unreasonable cost, we can replace the function W by the constant function 1. In other words, we want to show that

$$\mathbb{E}_{x,h_1,h_2} f_0(x) f_0(x+h_1) f_0(x+h_2) f_0(x+h_1+h_2)(W(x,h_1,h_2)-1)$$

is small. But by Cauchy-Schwarz and the boundedness assumption on f_0, we have that the latter expression is bounded by

$$(\mathbb{E}_{x,h_1,h_2} \nu(x)\nu(x+h_1)\nu(x+h_2)\nu(x+h_1+h_2))^{1/2}$$
$$(\mathbb{E}_{x,h_1,h_2} \nu(x)\nu(x+h_1)\nu(x+h_2)\nu(x+h_1+h_2)(W(x,h_1,h_2)-1)^2)^{1/2}.$$

Now $(W(x,h_1,h_2)-1)^2 = W(x,h_1,h_2)^2 - 2W(x,h_1,h_2) + 1$, so each of the four expectations involved is of a form to which the linear forms condition applies. In particular, each of the four expectations equals $1+o(1)$, giving a final bound of $o(1)$ as desired. □

It can easily be checked that if ν is a $(k+1)$-pseudorandom measure, then so is $\frac{1}{2}(1+\nu)$, and hence Proposition 13 also applies to functions bounded by $1+\nu$ (up to constant factors). This is a fact which we shall need in the next section, where we shall decompose $f = \widetilde{\Lambda}$, which is majorized by ν, as $f = g + h$, where g is bounded by 1 and h has small U^k norm. In this case Proposition 13 says that contributions from h are negligible to any average counting arithmetic progressions in f.

3.2. The transference principle

We now turn to the heart of the proof of the Green-Tao theorem, namely the transference principle. To reiterate, the main idea is that if a theorem (such as Szemerédi's theorem or the inverse theorem) is true for bounded functions, then it should be true for functions that are majorized by a pseudorandom measure. Indeed, a first step in this direction was taken by Kohayakawa, Łuczak and Rödl **[35]**, who proved the existence of 3-term progressions in dense subsets of a (truly) random set. Since then, a number of results from additive combinatorics have been transferred to the context of dense subsets of random sets (see for example **[32]**). With the benefit of hindsight therefore, establishing a similar result for dense subsets of *pseudo*random sets seems rather natural, even though at the time it represented a clear conceptual leap that marks the main advance in the work of Green and Tao (and the earlier paper by Green on Roth's theorem in the primes **[22]**). The transference principle was not stated explicitly in **[25]** but was for the first time isolated in **[50]**. Both papers proceed by decomposing $\widetilde{\Lambda}$ into an almost periodic and a weakly mixing part, reminiscent of Furstenberg's ergodic structure theorem **[14]**.[(28)] Here we shall follow the more recent analytic approach due to Gowers **[18]**, which was also simultaneously and independently discovered by Reingold, Trevisan, Tulsiani and Vadhan **[40]** in the context of theoretical computer science. This rather general and powerful method is based on the *Hahn-Banach theorem*, or the *duality of linear programming* as it is known to computer scientists.[(29)]

THEOREM 14 (Transference principle). — *For every $\delta, \eta > 0$, there exists $\epsilon > 0$ with the following property. Let ν be a $(k+1)$-pseudorandom measure satisfying $\|\nu - 1\|_{U^k} \leq \epsilon$, and suppose that $f : \mathbb{Z}_N \to \mathbb{R}$ is a function satisfying $0 \leq f \leq \nu$. Then there exists a function g such that $0 \leq g \leq (1-\delta)^{-1}$ and $\|f - g\|_{U^k} \leq \eta$.*

(28) Specifically, one writes $f = \mathbb{E}(f|\mathcal{B}) + (f - \mathbb{E}(f|\mathcal{B}))$ for a σ-algebra $\mathcal{B}$ constructed out of level sets of dual functions using an energy increment strategy. For further details see Host's exposition **[33]**.
(29) For example, this method gives strongly quantitative quadratic decomposition theorems, which resemble an inversion formula for a type of "quadratic Fourier transform" **[20]**. It has also found important, more combinatorial applications in recent work of Conlon and Gowers **[8]**.

Let us first sketch how this result, together with the relative generalized von Neumann theorem and Szemerédi's theorem, implies the Green-Tao theorem.

SKETCH PROOF OF THEOREM 1: We wish to estimate from below the average

$$\mathbb{E}_{x,d} f(x) f(x+d) \dots f(x+kd),$$

where $f(n) = \widetilde{\Lambda}(n)$, the modified von Mangoldt function defined above, is bounded above by (a constant time) ν. Theorem 14 gives us a function $0 \leq g \leq 1+\delta$ such that $\|f-g\|_{U^k} \leq \eta$, and so by Proposition 13 the above average is, up to small error terms, equal to (a constant multiple of)

$$\mathbb{E}_{x,d} g(x) g(x+d) \dots g(x+kd).$$

Now g is an (almost) bounded function, so Theorem 3 applies, implying that the latter average is bounded below by a constant depending only on the density of g. But the density of g is strictly positive since $\mathbb{E}g = \mathbb{E}f - \mathbb{E}(f-g)$, and $|\mathbb{E}(f-g)| \leq \|f-g\|_{U^k} \leq \eta$, concluding the proof for an appropriate choice of the parameters δ and η. □

Before we delve into a sketch proof of Theorem 14, we first verify that when ν is a $(k+1)$-pseudorandom measure, the U^k norm cannot distinguish it from the constant function 1, and so the additional hypothesis in Theorem 14 is always satisfied.

LEMMA 15. — *If ν is a $(k+1)$-pseudorandom measure, then $\|\nu - 1\|_{U^k} = o(1)$.*

Proof. — We simply rearrange the definition of the Gowers norm and write

$$\|\nu-1\|_{U^k}^{2^k} = \mathbb{E}_{x,h_1,\dots,h_k} \prod_{\omega \in \{0,1\}^k} (\nu(x+\omega\cdot h)-1) = \mathbb{E}_{x,h_1,\dots,h_k} \sum_{S \subseteq \{0,1\}^k} (-1)^{|S|} \prod_{\omega \in S} \nu(x+\omega\cdot h).$$

The latter expression equals

$$\sum_{S \subseteq \{0,1\}^k} (-1)^{|S|} \mathbb{E}_{h_1,\dots,h_k} (\mathbb{E}_x \prod_{\omega \in S} \nu(x+\omega\cdot h)),$$

where the expectation in x can be written as $\mathbb{E}_x \nu(\psi_1(x)) \cdots \nu(\psi_m(x))$ for some linear forms ψ_i. For each $h_1, \dots, h_k$, this expectation evaluates to $1 + o(1)$ by the linear forms condition, and the result follows on summing. □

Let us recall the statement of the classical Hahn-Banach theorem in the context of finite dimensional vector spaces over $\mathbb{R}$, which will be used as a black box in the proof of the transference principle. Consequently, the function g that appears in its statement is not constructed explicitly, but rather its existence is the result of an argument by contradiction.

THEOREM 16 (Hahn-Banach theorem). — *Let $X = (\mathbb{R}^N, \|\cdot\|)$ be a normed space and let $x \in X$ be a vector with $\|x\| \geq 1$. Then there is a vector $z \in \mathbb{R}^N$ such that $\langle x, z\rangle > 1$ and such that $|\langle y, z\rangle| \leq 1$ whenever $\|y\| \leq 1$.*

Defining the dual norm of $\|\cdot\|$ by $\|x\|^* = \sup\{|\langle x, y\rangle| : \|y\| \leq 1\}$, we can reformulate the statement of the theorem as follows: if $x \in X$ satisfies $\|x\| \geq 1$, then there exists z such that $\langle x, z\rangle > 1$ and $\|z\|^* \leq 1$. We shall actually need a slightly modified version. For $i = 1, 2$, let $\|\cdot\|_i$ be a norm defined on a subspace V_i of $\mathbb{R}^N$, and assume that $V_1 + V_2 = \mathbb{R}^N$. For $\alpha_1, \alpha_2 \in \mathbb{R}^+$, define the norm

$$\|x\| = \inf\{\alpha_1\|x_1\|_1 + \alpha_2\|x_2\|_2 : x = x_1 + x_2\},$$

whose dual is easily seen to be[30]

$$\|z\|^* = \max\{\alpha_1^{-1}\|z\|_1^*, \alpha_2^{-1}\|z\|_2^*\}.$$

With this notation we immediately deduce the following corollary.

COROLLARY 17. — *Let V_1, V_2 be as above, and let $\alpha_1, \alpha_2 \in \mathbb{R}^+$. Suppose that it is* not *possible to write the vector $x \in \mathbb{R}^N$ as $x_1 + x_2$ in such a way that $x_i \in V_i$ for each i, and $\alpha_1\|x_1\|_1 + \alpha_2\|x_2\|_2 \leq 1$. Then there exists a vector $z \in \mathbb{R}^N$ such that $\langle x, z\rangle > 1$ and such that $\|z\|_i^* \leq \alpha_i$ for $i = 1, 2$.*

We are now in a position to sketch the deduction of the transference principle from Corollary 17. Note that the statement of the transference principle can indeed be viewed as a kind of decomposition theorem: we want to write the function f, which is bounded by the pseudorandom measure ν, as a sum $g + h$, where $h = f - g$ is small in U^k and g is (almost) bounded.

SKETCH PROOF OF THEOREM 14: Suppose for the sake of contradiction that it is not possible to write $f = g + h$ with $\|g\|_\infty \leq (1-\delta)^{-1}$ and $\|h\|_{U^k} \leq \eta$. Then the hypotheses of Corollary 17 are satisfied with $V_1 = (\mathbb{R}^N, \|\cdot\|_\infty)$, $V_2 = (\mathbb{R}^N, \|\cdot\|_{U^k})$, $\alpha_1 = 1 - \delta$ and $\alpha_2 = \eta^{-1}$. Therefore there exists $\phi \in \mathbb{R}^N$ satisfying $\|\phi\|_\infty^* = \|\phi\|_1 \leq 1 - \delta$ and $\|\phi\|_{U^k}^* \leq \eta^{-1}$ but $\langle f, \phi\rangle > 1$.

Now suppose for a moment (and this is a significant oversimplification) that ϕ only took positive values. Then we would have

$$1 < \langle f, \phi\rangle \leq \langle \nu, \phi\rangle = \langle 1, \phi\rangle + \langle \nu - 1, \phi\rangle.$$

Now since ϕ has U^k dual norm bounded by η^{-1} and $\|1 - \nu\|_{U^k} \leq \epsilon$, this is, up to an error of $\epsilon\eta^{-1}$, equal to

$$\langle 1, \phi\rangle = \|\phi\|_1 \leq 1 - \delta,$$

giving the desired contradiction if ϵ is chosen sufficiently small. Of course, in reality things are not quite as easy as that. Denoting by $\phi_+ = \max(0, \phi)$ the positive part of ϕ, we have

$$1 < \langle f, \phi\rangle \leq \langle f, \phi_+\rangle \leq \langle \nu, \phi_+\rangle.$$

[30] Here $\|z\|_i^* = \sup\{|\langle z, x\rangle| : x \in V_i, \|x\|_i \leq 1\}$ is a semi-norm.

In order to proceed as above, that is, to say that the latter inner product is approximately equal to $\langle 1, \phi_+ \rangle$, we would need to know that ϕ_+ is bounded in the U^k dual norm. But a priori we know nothing about where ϕ is positive or negative, just that its U^k dual norm is bounded. One can, however, use the Weierstrass approximation theorem to obtain a polynomial P such that $\phi_+ \approx P(\phi)$. So it suffices to show that $\|P(\phi)\|^*_{U^k}$ is bounded whenever $\|\phi\|^*_{U^k}$ is. For, in that case, we could continue the above string of inequalities, up to an error depending on the quality of the polynomial approximation, via

$$\langle \nu, P(\phi) \rangle \approx \langle 1, P(\phi) \rangle \approx \langle 1, \phi_+ \rangle \leq 1 - \delta.$$

In order to show that $\|P(\phi)\|^*_{U^k}$ is bounded under the assumption that $\|\phi\|^*_{U^k}$ is, it suffices by linearity to consider the case where P is a monomial $x_1 x_2 \cdots x_m$. Since $\|\phi\|^*_{U^k}$ is small, we know that ϕ is essentially a kth order dual function (although of course this needs to be made precise). Thus, in order to complete the proof of the transference principle it is enough to prove a lemma to the effect that products of dual functions have bounded U^k dual norm. Since the function h that gave rise to the bound $\|\phi\|^*_{U^k} \leq \eta^{-1}$ is only bounded by $1 + \nu$ instead of 1, we need to consider dual functions of functions that are bounded by a pseudorandom measure.[31]

LEMMA 18. — *Suppose that the functions $f_1, \ldots, f_m : \mathbb{Z}_N \to \mathbb{C}$ satisfy $|f_j(x)| \leq \nu(x)$ for each $j = 1, \ldots, m$, where ν is a $(k+1)$-pseudorandom measure. Then there exists a constant $C(m)$ such that*

$$\|\mathcal{D}f_1 \cdots \mathcal{D}f_m\|^*_{U^k} \leq C(m),$$

where $\mathcal{D}$ refers to the kth order dual operator introduced in Definition 2.

Proof. — We need to show that $\langle g, \mathcal{D}f_1 \cdots \mathcal{D}f_m \rangle$ is bounded by a constant depending on m whenever $g : \mathbb{Z}_N \to \mathbb{C}$ is such that $\|g\|_{U^k} \leq 1$. The inner product can be written as

$$\mathbb{E}_{x \in \mathbb{Z}_N} g(x) \prod_{j=1}^{m} \mathbb{E}_{h^{(j)} \in \mathbb{Z}_N^k} \prod_{\omega \in \{0,1\}^k \setminus 0^k} f_j(x + \omega \cdot h^{(j)}),$$

and with a change of variable $h^{(j)} = h + H^{(j)}$ becomes

$$\mathbb{E}_{x \in \mathbb{Z}_N} g(x) \prod_{j=1}^{m} \mathbb{E}_{h, H^{(j)} \in \mathbb{Z}_N^k} \prod_{\omega \in \{0,1\}^k \setminus 0^k} f_j(x + \omega \cdot h + \omega \cdot H^{(j)}).$$

(31) In order to make this step precise, one defines another norm $\|f\|_{BAC} = \max\{|\langle f, \mathcal{D}g \rangle| : 0 \leq g \leq 1 + \nu\}$, whose dual $\|f\|^*_{BAC} = \inf\{\sum_i |\lambda_i| : f = \sum_i \lambda_i \mathcal{D}f_i, 0 \leq f_i \leq 1 + \nu\}$ measures the extent to which a function f can be written as a small linear combination of dual functions. A sufficient condition for $\|h\|_{U^k}$ to be at most η is then that $\|h\|_{BAC} \leq \eta$, and hence the actual condition on ϕ we end up with is $\|\phi\|^*_{BAC} \leq \eta^{-1}$. This, together with the remark following the proof of Proposition 13, explains the hypotheses in Lemma 18.

Rearranging, we obtain

$$\mathbb{E}_{H=(H^{(1)},\dots,H^{(m)})\in(\mathbb{Z}_N^k)^m}\mathbb{E}_{x\in\mathbb{Z}_N}g(x)\prod_{\omega\in\{0,1\}^k\setminus 0^k}\mathbb{E}_{h\in\mathbb{Z}_N^k}\prod_{j=1}^{m}f_j(x+\omega\cdot h+\omega\cdot H^{(j)}),$$

which, with $g_{\omega,H}(x)=\prod_{j=1}^m f_j(x+\omega\cdot H^{(j)})$, equals a Gowers inner product of the form

$$\mathbb{E}_{H\in(\mathbb{Z}_N^k)^m}\mathbb{E}_{x\in\mathbb{Z}_N}g(x)\mathbb{E}_{h\in\mathbb{Z}_N^k}\prod_{\omega\in\{0,1\}^k\setminus 0^k}g_{\omega\cdot H}(x+\omega\cdot h).$$

By the Gowers-Cauchy-Schwarz inequality from Section 2.1, we find that this expression is bounded above in absolute value by

$$\mathbb{E}_{H\in(\mathbb{Z}_N^k)^m}\|g\|_{U^k}\prod_{\omega\in\{0,1\}^k\setminus 0^k}\|g_{\omega\cdot H}\|_{U^k}\leq\sup_{\omega\in\{0,1\}^k\setminus 0^k}\mathbb{E}_{H\in(\mathbb{Z}_N^k)^m}\|g_{\omega\cdot H}\|_{U^k}^{2^k}.$$

But for fixed $\omega\in\{0,1\}^k\setminus 0^k$, as H runs through $(\mathbb{Z}_N^k)^m$, $\omega\cdot H$ runs through all values of $\mathbb{Z}_N$, so the expression we are trying to bound can be rewritten as

$$\mathbb{E}_{u^{(1)},\dots,u^{(m)}\in\mathbb{Z}_N}\mathbb{E}_{x\in\mathbb{Z}_N,h\in\mathbb{Z}_N^k}\prod_{\omega'\in\{0,1\}^k}\prod_{j=1}^{m}f_j(x+u^{(j)}+\omega'\cdot h)$$
$$=\mathbb{E}_{x\in\mathbb{Z}_N,h\in\mathbb{Z}_N^k}\prod_{j=1}^{m}(\mathbb{E}_{u^{(j)}\in\mathbb{Z}_N}\prod_{\omega'\in\{0,1\}^k}f_j(x+u^{(j)}+\omega'\cdot h)).$$

By hypothesis on f_j and Hölder's inequality the latter average is bounded above by

$$\mathbb{E}_{x\in\mathbb{Z}_N,h\in\mathbb{Z}_N^k}(\mathbb{E}_{u\in\mathbb{Z}_N}\prod_{\omega'\in\{0,1\}^k}\nu(x+u+\omega'\cdot h))^m=\mathbb{E}_{h\in\mathbb{Z}_N^k}(\mathbb{E}_{y\in\mathbb{Z}_N}\prod_{\omega'\in\{0,1\}^k}\nu(y+\omega'\cdot h))^m,$$

which, by the correlation condition in Definition 4 applied to the expectation in y and the triangle inequality, is bounded by

$$\mathbb{E}_{h\in\mathbb{Z}_N^k}(\sum_{\omega'\neq\omega''\in\{0,1\}^k}\tau(h\cdot(\omega'-\omega'')))^m\leq C(m)\sup_{\omega'\neq\omega''\in\{0,1\}^k}\mathbb{E}_{h\in\mathbb{Z}_N^k}\tau(h\cdot(\omega'-\omega''))^m.$$

But for $\omega'\neq\omega''$, $h\cdot(\omega'-\omega'')$ uniformly covers $\mathbb{Z}_N$ as h runs through $\mathbb{Z}_N^k$, and τ is assumed to have bounded moments of all order, so the proof is complete. □

Even though Green and Tao's original approach to the transference principle is quite different, this lemma also played a crucial role there. It is the only point in the argument where the correlation condition is used, and it concludes our sketch of the proof of the transference principle. □

3.3. Existence of a pseudorandom measure

In order to complete the proof of the Green-Tao theorem, it remains to establish the existence of a pseudorandom measure that satisfies the conditions given in Definition 4. That is, we need to find a majorant for the modified von Mangoldt function that averages roughly 1 and satisfies the linear forms and correlation conditions. As we observed before, the Siegel-Walfisz theorem tells us that $\mathbb{E}_{n\in[N]}\widetilde{\Lambda}(n) = 1 + o(1)$, but higher-order correlations, such as the ones necessary for the linear forms and correlation conditions, are in general rather poorly understood. However, it is possible to replace Λ, which can be written as

$$\Lambda(n) = \sum_{d|n} \mu(d) \log \frac{n}{d},$$

where μ is the Möbius function[32], by a truncated version which is localized not to primes but to "almost" primes, meaning integers with no small prime factors. For example, averages such as

$$\mathbb{E}_{n\in[N]}\Lambda_R(n+h_1)\cdots\Lambda_R(n+h_m)$$

for arbitrary $h_1,\ldots,h_m$, where the truncated divisor sum Λ_R (for R a small power of N) is defined by

$$\Lambda_R(n) = \sum_{d|n,d<R} \mu(d) \log \frac{R}{d} = \sum_{d|n} \mu(d) \log_+ \frac{R}{d},$$

are much better understood as they fall within the remit of sieve theory methods. In particular, concrete asymptotics for such averages were given in the work of Goldston and Yıldırım **[16]** on small gaps between primes. A promising candidate for a majorant for $\widetilde{\Lambda}$ is thus the function

$$\nu(n) = \frac{\phi(W)}{W} \frac{\Lambda_R(Wn+1)^2}{\log R}$$

where W is the product of small primes introduced at the start of Section 3.[33]

It was observed by Tao ([**49**], see also **[33]** and Appendix D of **[31]**) that the necessary correlation estimates can, in a qualitative sense, be deduced from a quite elementary fact about the Riemann zeta function, namely that it has a simple pole of

[32] The Möbius function $\mu(n)$ is defined to be 1 when n is the product of an even number of distinct prime factors, -1 when it is the product of an odd number of distinct prime factors, and 0 otherwise. By definition $\mu(0) = 1$.

[33] One can check that if $Wn + 1$ is prime and n is of order comparable to N, with N sufficiently large, then $\Lambda_R(Wn + 1) = \log R$.

order 1 at $z = 1$. The trick is to smooth the truncated divisor sum Λ_R even further, introducing a mollifier $\chi : \mathbb{R} \to \mathbb{R}$ supported on $[-1, 1]$ and writing

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

Replacing Λ_R by $\Lambda_{R,\chi}$ in the definition of the majorant ν above (and renormalizing by a factor of $(\log R)^2$) makes the latter rather easier to handle. In the original proof of Green and Tao the function $\log_+ x$, which appears in the definition of Λ_R above, was expressed as a contour integral

$$\log_+ x = \frac{1}{2\pi i} \int \frac{x^z}{z^2}\, dz$$

along the vertical line $\frac{1}{\log R} + it$, which allowed them to estimate correlations of Λ_R using a multiple contour integral involving the Riemann zeta function ζ. This required knowledge of the classical zero-free region of ζ to the left of the line $z = 1$. The mollifier χ, on the other hand, can be expressed in terms of its Fourier transform, which decays rapidly and allows one to truncate the integrals involved to a short range, so that more elementary information about ζ is sufficient for their estimation. From a number-theoretic point of view the verification of the linear forms and correlation conditions is not particularly instructive but rather technical, so we shall give no further details here.[34]

3.4. Asymptotics and solutions to systems of linear equations

Looking back at the global sketch of the argument used to prove the Green-Tao theorem immediately following Theorem 14, we observe that this strategy, employed in **[25]**, only yields a lower bound on the number of k-term arithmetic progressions in the primes belonging to the interval $[N]$. It was subsequently shown in **[31]** that the transference principle together with the inverse theorem for the U^3 norm (Theorem 9) implies that there are asymptotically

$$\mathfrak{S}_4 \frac{N^2}{\log^4 N}$$

4-term progressions in the primes up to N. The term $N^2/\log^4 N$ arises from Cramer's probabilistic model of the primes, where we imagine each integer being prime independently with probability $1/\log N$. The local factor

$$\mathfrak{S}_4 = \frac{3}{4} \prod_{p \geq 5} \frac{p^2(p-3)}{(p-1)^3} \approx 0.4764$$

contains the arithmetical information.

[34] For an elementary verification of the fact that $\mathbb{E}\nu = 1 + o(1)$, see Host's treatment **[33]**.

Asymptotics for arithmetic progressions of length $k+1$ for $k > 3$ were given conditional on the U^k inverse theorem (Theorem 10), which was only conjectured at the time. Furthermore, Green and Tao required an additional fact regarding the Möbius function, namely the *Möbius and nilsequences conjecture*, which they were able to resolve shortly afterwards [**28**].

THEOREM 19 (Möbius is orthogonal to nilsequences). — *Let G/Γ be a k-step nilmanifold and $F(g^n x)$ be a k-step nilsequence. Then for any $C > 0$, we have*

$$|\mathbb{E}_{n\in[N]}\mu(n)F(g^n x)| \ll (\log N)^{-C},$$

where the implied constant depends on G/Γ, k, C and the Lipschitz constant of F (but not on x or g).

Let us briefly outline how to use the transference principle, the inverse theorem and Theorem 19 to obtain asymptotics for k-term arithmetic progressions in the primes. First, note that the transference principle and the inverse theorem can be combined to give a *transferred inverse theorem*, that is, an inverse theorem for functions that are bounded by a pseudorandom measure instead of a constant.(35) Using the decomposition of the von Mangoldt function in terms of the Möbius function together with relatively standard methods from analytic number theory, Theorem 19 implies that $\Lambda - 1$ is almost orthogonal to all $(k-1)$-step nilsequences. One immediately deduces that $\Lambda - 1$ has small U^k norm: otherwise $\Lambda - 1$ correlates with one of the nilsequences by the transferred inverse theorem, which leads to a contradiction. Then, by the relative generalized von Neumann theorem (Proposition 13), the average over a product of different instances of $\Lambda = 1 + (\Lambda - 1)$ is essentially, up to local factors, just the product over the constant part. The error terms remain ineffective without GRH.

Using a relative version of the more general Proposition 7 together with the transferred inverse theorem, one obtains asymptotics not only for long arithmetic progressions in the primes, but more generally for all systems of affine linear forms of finite complexity, no two of which are linearly dependent.(36) This solves the *generalized Hardy-Littlewood prime tuples conjecture*, excluding only notoriously hard problems such as the twin primes and Goldbach's conjecture (which have infinite complexity). A quirky consequence, for example, is a strengthening of Vinogradov's result [**53**] that every sufficiently large odd integer can be written as a sum of three primes $p_1+p_2+p_3$: one may impose the additional constraint that $p_1 - p_2$ be equal to a prime minus 1.(37)

(35) Such a statement was first given as Proposition 10.1 in [**31**].

(36) One needs to adjust the majorant slightly since it is now necessary to simultaneously control $\widetilde{\Lambda}_{b_1}, \ldots, \widetilde{\Lambda}_{b_t}$ for different values of $b_1, \ldots, b_t$, where $\widetilde{\Lambda}_b(n) = \frac{\phi(W)}{W}\log(Wn+b)$ whenever $Wn+b$ is prime, and zero otherwise.

(37) For a precise statement of the generalized Hardy-Littlewood prime tuples conjecture, see Conjecture 1.4 in [**31**].

4. POLYNOMIAL PROGRESSIONS IN THE PRIMES

The general philosophy of the proof of Theorem 4 is the same as that of the Green-Tao theorem. However, there are two immediate and serious obstacles to a straightforward adaptation of the argument. The first is that, as mentioned in Section 2.4, there is no quantitative version of a polynomial Szemerédi theorem, and we only have the qualitative Bergelson-Leibman theorem (Theorem 12) at our disposal. However, it is possible to obtain a "pseudo"-quantitative, finitary version of Theorem 12. This remains completely ineffective, which is why the final statement in Theorem 4 contains no explicit bounds. We discuss this step in more detail in Section 4.1.

The second and in some sense more serious difficulty is a direct consequence of the polynomial nature of the problem. Recall that when we were dealing with linear systems such as $x, x+d, x+2d$, if the variable x ranged over the interval $[N]$, then the range of the variable d was still comparable to $[N]$. However, in the case of polynomial systems such as $x, x+m^2$, the range of the variable m has to be restricted to a smaller range $[M] = [\sqrt{N}]$. This means that we can no longer control our averages by the uniformity norms introduced in Section 2.1. Instead, one has to define local versions of these norms, in which the ranges of some of the parameters are restricted to smaller scales. We shall give the precise definition of these norms in Section 4.2.

The resulting modifications of the remainder of the argument are somewhat more routine, but contain numerous subtleties and are technically formidable. In Section 4.3 we show how to explicitly derive a relative generalized von Neumann theorem for the specific case of the polynomial system $x, x+m^2$ mentioned above. This involves linearizing the polynomial system by the so-called *Polynomial Exhaustion Theorem* (or *PET induction*), which is an inductive process on polynomial systems that was crucial to the proof of Theorem 12.[38] A key ingredient for showing that these local uniformity norms are good for transference will be established, again in a special case, in Section 4.4. Of course, one requires somewhat more stringent conditions on the majorizing pseudorandom measure ν in order to be able to handle polynomial systems, labelled the *polynomial forms* and the *polynomial correlation condition.* Because of their technical complexity, we shall not state these formally but only describe some of the ideas that go into verifying these properties in Section 4.5.

The generalization of the argument to arbitrary families of polynomials brings with it an intimidating amount of notation, which we shall try to avoid as much as possible by presenting only special cases of (parts of) the argument. There are many different scales of parameters involved in the proof of Theorem 4: a coarse scale (the range M of m above, which should be thought of as a small power of N

[38] See also **[54]** for a combinatorial application.

depending on the maximal degree of the family of polynomials $P_1, \dots, P_k$), the degree of pseudorandomness required of the measure, the sieve level R, the error term in the transference principle, the "complexity" of the polynomial system and a fine scale H arising from applications of van der Corput's lemma. Since we cannot accurately represent the dependencies between these parameters in the remaining pages, we have chosen to refer to all error terms by $o(1)$ and ask the reader to take on trust (or to check for themselves) that these errors can be made to depend on each other in the way required.

4.1. The quantitative Bergelson-Leibman theorem

First we shall outline very briefly how to obtain a quantitative but ineffective polynomial Szemerédi theorem from the Bergelson-Leibman theorem. This approach is very much in line with Tao's general efforts to bring results from ergodic theory to the discrete setting. To save space and remain closer to the original presentation of Tao and Ziegler, we shall from now on adopt ergodic theory notation, setting $X = \mathbb{Z}_N$, using the integral $\int_X$ instead of the expectation $\mathbb{E}_{x \in X}$ and writing T for the shift map on X, which acts by $Tf(x) = f(x-1)$.

THEOREM 20 (Quantitative Bergelson-Leibman theorem). — *Let $\delta > 0$, and let $g : X \to \mathbb{R}$ be any function satisfying $0 \leq g \leq 1 + o(1)$ as well as $\int_X g \geq \delta - o(1)$. Then*

$$\mathbb{E}_{m \in [M]} \int_X T^{P_1(Wm)/W} g \cdots T^{P_k(Wm)/W} g \geq c(\delta) - o(1),$$

for some $c(\delta)$ depending on $\delta, P_1, \dots, P_k$ (but not on N or W).

The appearance of the parameter W in the statement may appear odd, but its role is identical to that in the linear case: it helps us deal with non-uniform distribution modulo small primes. In fact, the above statement with g replaced by the normalized counting function $f : X \to \mathbb{R}^+$, defined by

$$f(n) = \begin{cases} \frac{\phi(W)}{W} \log R & \text{whenever } n \in [N/2], Wn + b \in A, \\ 0 & \text{otherwise,} \end{cases}$$

where A is a subset of the primes of positive relative density, precisely corresponds to the "moreover" part of Theorem 4. As before, the parameter R denotes the sieve level and the ratio between $\log R$ and $\log N$ represents the relative density between the primes and almost primes. The integer b is chosen by the pigeonhole principle to ensure that

$$|\{n \in [N/2] : Wn + b \in A\}| \gg \frac{W}{\phi(W)} \frac{N}{\log N}.$$

SKETCH PROOF OF THEOREM 20: The deduction of Theorem 20 from (a multi-dimensional version of) Theorem 12 proceeds by a non-standard application of the

Furstenberg correspondence principle [14]. One argues by contradiction and supposes that Theorem 20 fails. An averaging argument attributed to Varnavides [52] tells us that it does so in a strong way.[39] After lifting to a higher-dimensional setting to eliminate the dependence on W and m, one invokes Furstenberg's correspondence principle to pass from the (quasi-)finite formulation to a dynamical system involving several commuting transformations, whose recurrence properties correspond to the structural information we already possess about the function g. Finally, weak sequential compactness is used to derive a contradiction with the multidimensional version of Theorem 12. □

This procedure can be written up in detail in less than two pages, and naturally gives no explicit bounds for $c(\delta)$ in terms of δ.[40]

4.2. Local uniformity norms

Motivated by the discussion in the introduction to this section, we need to introduce several scales into our usual definition of the U^k norm.

DEFINITION 5 (Averaged local uniformity norms). — *For steps $a = (a_1, \dots, a_d) \in \mathbb{Z}^d$, define the* local uniformity norm $U^a_{\sqrt{M}}$ *of $f : \mathbb{Z}_N \to \mathbb{R}$*[41] *by the formula*

$$\|f\|^{2^d}_{U^a_{\sqrt{M}}} = \mathbb{E}_{m^{(0)}, m^{(1)} \in [\sqrt{M}]^d} \int_X \prod_{\omega \in \{0,1\}^d} T^{a \cdot m^{(\omega)}} f,$$

where $\omega = (\omega_1, \dots, \omega_d)$, $m^{(i)} = (m_1^{(i)}, \dots, m_d^{(i)})$ and $a \cdot m^{(\omega)} = a_1 m_1^{(\omega_1)} + \cdots + a_d m_d^{(\omega_d)}$.

Moreover, given any integer $t \geq 0$ and any $Q = (Q_1, \dots, Q_d) \in \mathbb{Z}[h_1, \dots, h_t, W]^d$, write $Q(h_1, \dots, h_t, W)$ for the d-tuple $(Q_1(h_1, \dots, h_t, W), \dots, Q_d(h_1, \dots, h_t, W))$ of polynomials. Define the averaged local uniformity norm $U^{Q([H]^t, W)}_{\sqrt{M}}$ *as*

$$\|f\|^{2^d}_{U^{Q([H]^t, W)}_{\sqrt{M}}} = \mathbb{E}_{h_1, \dots, h_t \in [H]} \|f\|^{2^d}_{U^{Q(h_1, \dots, h_t, W)}_{\sqrt{M}}}.$$

While the global uniformity norms in Definition 1 measure the extent to which a function f behaves like a polynomial of degree at most k on all of $\mathbb{Z}_N$, the local norms measure the extent to which f is polynomial on arithmetic progressions of the form $\{x + a \cdot m : m_1, \dots, m_d \in [\sqrt{M}]\}$. The averaged version arises from various applications of van der Corput's lemma (Lemma 21 below). It turns out that these averages need to be taken over tiny ranges H (H should be thought of as a *very* small power of N, basically depending on all other parameters in the problem). This is one

[39] Whenever we can assert the existence of one non-trivial progression in a subset $A \subseteq [N]$ of density α, it is easy to show by averaging that we actually have $c(\alpha)N^2$ many.
[40] See Appendix B of [50].
[41] We shall omit the complex conjugate signs here since the functions we will be dealing with only take positive real values.

of the subtleties that make the adaptation of the argument from the linear case not entirely straightforward.[42]

It can be verified without too much trouble that each of the formulae in Definition 5 does indeed define a norm. An example which will make an appearance later on is the averaged local norm defined by

$$\|f\|_{ex}^2 = \mathbb{E}_{h,h' \in [H]} \mathbb{E}_{k,k' \in [\sqrt{M}]} \int_X T^{2(h'-h)k} f T^{2(h'-h)k'} f.$$

Here $t = 2, d = 1$ and $Q(h, h', W) = 2(h' - h)$.

Contrary to the global U^k norms, which are by now rather well understood, we know relatively little about the local ones. In the next section we shall use them to bound above a polynomial average in the spirit of the relative generalized von Neumann theorem.

4.3. The relative generalized von Neumann theorem for polynomials

The proof of the relative generalized von Neumann theorem for polynomial systems contains an important new ingredient over the linear case, borrowed from relevant work in ergodic theory (in particular, the proof of Theorem 12). PET induction is needed to linearize the polynomial system in successive stages before the Cauchy-Schwarz inequality can take over to bound the resulting linear system by an expression resembling a local uniformity norm. In a final step one has to get rid of a certain weight function introduced by this procedure, by applying the Cauchy-Schwarz inequality one more time, similarly to what we saw in Proposition 13. In order to carry out the linearization procedure, one requires a slight but well-known variant of the Cauchy-Schwarz inequality which allows one to replace coarse-scale averages by coarse-scale averages of shifts over much smaller scales.

LEMMA 21 (van der Corput). — *Let N, M, H be as above. Let $(x_m)_{m \in \mathbb{Z}}$ be a sequence of reals satisfying $x_m \ll_\epsilon N^\epsilon$ for all $\epsilon > 0$, $m \in \mathbb{Z}$. Then*

$$\mathbb{E}_{m \in [M]} x_m = \mathbb{E}_{h \in [H]} \mathbb{E}_{m \in [M]} x_{m+h} + o(1)$$

and

$$|\mathbb{E}_{m \in [M]} x_m|^2 \ll \mathbb{E}_{h,h' \in [H]} \mathbb{E}_{m \in [M]} x_{m+h} x_{m+h'} + o(1).$$

Proof. — By assumption on the growth of x_m,

$$\mathbb{E}_{m \in [M]} x_m = \mathbb{E}_{m \in [M]} x_{m+h} + o(1)$$

[42] The parameter W will not appear in our sample calculations, but is necessary for the final result (see Theorem 20 and the discussion that follows).

for all $h \in [H]$. The same is certainly true for the average in h, and so the first part of the result follows. The second part is obtained by applying the Cauchy-Schwarz inequality to the first. □

We shall use the very simple example of a polynomial system $x, x+m^2$ to illustrate how to combine PET linearization and Cauchy-Schwarz to prove a polynomial version of the relative generalized von Neumann theorem.[43] It should give the reader an idea of what kinds of polynomial averages of ν need to be controlled by the polynomial forms condition, which we shall not state formally in this text.

EXAMPLE 2. — *Let $f_1, f_2 : \mathbb{Z}_N \to \mathbb{R}$ be any functions satisfying $|f_j(x)| \leq \nu(x)$ for $j = 1, 2$, where ν satisfies the polynomial forms condition. Then*

$$|\mathbb{E}_{m\in[M]} \int_X f_1 T^{m^2} f_2| \ll \|f_2\|_{ex}^{1/2} + o(1).$$

Proof of Example 2. — The first step is reminiscent of Weyl differencing: we turn the polynomial system into one that is somewhat more linear. We begin with

$$|\mathbb{E}_{m\in[M]} \int_X f_1 T^{m^2} f_2|^2 \leq (\int_X \nu)(\int_X \nu |\mathbb{E}_{m\in[M]} T^{m^2} f_2|^2),$$

which equals

$$(1 + o(1)) \int_X \nu |\mathbb{E}_{m\in[M]} T^{m^2} f_2|^2,$$

by assumption on ν. By the second part of van der Corput's lemma[44], we have

$$\int_X \nu |\mathbb{E}_{m\in[M]} T^{m^2} f_2|^2 = \mathbb{E}_{h,h'\in[H]} \mathbb{E}_{m\in[M]} \int_X \nu T^{(m+h)^2} f_2 T^{(m+h')^2} f_2 + o(1).$$

As mentioned above, it turns out that the range H needs to be very small compared with M (we shall see why in Example 3 below). Exploiting the translation invariance of the integral over X, we shall now shift the entire integrand by $-(m+h)^2$ to obtain new shifts $R_1(m,h,h') = -(m+h)^2$, $R_2(m,h,h') = 0$, $R_2'(m,h,h') = 2m(h'-h) + (h'^2 - h^2)$, the latter now being linear in m. So we need to estimate from above the expression

$$\mathbb{E}_{h,h'\in[H]} \mathbb{E}_{m\in[M]} \int_X f_2 T^{R_1(m,h,h')} \nu T^{2m(h'-h)+(h'^2-h^2)} f_2.$$

---

[43] Of course, this system is covered by the strong bound in Sárközy's theorem (Theorem 11) and therefore the procedure that follows is not strictly speaking necessary in this case. However, even the only slightly more complicated system $x, x+m, x+m^2$ results in a local uniformity norm involving ten different variables, for which the amount of notation required would be almost as off-putting as the general case, given by Proposition 5.9 in **[50]**.

[44] Here we need the estimate $\nu \ll_\epsilon N^\epsilon$ for every $\epsilon > 0$. Fortunately this is a simple consequence of the well-known fact that the number of divisors of an integer N is $\ll_\epsilon N^\epsilon$ for any $\epsilon > 0$.

Let $k \in [\sqrt{M}]$, and observe that the first part of van der Corput's lemma yields that the above expression is equal to

$$\mathbb{E}_{h,h' \in [H]} \mathbb{E}_{m \in [M]} \int_X \mathbb{E}_{k \in [\sqrt{M}]} f_2 g_1 T^{2(m-k)(h'-h)+(h'^2-h^2)} f_2 + o(1),$$

where $g_1 = T^{R_1(m-k,h,h')}\nu$. Letting $Q_0 = Q_0(k,h,h') = 2(h'-h)k$ and shifting again, we have

$$\mathbb{E}_{h,h' \in [H]} \mathbb{E}_{m \in [M]} \int_X \mathbb{E}_{k \in [\sqrt{M}]} T^{Q_0(k,h,h')} f_2 T^{Q_0(k,h,h')} g_1 T^{2m(h'-h)+(h'^2-h^2)} f_2 + o(1),$$

the crucial point being that the final instance of f_2 is independent of k. Writing $g_2 = T^{2m(h'-h)+(h'^2-h^2)}\nu$ and applying the Cauchy-Schwarz inequality one more time, we obtain an upper bound of

$$(\mathbb{E}_{h,h' \in [H]} \mathbb{E}_{m \in [M]} \int_X |\mathbb{E}_{k \in [\sqrt{M}]} T^{Q_0} f_2 T^{Q_0} g_1|^2 g_2)^{1/2} (\mathbb{E}_{h,h' \in [H]} \mathbb{E}_{m \in [M]} \int_X g_2)^{1/2} + o(1).$$

The average in g_2 is exactly of a form controlled by the polynomial forms condition, and hence bounded above by a constant. The first average, on the other hand, can be rewritten (ignoring the square-root) as

$$\mathbb{E}_{h,h' \in [H]} \mathbb{E}_{k,k' \in [\sqrt{M}]} \int_X T^{Q_0(k,h,h')} f_2 T^{Q_0(k',h,h')} f_2 W(h,h',k,k'),$$

where $W(h,h',k,k') = \mathbb{E}_{m \in [M]} g_2 T^{Q_0(k,h,h')} g_1 T^{Q_0(k',h,h')} g_1'$, and g_1' is identical to g_1 except with k replaced by k'. Now clearly

$$\mathbb{E}_{h,h' \in [H]} \mathbb{E}_{k,k' \in [\sqrt{M}]} \int_X T^{Q_0(k,h,h')} f_2 T^{Q_0(k',h,h')} f_2 = \|f_2\|_{ex}^2,$$

and it thus remains to replace the weight function W, consisting of averages of shifts of ν, by 1 asymptotically. This is done by applying the Cauchy-Schwarz inequality one more time to the difference between the expression we have and the local uniformity norm we are aiming for, and expanding out the square $(W(h,h',k,k')-1)^2$ similarly to the end of the proof of Proposition 13. All resulting averages of shifts of ν are controlled by the polynomial forms condition. □

The general PET induction scheme assigns a weight vector to each family of polynomials, and the aim is to reduce the weight of the polynomials that appear in the average (in some ordering) until the point where one reaches linearity in the variable m for all "active" functions (those that have not been replaced by ν).

4.4. Transference in the polynomial case

In order to apply the transference technique from Section 3.2, we need to check that the local uniformity norms satisfy the required properties (in the terminology of Gowers [18], that they are *quasi-algebra predual norms*). In particular, we need an analogue of Lemma 18 for the local norms. As before, we shall only do a toy calculation so as to avoid overburdening the reader with notation. As a simple example of a dual function consider

$$\mathcal{D}_{ex} f = \mathbb{E}_{h,h' \in [H]} \mathbb{E}_{k,k' \in [\sqrt{M}]} T^{2(h-h')(k-k')} f,$$

which corresponds to the local uniformity norm $\|\cdot\|_{ex}$ we obtained when bounding the system $x, x+m^2$ in the preceding section. While in Section 3 we showed directly that the U^k dual norm of a product of these dual functions was small, we shall proceed more indirectly here and prove the consequence of this fact that we actually need, namely that the product of dual functions is practically orthogonal to $\nu - 1$.[45] Again, our hope is that following the calculation in this example will convey a good picture of the type of polynomial correlation condition the majorant ν needs to satisfy.

EXAMPLE 3. — *Let $f_1, \dots, f_m : \mathbb{Z}_N \to \mathbb{R}$ be any functions satisfying $|f_j(x)| \leq \nu(x)$ for $j = 1, \dots, m$, where ν satisfies the polynomial correlation condition. Then*

$$\int_X (\mathcal{D}_{ex} f_1 \cdots \mathcal{D}_{ex} f_m)(\nu - 1) = o(1),$$

where the dual operator $\mathcal{D}_{ex}$ is the one defined above.

Before we begin the proof, note that if one simply applied absolute value signs and bounded everything above by ν, then the pseudorandomness properties of ν would only imply an upper bound of $O(1)$. One might also naively want to proceed by applying the Cauchy-Schwarz inequality many times, but this would result in problems since m could be too large to be multiplied with some of the $o(1)$ error terms.

Proof of Example 3. — The left-hand side can be rewritten as

$$\mathbb{E}_{h_1,\dots,h_m,h'_1,\dots,h'_m} \mathbb{E}_{k_1,\dots,k_m,k'_1,\dots,k'_m} \int_X \prod_{j=1}^{m} T^{2(h_j-h'_j)(k_j-k'_j)} f_j(\nu - 1).$$

Here the range of the variables h will be the very short interval $[H]$, while that of the variables k is the coarse scale $[\sqrt{M}]$. Setting $u_j = \prod_{l \neq j}(h_l - h'_l)$ and shifting

[45] The general statement to this effect is Proposition 6.5 in [50].

each variable k_j by $u_j(n-n')$ for $n, n' \in [M^{1/4}]$, this expression equals, by van der Corput's lemma

$$\mathbb{E}_{n,n' \in [M^{1/4}]} \mathbb{E}_{h_1,\dots,h_m,h'_1,\dots,h'_m} \mathbb{E}_{k_1,\dots,k_m,k'_1,\dots,k'_m} \int_X \prod_{j=1}^m T^{2(h_j-h'_j)((k_j-k'_j)+u_j(n-n'))} f_j(\nu-1) + o(1).$$

Note that this was only possible because each h belonged to the narrow range $[H]$. The introduction of the new variables now allows us to get away with a single application of Cauchy-Schwarz by separating n and n'. With this in mind, we shift the entire integral by $n'u$, where $u = 2\prod_{l=1}^m (h_l - h'_l)$ is again small compared with $M^{1/4}$, and find that the above is equal to

$$\mathbb{E}_{n,n' \in [M^{1/4}]} \mathbb{E}_{h_1,\dots,h_m,h'_1,\dots,h'_m} \mathbb{E}_{k_1,\dots,k_m,k'_1,\dots,k'_m} \int_X T^{nu}(\prod_{j=1}^m T^{2(h_j-h'_j)(k_j-k'_j)} f_j) T^{n'u}(\nu-1) + o(1).$$

We can now factor this expression as

$$\mathbb{E}_{h_1,\dots,h_m,h'_1,\dots,h'_m} \int_X \mathbb{E}_{n \in [M^{1/4}]} \prod_{j=1}^m \mathbb{E}_{k,k'} T^{nu+2(k-k')(h_j-h'_j)} f_j \times \mathbb{E}_{n' \in [M^{1/4}]} T^{n'u}(\nu-1) + o(1),$$

which by Cauchy-Schwarz and the assumption on f_j is at most

$$(\mathbb{E}_{h_1,\dots,h_m,h'_1,\dots,h'_m} \int_X (\mathbb{E}_{n \in [M^{1/4}]} \prod_{j=1}^m \mathbb{E}_{k,k'} T^{nu+2(k-k')(h_j-h'_j)} \nu)^2)^{1/2}$$
$$\times (\mathbb{E}_{h_1,\dots,h_m,h'_1,\dots,h'_m} \int_X (\mathbb{E}_{n' \in [M^{1/4}]} T^{n'u}(\nu-1))^2)^{1/2} + o(1),$$

It turns out that this is precisely the kind of average that is controlled by the polynomial correlation condition.[46] In particular, upon expanding the square, the first average equals $1 + o(1)$, while the second splits into three separate averages over ν whose constant contributions cancel, leaving us with a final estimate of $o(1)$. □

4.5. Existence of the pseudorandom measure

Finally, one needs to look for a measure ν that majorizes the function f defined in Section 4.1 pointwise, and satisfies the polynomial forms and polynomial correlation conditions, of which we hope to have given the reader a taste in the preceding two sections. In any case, the exact conditions are somewhat artificial and a compromise between what is needed and what can be proved about the primes.

[46] It is important here that m can be arbitrarily large (otherwise this statement reduces to a special case of the polynomial forms condition), and that the inside averages are over coarse scales.

As in the linear case, the construction of ν can be given completely explicitly by

$$\nu(n) = \frac{\phi(W)}{W} \log R \left(\sum_{d|Wn+b} \mu(d) \chi\left(\frac{\log d}{\log R}\right) \right)^2,$$

where χ is a smooth compactly supported cutoff obeying the normalization condition $\int_0^1 |\chi'(t)|^2 dt = 1$. Correlations over coarse scales are controlled rather well, either by the methods of Goldston and Yıldırım, or using the more elementary approach described in Section 3.3 which exploits the smoothness of χ. However, those averages that involve parameters on the finer scale $[H]$, which we were forced to introduce in Example 2 and whose magnitude was determined by the calculation in Example 3, are no longer controlled by elementary sieve theory methods. Indeed, some of the estimates required would be equivalent to understanding the distribution of primes in short intervals, which are beyond current techniques.

The way to handle this situation is to first fix the fine-scale parameters and estimate the remaining sums using sieve theory methods, before the resulting (now tractable) sums are averaged over finer scales. In the process of verifying the polynomial forms and correlation conditions, one needs to control the density of varieties such as $\{x \in \mathbb{F}_p^d : WP(x) + b = 0\}$. In the most general case this requires the Weil conjectures, but here the polynomial P will always be linear in at least one of the coarse-scale variables and one can get away with much more elementary estimates, avoiding the modern tools from arithmetic geometry altogether. An adaptation of the *combinatorial Nullstellensatz* in **[1]** to the case of several jointly coprime polynomials turns out to be sufficient.(47) The verification of the pseudorandomness conditions for ν is nevertheless highly technical, and we shall be able to say no more about it in the present exposition.(48)

5. CONCLUDING REMARKS

The proof of Theorem 4 shows that there are $cNM/\log^k N$ polynomial progressions $x + P_1(m), \ldots, x + P_k(m)$ in the primes with $x \in [N], m \in [M]$, for some constant c. Moreover, it is true that there are infinitely many "short" progressions, that is progressions for which m is comparable to $x^{1/2d}$, where d is the maximal degree of the family of polynomials. Indeed, disregarding the values of x of size $\ll N$, we can assume that x is comparable to N and similarly, m is comparable to M. For fixed k, one therefore obtains infinitely many progressions of the form $x + P_1(m), \ldots, x + P_k(m)$ in which

(47) See Appendix D of **[50]**.
(48) The interested reader is in the first instance referred to the excellent discussion on page 270 of **[50]**.

m is comparable to $x^{1/2d}$. By a diagonalization argument one can actually obtain infinitely many such progressions with $m = x^{o(1)}$. In a subset A of the integers of positive upper density something stronger is true: there is a fixed $m \neq 0$ for which the set $\{x : x + P_j(m) \in A \text{ for all } j \in [k]\}$ is infinite, and in fact of positive upper density (this follows from the Bergelson-Leibman theorem). A similar result for the primes, even in the simplest case, amounts to saying that the primes have bounded gaps arbitrarily often, which even by the strong results of Goldston, Pintz and Yıldırım is not known unconditionally.

The conjectured asymptotics for polynomial configurations in the primes are given by the Bateman-Horn conjecture [**2**].

CONJECTURE 22 (Bateman-Horn conjecture). — *Let $P_1, \dots, P_k \in \mathbb{Z}[m]$ be irreducible polynomials with positive leading coefficient, of degree $d_1, \dots, d_k$, respectively, such that the product $P_1 \cdots P_k$ has no non-trivial constant factor.*[49] *Then the number of positive integers $m \in [N]$ such that $P_1(m), \dots, P_k(m)$ are all prime is asymptotic to*

$$\frac{C(P_1, \dots, P_k)}{d_1 \cdots d_k} \int_2^N \frac{du}{\log^k u},$$

where

$$C(P_1, \dots, P_k) = \prod_p \frac{1 - \frac{\omega(p)}{p}}{(1 - \frac{1}{p})^k},$$

and $\omega(p)$ is the number of solutions to the congruence $P_1(x) \cdots P_k(x) \equiv 0 \bmod p$.

While the work of Tao and Ziegler is consistent with these predictions, these asymptotics are out of reach for now. This is largely due to the lack of an inverse theorem for the local Gowers norms (see Section 3.4 for a discussion of how such an inverse theorem for the global Gowers norms implies asymptotics for certain linear configurations).

Some interesting work has also been done on this sphere of problems in the function field setting. In particular, Lê [**37**] proved a version of the Green-Tao theorem for function fields: for every k, there exists $f, g \in \mathbb{F}_q[t]$, $g \neq 0$ such that the elements of $\{f + Pg : P \in \mathbb{F}_q[t], \deg(P) < k\}$ are all irreducible. While the function field setting appears to be a useful model for polynomial problems in the integers, it turns out that the naive analogue of the Bateman-Horn conjecture is false in this context.[50]

[49] If a prime p divides $P_1 \cdots P_k$, it can be the only prime amongst the values of the polynomials.
[50] For a discussion see [**9**].

Acknowledgements

The author would like to thank Tim Gowers, Ben Green and Terence Tao for their thoughts on the manuscript as well as the many insights they have generously shared over the years. She is also greatly indebted to Bernard Host and Tamar Ziegler for numerous useful comments and corrections, and Tom Sanders for helpful conversations and moral support.

REFERENCES

[1] N. ALON – "Combinatorial Nullstellensatz", *Combin. Probab. Comput.* **8** (1999), nos. 1-2, p. 7–29.

[2] P. T. BATEMAN & R. A. HORN – "A heuristic asymptotic formula concerning the distribution of prime numbers", *Math. Comp.* **16** (1962), p. 363–367.

[3] F. A. BEHREND – "On sets of integers which contain no three terms in arithmetical progression", *Proc. Nat. Acad. Sci. U. S. A.* **32** (1946), p. 331–332.

[4] V. BERGELSON – "Weakly mixing PET", *Ergodic Theory Dynam. Systems* **7** (1987), no. 3, p. 337–349.

[5] V. BERGELSON, B. HOST & B. KRA – "Multiple recurrence and nilsequences", *Invent. Math.* **160** (2005), no. 2, p. 261–303.

[6] V. BERGELSON & A. LEIBMAN – "Polynomial extensions of van der Waerden's and Szemerédi's theorems", *J. Amer. Math. Soc.* **9** (1996), no. 3, p. 725–753.

[7] J. BOURGAIN – "On triples in arithmetic progression", *Geom. Funct. Anal.* **9** (1999), no. 5, p. 968–984.

[8] D. CONLON & W. T. GOWERS – "Combinatorial theorems in sparse random sets", preprint arXiv:1011.4310.

[9] K. CONRAD – "Irreducible values of polynomials: a non-analogy", in *Number fields and function fields—two parallel worlds*, Progr. Math., vol. 239, Birkhäuser, 2005, p. 71–85.

[10] J.-P. CONZE & E. LESIGNE – "Sur un théorème ergodique pour des mesures diagonales", *C. R. Acad. Sci. Paris Sér. I Math.* **306** (1988), no. 12, p. 491–493.

[11] J. G. VAN DER CORPUT – "Über Summen von Primzahlen und Primzahlquadraten", *Math. Ann.* **116** (1939), p. 1–50.

[12] P. ERDŐS & P. TURÁN – "On Some Sequences of Integers", *J. London Math. Soc.* **S1-11** , no. 4, p. 261–264.

[13] G. A. FREĬMAN – *Foundations of a structural theory of set addition*, Translations of Mathematical Monographs, vol. 37, Amer. Math. Soc., 1973.

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

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

[16] D. A. GOLDSTON & C. Y. YILDIRIM – "Higher correlations of divisor sums related to primes. III. Small gaps between primes", *Proc. Lond. Math. Soc.* **95** (2007), no. 3, p. 653–686.

[17] W. T. GOWERS – "A new proof of Szemerédi's theorem", *Geom. Funct. Anal.* **11** (2001), no. 3, p. 465–588.

[18] ———, "Decompositions, approximate structure, transference, and the Hahn-Banach theorem", *Bull. Lond. Math. Soc.* **42** (2010), no. 4, p. 573–606.

[19] W. T. GOWERS & J. WOLF – "The true complexity of a system of linear equations", *Proc. Lond. Math. Soc.* **100** (2010), no. 1, p. 155–176.

[20] ———, "Linear forms and quadratic uniformity for functions on $\mathbb{Z}_N$", *J. Anal. Math.* **115** (2011), p. 121–186.

[21] B. GREEN – "On arithmetic structures in dense sets of integers", *Duke Math. J.* **114** (2002), no. 2, p. 215–238.

[22] ———, "Roth's theorem in the primes", *Ann. of Math.* **161** (2005), no. 3, p. 1609–1636.

[23] ———, "Long arithmetic progressions of primes", in *Analytic number theory*, Clay Math. Proc., vol. 7, Amer. Math. Soc., 2007, p. 149–167.

[24] B. GREEN & T. TAO – "An inverse theorem for the Gowers $U^3(G)$ norm", *Proc. Edinb. Math. Soc.* **51** (2008), no. 1, p. 73–153.

[25] ———, "The primes contain arbitrarily long arithmetic progressions", *Ann. of Math.* **167** (2008), no. 2, p. 481–547.

[26] ———, "New bounds for Szemerédi's theorem. II. A new bound for $r_4(N)$", in *Analytic number theory*, Cambridge Univ. Press, 2009, p. 180–204.

[27] ———, "An arithmetic regularity lemma, an associated counting lemma, and applications", in *An irregular mind*, Bolyai Soc. Math. Stud., vol. 21, János Bolyai Math. Soc., 2010, p. 261–334.

[28] ———, "The Möbius function is strongly orthogonal to nilsequences", *Ann. of Math.* **175** (2012), no. 2, p. 541–566.

[29] B. GREEN, T. TAO & T. ZIEGLER – "An inverse theorem for the Gowers $U^{s+1}[N]$-norm", *Electron. Res. Announc. Math. Sci.* **18** (2011), p. 69–90.

[30] ———, "An inverse theorem for the Gowers $U^{s+1}[N]$-norm", *Ann. of Math.* **176** (2012), no. 2, p. 1231–1372.

[31] B. GREEN & T. TAO – "Linear equations in primes", *Ann. of Math.* **171** (2010), no. 3, p. 1753–1850.

[32] M. HAMEL & I. ŁABA – "Arithmetic structures in random sets", *Integers* **8** (2008), p. A04, 21.

[33] B. HOST – "Progressions arithmétiques dans les nombres premiers (d'après B. Green et T. Tao)", Séminaire Bourbaki, vol. 2004/2005, exposé n° 944, *Astérisque* **307** (2006), p. 229–246.

[34] B. HOST & B. KRA – "Nonconventional ergodic averages and nilmanifolds", *Ann. of Math.* **161** (2005), no. 1, p. 397–488.

[35] Y. KOHAYAKAWA, T. ŁUCZAK & V. RÖDL – "Arithmetic progressions of length three in subsets of a random set", *Acta Arith.* **75** (1996), no. 2, p. 133–163.

[36] B. KRA – "The Green-Tao theorem on arithmetic progressions in the primes: an ergodic point of view", *Bull. Amer. Math. Soc. (N.S.)* **43** (2006), no. 1, p. 3–23.

[37] T. H. LÊ – "Green-Tao theorem in function fields", *Acta Arith.* **147** (2011), no. 2, p. 129–152.

[38] J. PINTZ, W. L. STEIGER & E. SZEMERÉDI – "On sets of natural numbers whose difference set contains no squares", *J. London Math. Soc.* **37** (1988), no. 2, p. 219–231.

[39] O. RAMARÉ & I. Z. RUZSA – "Additive properties of dense subsets of sifted sequences", *J. Théor. Nombres Bordeaux* **13** (2001), no. 2, p. 559–581.

[40] O. REINGOLD, L. TREVISAN, M. TULSIANI & S. VADHAN – "Dense subsets of pseudorandom sets", in *49th Annual IEEE Symposium on Foundations of Computer Science (FOCS), Philadephia*, 2008, p. 76–85.

[41] K. F. ROTH – "On certain sets of integers", *J. London Math. Soc.* **28** (1953), p. 104–109.

[42] I. Z. RUZSA – "Generalized arithmetical progressions and sumsets", *Acta Math. Hungar.* **65** (1994), no. 4, p. 379–388.

[43] T. SANDERS – "On Roth's theorem on progressions", *Ann. of Math.* **174** (2011), no. 1, p. 619–636.

[44] ———, "On the Bogolyubov-Ruzsa lemma", *Anal. PDE* **5** (2012), no. 3, p. 627–655.

[45] A. SÁRKŐZY – "On difference sets of sequences of integers. I", *Acta Math. Acad. Sci. Hungar.* **31** (1978), nos. 1–2, p. 125–149.

[46] T. SCHOEN & I. SHKREDOV – "Roth's theorem in many variables", preprint arXiv:1106.1601.

[47] E. SZEMERÉDI – "On sets of integers containing no k elements in arithmetic progression", *Acta Arith.* **27** (1975), p. 199–245.

[48] T. TAO – "The dichotomy between structure and randomness, arithmetic progressions, and the primes", in *International Congress of Mathematicians. Vol. I*, Eur. Math. Soc., Zürich, 2007, p. 581–608.

[49] ———, "A remark on Goldston-Yıldırım correlation estimates", `http://www.math.ucla.edu/~tao/preprints/Expository/gy-corr.dvi`.

[50] T. TAO & T. ZIEGLER – "The primes contain arbitrarily long polynomial progressions", *Acta Math.* **201** (2008), no. 2, p. 213–305.

[51] L. TREVISAN, M. TULSIANI & S. VADHAN – "Regularity, boosting, and efficiently simulating every high-entropy distribution", in *24th Annual IEEE Conference on Computational Complexity, Paris*, 2009, p. 126–136.

[52] P. VARNAVIDES – "On certain sets of positive density", *J. London Math. Soc.* **34** (1959), p. 358–360.

[53] I. M. VINOGRADOV – "Some theorems concerning the primes", *Mat. Sbornik* **2** (1937), p. 179–195.

[54] M. WALTERS – "Combinatorial proofs of the polynomial van der Waerden theorem and the polynomial Hales-Jewett theorem", *J. London Math. Soc.* **61** (2000), no. 1, p. 1–12.

[55] T. ZIEGLER – "Universal characteristic factors and Furstenberg averages", *J. Amer. Math. Soc.* **20** (2007), no. 1, p. 53–97.

Julia WOLF

École Polytechnique
Centre de Mathématiques Laurent Schwartz
91128 Palaiseau Cedex
E-mail : `julia.wolf@cantab.net`