The dichotomy between structure and randomness, arithmetic progressions, and the primes

Terence Tao*

Abstract. A famous theorem of Szemerédi asserts that all subsets of the integers with positive upper density will contain arbitrarily long arithmetic progressions. There are many different proofs of this deep theorem, but they are all based on a fundamental dichotomy between structure and randomness, which in turn leads (roughly speaking) to a decomposition of any object into a structured (low-complexity) component and a random (discorrelated) component. Important examples of these types of decompositions include the Furstenberg structure theorem and the Szemerédi regularity lemma. One recent application of this dichotomy is the result of Green and Tao establishing that the prime numbers contain arbitrarily long arithmetic progressions (despite having density zero in the integers). The power of this dichotomy is evidenced by the fact that the Green–Tao theorem requires surprisingly little technology from analytic number theory, relying instead almost exclusively on manifestations of this dichotomy such as Szemerédi's theorem. In this paper we survey various manifestations of this dichotomy in combinatorics, harmonic analysis, ergodic theory, and number theory. As we hope to emphasize here, the underlying themes in these arguments are remarkably similar even though the contexts are radically different.

Mathematics Subject Classification (2000). Primary 11P32, 37A45, 05C65, 05C75, 42A99.

Keywords. Szemerédi's theorem, ergodic theory, graph theory, hypergraph theory, arithmetic combinatorics, arithmetic progressions, prime numbers.

1. Introduction

In 1975, Szemerédi [53] proved the following deep and enormously influential theorem:

Theorem 1.1 (Szemerédi's theorem). *Let A be a subset of the integers $\mathbb{Z}$ of positive upper density, thus* $\limsup_{N\to\infty} \frac{|A\cap[-N,N]|}{|[-N,N]|} > 0$. *Here $|A|$ denotes the cardinality of a set A, and $[-N, N]$ denotes the integers between $-N$ and N. Then for any $k \geq 3$, A contains infinitely many arithmetic progressions of length k.*

Several proofs of this theorem are now known. The original proof of Szemerédi [53] was combinatorial. A later proof of Furstenberg [11], [13] used ergodic theory and has led to many extensions. A more quantitative proof of Gowers [19], [20] was based on Fourier analysis and arithmetic combinatorics (extending a much older

*The author is supported by a grant from the Packard foundation.

Proceedings of the International Congress
of Mathematicians, Madrid, Spain, 2006

argument of Roth [50] handling the $k = 3$ case). A fourth proof by Gowers [21] and Rödl, Nagle, Schacht, and Skokan [46], [47], [48], [49] relied on the structural theory of hypergraphs. These proofs are superficially all very different (with each having their own strengths and weaknesses), but have a surprising number of features in common. The main difficulty in all of the proofs is that one *a priori* has no control on the behaviour of the set A other than a lower bound on its density; A could range from being a very random set, to a very structured set, to something in between. In each of these cases, A will contain many arithmetic progressions – but the *reason* for having these progressions varies from case to case. Let us illustrate this by informally discussing some representative examples:

- (Random sets) Let $0 < \delta < 1$, and let A be a random subset of $\mathbb{Z}$, which each integer n lying in A with an independent probability of δ. Then A almost surely has upper density δ, and it is easy to establish that A almost surely has infinitely many arithmetic progressions of length k, basically because each progression of length k in $\mathbb{Z}$ has a probability of δ^k of also lying in A. A more refined version of this argument also applies when A is *pseudorandom* rather than random – thus we allow A to be deterministic, but require that a suitable number of correlations (e.g. pair correlations, or higher order correlations) of A are negligible. The argument also extends to sparse random sets, for instance one where $\mathbf{P}(n \in A) \sim 1/\log n$.

- (Linearly structured sets) Consider a quasiperiodic set such as $A := \{n : \{\alpha n\} \leq \delta\}$, where $0 < \delta < 1$ is fixed, α is a real number (e.g. $\alpha = \sqrt{2}$) and $\{x\}$ denotes the fractional part of x. Such sets are "almost periodic" because there is a strong correlation between the events $n \in A$ and $n + L \in A$, thanks to the identity $\{\alpha(n + L)\} - \{\alpha n\} = \{\alpha L\} \mod 1$. An easy application of the Dirichlet approximation theorem (to locate an approximate period L with $\{\alpha L\}$ small) shows that such sets still have infinitely many progressions of any given length k. Note that this argument works regardless of whether α is rational or irrational.

- (Quadratically structured sets) Consider a "quadratically quasiperiodic" set of the form $A := \{n : \{\alpha n^2\} \leq \delta\}$. If α is irrational, then this set has upper density δ, thanks to Weyl's theorem on equidistribution of polynomials. (If α is rational, one can still obtain some lower bound on the upper density.) It is not linearly structured (there is no asymptotic correlation between the events $n \in A$ and $n + L \in A$ as $n \to \infty$ for any fixed non-zero L), however it has quadratic structure in the sense that there is a strong correlation between the events $n \in A$, $n + L \in A$, $n + 2L \in A$, thanks to the identity

$$\{\alpha n^2\} - 2\{\alpha(n + L)^2\} + \{\alpha(n + 2L)^2\} = 2\{\alpha L^2\} \mod 1.$$

  In particular A does not behave like a random set. Nevertheless, the quadratic structure still ensures that A contains infinitely many arithmetic progressions

of any length k, as one first locates a "quadratic period" L with $\{\alpha L^2\}$ small, and then for suitable $n \in A$ one locates a much smaller "linear period" M with $\{\alpha LMn\}$ small. If this is done correctly, the progression $n, n + LM, \ldots, n + (k-1)LM$ will be completely contained in A. The same arguments also extend to a more general class of quadratically structured sets, such as the "2-step nilperiodic" set $A = \{n : \{\lfloor\sqrt{2}n\rfloor\sqrt{3}n \leq \delta\}$, where $\lfloor x\rfloor$ is the greatest integer function.

- (Random subsets of structured sets) Continuing the previous example $A := \{n : \{\alpha n^2\} \leq \delta\}$, let A' be a random subset of A with each $n \in A$ lying in A' with an independent probability of δ' for some $0 < \delta' < 1$. Then this set A' almost surely has a positive density of $\delta\delta'$ if α is irrational. The set A' almost surely has infinitely many progressions of length k, since A already starts with infinitely many such progressions, and each such progression as a probability of $(\delta')^k$ of also lying in A'. One can generalize this example to random sets $\tilde{A}$ where the events $n \in \tilde{A}$ are independent as n varies, and the probability $\mathbf{P}(n \in \tilde{A})$ is a "quadratically almost periodic" function of n such as $\mathbf{P}(n \in \tilde{A}) = F(\{\alpha n^2\})$ for some nice (e.g. piecewise continuous) function F taking values between 0 and 1; the preceding example is the case where $F(x) := \delta' 1_{x<\delta}$. It is also possible to adapt this argument to (possibly sparse) pseudorandom subsets of structured sets, though one needs to take some care in defining exactly what "pseudorandom" means here.

- (Sets containing random subsets of structured sets) Let A'' be any set which contains the set A' (or $\tilde{A}$) of the previous example. Since A' contains infinitely many progressions of length k, it is trivial that A'' does also.

As the above examples should make clear, the reason for the truth of Szemerédi's theorem is very different in the cases when A is random, and when A is structured. These two cases can then be combined to handle the case when A is (or contains) a large (pseudo-)random subset of a structured set. Each of the proofs of Szemerédi's theorem now hinge on a *structure theorem* which, very roughly speaking, asserts that *every* set of positive density is (or contains) a large pseudorandom subset of a structured set; each of the four proofs obtains a structure theorem of this sort in a different way (and in a very different language). These remarkable structural results – which include the Furstenberg structure theorem and the Szemerédi regularity lemma as examples – are of independent interest (beyond their immediate applications to arithmetic progressions), and have led to many further developments and insights. For instance, in [27] a "weighted" structure theorem (which was in some sense a hybrid of the Furstenberg structure theorem and the Szemerédi regularity lemma) was the primary new ingredient in proving that the primes $P := \{2, 3, 5, 7, \ldots\}$ contained arbitrarily long arithmetic progressions. While that latter claim is ostensibly a number-theoretical result, the method of proof in fact uses surprisingly little from number theory, being much closer in spirit to the proofs of Szemerédi's theorem (and

in fact Szemerédi's theorem is a crucial ingredient in the proof). This can be seen from the fact that the argument in [27] in fact proves the following stronger result:

Theorem 1.2 (Szemerédi's theorem in the primes [27]). *Let A be a subset of the primes P of positive* relative *upper density, thus* $\limsup_{N\to\infty} \frac{|A\cap[-N,N]|}{|P\cap[-N,N]|} > 0$. *Then for any $k \geq 3$, A contains infinitely many arithmetic progressions of length k.*

This result was first established in the $k = 3$ case by Green [22], the key step again being a (Fourier-analytic) structure theorem, this time for subsets of the primes. The arguments used to prove this theorem do not directly address the important question of whether the primes P (or any subset thereof) have any pseudorandomness properties (but see Section 5 below). However, the structure theorem does allow one to (essentially) describe any dense subset of the primes as a (sparse) pseudorandom subset of some unspecified dense set, which turns out to be sufficient (thanks to Szemerédi's theorem) for the purpose of establishing the existence of arithmetic progressions.

There are now several expositions of Theorem 1.2; see for instance [42], [25], [55], [56], [37]. Rather than give yet another exposition of this result, we have chosen to take a broader view, surveying the collection of structural theorems which underlie the proof of such results as Theorem 1.1 and Theorem 1.2. These theorems have remarkably varied contexts – measure theory, ergodic theory, graph theory, hypergraph theory, probability theory, information theory, and Fourier analysis – and can be either qualitative (infinitary) or quantitative (finitary) in nature. However, their *proofs* tend to share a number of common features, and thus serve as a kind of "Rosetta stone" connecting these various fields. Firstly, for a given class of objects, one quantifies what it means for an object to be "(pseudo-)random" and an object to be "structured". Then, one establishes a *dichotomy between randomness and structure*, which typically looks something like this:

> *If an object is not* (*pseudo-*)*random, then it* (*or some non-trivial component of it*) *correlates with a structured object.*

One can then iterate this dichotomy repeatedly (e.g. via a stopping time argument, or by Zorn's lemma), to extract out all the correlations with structured objects, to obtain a *weak structure theorem* which typically looks as follows:

> *If A is an arbitrary object, then A* (*or some non-trivial component of A*) *splits as the sum of a structured object, plus a pseudorandom error.*

In many circumstances, we need to improve this result to a *strong structure theorem*:

> *If A is an arbitrary object, then A* (*or some non-trivial component of A*) *splits as the sum of a structured object, plus a small error, plus a* ***very*** *pseudorandom error.*

When one is working in an infinitary (qualitative) setting rather than a finitary (quantitative) one – which is for instance the case in the ergodic theory approach – one works instead with an *asymptotic structure theorem*:

If A is an arbitrary object, then A (or some non-trivial component of A) splits as the sum of a "compact" object (the limit of structured objects), plus an ***infinitely*** *pseudorandom error.*

The reason for the terminology "compact" to describe the limit of structured objects is in analogy to how a compact operator can be viewed as the limit of finite rank operators; see [12] for further discussion.

In many applications, the small or pseudorandom errors in these structure theorems are negligible, and one then reduces to the study of structured objects. One then exploits the structure of these objects to conclude the desired application.

Our focus here is on the structure theorems related to Szemerédi's theorem and related results such as Theorem 1.2; we will not have space to describe all the generalizations and refinements of these results here. However, these types of structural theorems appear in other contexts also, for instance the Komlós subsequence principle [40] in probability theory. The Lebesgue decomposition of a spectral measure into pure point, singular continuous, and absolutely continuous spectral components can also be viewed as a structure theorem of the above type. Also, the stopping time arguments which underlie the structural theorems here are also widely used in harmonic analysis, in particular obtaining fundamental decompositions such as the Calderón–Zygmund decomposition or the atomic decomposition of Hardy spaces (see e.g. [52]), as well as the tree selection arguments used in multilinear harmonic analysis (see e.g. [43]). It may be worth investigating whether there are any concrete connections between these disparate structural theorems.

2. Ergodic theory

We now illustrate the above general strategy in a number of contexts, beginning with the ergodic theory approach to Szemerédi's theorem, where the dichotomy between structure and randomness is particularly clean and explicit. Informally speaking, the ergodic theory approach seeks to understand the set A of integers by analyzing the asymptotic correlations of the shifts $A+n := \{a+n : a \in A\}$ (or of various asymptotic averages of these shifts), and treating these shifts as occurring on an abstract measure space. More formally, let X be a measure space with probability measure $d\mu$, and let $T\colon X \to X$ be a bijection such that T and T^{-1} are both measure-preserving maps. The associated shift operator $T\colon f \mapsto f \circ T^{-1}$ is thus a unitary operator on the Hilbert space $L^2(X)$ of complex-valued square-integrable functions with the usual inner product $\langle f, g\rangle := \int_X f\bar{g}\, d\mu$. A famous transference result known as the *Furstenberg correspondence principle*[1] (see [11], [13], [12]) shows that Szemerédi's

[1] Morally speaking, to deduce Szemerédi's theorem from Furstenberg's theorem, one takes X to be the integers $\mathbb{Z}$, T to be the standard shift $n \mapsto n+1$, and μ to be the density $\mu(A) = \lim_{N\to\infty} \frac{|A\cap[-N,N]|}{|[-N,N]|}$. This does not quite work because not all sets A have a well-defined density, however additional arguments (e.g. using the Hahn–Banach theorem) can fix this problem.

theorem is then equivalent to

Theorem 2.1 (Furstenberg recurrence theorem [11]). *Let X and T be as above, and let $f \in L^\infty(X)$ be any bounded non-negative function with $\int_X f\, d\mu > 0$. Then for any $k \geq 1$ we have*

$$\liminf_{N\to\infty} \mathbf{E}_{1\leq n\leq N} \int_X f T^n f \dots T^{(k-1)n} f\, d\mu > 0.$$

Here and in the sequel we use $\mathbf{E}_{n\in I} a_n$ as a shorthand for the average $\frac{1}{|I|}\sum_{n\in I} a_n$.

When $k = 2$ this is essentially the Poincaré recurrence theorem; by using the von Neumann ergodic theorem one can also show that the limit exists (thus the lim inf can be replaced with a lim). The $k = 3$ case can be proved by the following argument, as observed in [12]. We need to show that

$$\liminf_{N\to\infty} \mathbf{E}_{1\leq n\leq N} \int_X f T^n f T^{2n} f\, d\mu > 0 \tag{1}$$

whenever f is bounded, non-negative, and has positive integral.

The first key observation is that any sufficiently pseudorandom component of f will give a negligible contribution to (1) and can be dropped. More precisely, let us call f is *linearly pseudorandom* (or *weakly mixing*) with respect to the shift T if we have

$$\lim_{N\to\infty} \mathbf{E}_{1\leq n\leq N} |\langle T^n f, f\rangle|^2 = 0. \tag{2}$$

Such functions are negligible for the purpose of computing averages such as those in (1); indeed, if at least one of $f, g, h \in L^\infty(X)$ is linearly pseudorandom, then an easy application of van der Corput's lemma (which in turn is an application of Cauchy–Schwarz) shows that

$$\lim_{N\to\infty} \mathbf{E}_{1\leq n\leq N} \int_X f T^n g T^{2n} h\, d\mu = 0.$$

We shall refer to these types of results – that pseudorandom functions are negligible when averaged against other functions – as *generalized von Neumann theorems*.

In view of this generalized von Neumann theorem, one is now tempted to "quotient out" all the pseudorandom functions and work with a reduced class of "structured" functions. In this particular case, it turns out that the correct notion of structure is that of a *linearly almost periodic function*, which are in turn generated by the *linear eigenfunctions* of T. To make this more precise, we need the following dichotomy:

Lemma 2.2 (Dichotomy between randomness and structure). *Suppose that $f \in L^\infty(X)$ is not linearly pseudorandom. Then there exists an linear eigenfunction $g \in L^\infty(X)$ of T (thus $Tg = \lambda g$ for some $\lambda \in \mathbb{C}$) such that $\langle f, g\rangle \neq 0$.*

Remark 2.3. Observe that if g is a linear eigenfunction of T with $Tg = \lambda g$, then $|\lambda| = 1$ and $\lim_{N\to\infty} \mathbf{E}_{1\leq n\leq N} \int_X gT^n\bar{g}^2T^{2n}g\, d\mu = \int_X |g|^4$. Thus linear eigenfunctions can and do give nontrivial contributions to the expression in (1). One can view Lemma 2.2 as a converse to this observation.

The proof of this lemma follows easily from spectral theory and is omitted here. It has the following consequence. Let $\mathcal{Z}_1$ be the σ-algebra generated by all the eigenfunctions of T, this is known as the *Kronecker factor* of X, and roughly speaking encapsulates all the "linear structure" in the measure preserving system. Given every function $f \in L^2(X)$, we have the decomposition $f = f_{U^\perp} + f_U$, where $f_{U^\perp} := \mathbf{E}(f|\mathcal{Z}_1)$ is the conditional expectation of f with respect to the σ-algebra $\mathcal{Z}_1$ (i.e. the orthogonal projection from $L^2(X)$ to the $\mathcal{Z}_1$-measurable functions). By construction, $f_U := f - \mathbf{E}(f|\mathcal{Z}_1)$ is orthogonal to every eigenfunction of T, and is hence linearly pseudorandom by Lemma 2.2. In particular, we have established

Proposition 2.4 (Asymptotic structure theorem). *Let f be bounded and non-negative, with positive integral. Then we can split*[2] *$f = f_{U^\perp} + f_U$, where $f_{U^\perp}$ is bounded, non-negative, and $\mathcal{Z}_1$-measurable* (*and thus approximable in L^2 to arbitrary accuracy by finite linear combinations of linear eigenfunctions*)*, with positive integral, and f_U is linearly pseudorandom.*

This result is closely related to the Koopman–von Neumann theorem in ergodic theory. In the language of the introduction, it asserts (very roughly speaking) that any set A of integers can be viewed as a (linearly) pseudorandom set where the "probability" $f_{U^\perp}(n)$ that a given element n lies in A is a (linearly) almost periodic function of n.

Note that the linearly pseudorandom component f_U of f gives no contribution to (1), thanks to the generalized von Neumann theorem. Thus we may freely replace f by $f_{U^\perp}$ if desired; in other words, for the purposes of proving (1) we may assume without loss of generality that f is measurable with respect to the Kronecker factor $\mathcal{Z}_1$. In the notation of [14], we have just shown that the Kronecker factor is a *characteristic factor* for the recurrence in (1). (In fact it is essentially the universal factor for this recurrence, see [64], [39] for further discussion.)

We have reduced the proof of (1) to the case when f is structured, in the sense of being measurable in $\mathcal{Z}_2$. There are two ways to obtain the desired "structured recurrence" result. Firstly there is a "soft" approach, in which one observes that every $\mathcal{Z}_1$-measurable square-integrable function f is *almost periodic*, in the sense that for any $\varepsilon > 0$ there exists a set of integers n of positive density such that $T^n f$ is within ε of f in $L^2(X)$; from this it is easy to show that $\int_X fT^n fT^{2n} f\, d\mu$ is close to $\int_X f^3$ for a set of integers n of positive density, which implies (1). This almost periodicity can be verified by first checking it for polynomial combinations of linear eigenfunctions, and then extending by density arguments. There is also a "hard"

[2] The notation is from [27]; the subscript U stands for "Gowers uniform" (pseudorandom), and $U^\perp$ for "Gowers anti-uniform" (structured).

approach, in which one obtains algebraic and topological control on the Kronecker factor $\mathcal{Z}_1$. In fact, from a spectral analysis of T one can show that $\mathcal{Z}_1$ is the inverse limit of a sequence of σ-algebras, on each of which the shift T is isomorphic to a shift $x \mapsto x + \alpha$ on a compact abelian Lie group G. This gives a very concrete description of the functions f which are measurable in the Kronecker factor, and one can establish (1) by a direct argument similar to that used in the introduction for linearly structured sets. This "hard" approach gives a bit more information; for instance, it can be used to show that the limit in (1) actually converges, so one can replace the lim inf by a lim.

It turns out that these arguments extend (with some non-trivial effort) to the case of higher k. For sake of exposition let us just discuss the $k = 4$ case, though most of the assertions here extend to higher k. We wish to prove that

$$\liminf_{N\to\infty} \mathbf{E}_{1\leq n\leq N} \int_X f T^n f T^{2n} f T^{3n} f \, d\mu > 0 \tag{3}$$

whenever f is bounded, non-negative, and has positive integral. Here, it turns out that we must strengthen the notion of pseudorandomness (and hence generalize the notion of structure); linear pseudorandomness is no longer sufficient to imply negligibility. For instance, let f be a *quadratic eigenfunction*, in the sense that $Tf = \lambda f$, where λ is no longer constant but is itself a linear eigenfunction, thus $T\lambda = c\lambda$ for some constant c. As an example, if $X = (\mathbb{R}/\mathbb{Z})^2$ with the skew shift $T(x, y) = (x+\alpha, y+x)$ for some fixed number α, then the function $f(x, y) = e^{2\pi i y}$ is a quadratic eigenfunction but not a linear one. Typically such quadratic eigenfunctions will be linearly pseudorandom, but if $|\lambda| = |c| = 1$ (which is often the case) then we have the identity

$$\mathbf{E}_{1\leq n\leq N} \int_X f T^n \bar{f}^3 T^{2n} f^3 T^{3n} \bar{f} \, d\mu = \int_X |f|^8 \, d\mu \tag{4}$$

and so we see that these functions can give non-trivial contributions to expressions such as (1). The correct notion of pseudorandomness is now *quadratic pseudorandomness*, by which we mean that

$$\lim_{H\to\infty} \lim_{N\to\infty} \mathbf{E}_{1\leq n\leq N} \mathbf{E}_{1\leq h\leq H} |\langle T^h f \bar{f}, T^n (T^h f \bar{f})\rangle|^2 = 0.$$

In other words, f is quadratically pseudorandom if and only if $T^h f \bar{f}$ is asymptotically linearly pseudorandom on the average as $h \to \infty$. Several applications of van der Corput's lemma give a generalized von Neumann theorem, asserting that

$$\lim_{N\to\infty} \mathbf{E}_{1\leq n\leq N} \int_X f_0 T^n f_1 T^{2n} f_2 T^{3n} f_3 \, d\mu = 0$$

whenever f_0, f_1, f_2, f_3 are bounded functions with at least one function quadratically pseudorandom.

One would now like to construct a factor $\mathcal{Z}_2$ (presumably larger than the Kronecker factor $\mathcal{Z}_2$) which will play the role of the Kronecker factor for the average (3); in particular, we would like a statement of the form

Lemma 2.5 (Dichotomy between randomness and structure). *Suppose that $f \in L^\infty(X)$ is not linearly pseudorandom. Then there exists a $\mathcal{Z}$-measurable function $g \in L^\infty(X)$ such that $\langle f, g\rangle \neq 0$.*

which would imply[3]

Proposition 2.6 (Asymptotic structure theorem). *Let f be bounded and non-negative, with positive integral. Then we can split $f = f_{U^\perp} + f_U$, where $f_{U^\perp}$ is bounded, non-negative, and $\mathcal{Z}_2$-measurable, with positive integral,and f_U is quadratically pseudorandom.*

This reduces the proof of (3) to that of $\mathcal{Z}_2$-measurable f. The existence of such a factor $\mathcal{Z}_2$ (which would be a *characteristic factor* for this average) is trivial to construct, as we could just take $\mathcal{Z}_2$ to be the entire σ-algebra, and it is in fact easy (via Zorn's lemma) to show the existence of a "best" such factor, which embed into all other characteristic factors for this average (see [64]). Of course, for the concept of characteristic factor to be useful we would like $\mathcal{Z}_2$ to be smaller than this, and specifically for the functions in this factor to enjoy some structural properties. An obvious guess for $\mathcal{Z}_2$ would be the σ-algebra generated by all the linear and quadratic eigenfunctions, but this factor turns out to be a bit too small (see [14]; this is related to the example of the 2-step nilperiodic set in the introduction). A more effective candidate for $\mathcal{Z}_2$, analogous to the "soft" description of the Kronecker factor, is the space of all "quadratically almost periodic functions". This concept is a bit tricky to define rigorously (see e.g. [13], [12], [54]), but roughly speaking, a function f is linearly almost periodic if the orbit $\{T^n f : n \in \mathbb{Z}\}$ is precompact in $L^2(X)$ viewed as a Hilbert space, while a function f is quadratically almost periodic if the orbit is precompact in $L^2(X)$ viewed as a Hilbert *module* over the Kronecker factor $L^\infty(\mathcal{Z}_1)$; this can be viewed as a matrix-valued (or more precisely compact operator-valued) extension of the concept of a quadratic eigenfunction. Another rough definition is as follows: a function f is linearly almost periodic if $T^n f(x)$ is close to $f(x)$ for many constants n, whereas a function f is quadratically almost periodic if $T^{n(x)} f(x)$ is close to $f(x)$ for a function $n(x)$ which is itself linearly almost periodic. It turns out that with this "soft" proposal for $\mathcal{Z}_2$, it is easy to prove Lemma 2.5 and hence Proposition 2.6, essentially by obtaining a "relative" version of the proof of Lemma 2.2. The derivation of (3) in this soft factor is slightly tricky though, requiring either van der Waerden's theorem, or the color focusing argument used to prove van der Waerden's theorem; see [11], [13], [12], [54].

More recently, a more efficient "hard" factor $\mathcal{Z}_2$ was constructed by Conze–Lesigne [7], Furstenberg–Weiss [14], and Host–Kra [38]; the analogous factors for

[3] One can generalize this structure theorem to obtain similar characteristic factors $\mathcal{Z}_3$, $\mathcal{Z}_4$ for cubic pseudorandomness, quartic pseudorandomness, etc. Applying Zorn's lemma, one eventually obtains the *Furstenberg structure theorem*, which decomposes any measure preserving system as a weakly mixing extension of a distal system, and thus decomposes any function as a distal function plus an "infinitely pseudorandom" error; see [13]. However this decomposition is not the most "efficient" way to prove Szemerédi's theorem, as the notion of pseudorandomness is too strong, and hence the notion of structure too general. It does illustrate however that one does have considerable flexibility in where to draw the line between randomness and structure.

higher k are more difficult to construct, but this was achieved by Host–Kra in [39], and also subsequently by Ziegler [64]. This factor yields more precise information, including convergence of the limit in (3). Here, the concept of a 2-*step nilsystem* is used to define structure. A 2-step nilsystem is a compact symmetric space G/Γ, with G a 2-step nilpotent Lie group and Γ is a closed subgroup, together with a shift element $\alpha \in G$, which generates a shift $T(x\Gamma) := \alpha x\Gamma$. The factor $\mathcal{Z}_2$ constructed in these papers is then the inverse limit of a sequence of σ-algebras, on which the shift is equivalent to a 2-step nilsystem. This should be compared with the "hard" description of the Kronecker factor, which is the 1-step analogue of the above result. Establishing the bound (3) then reduces to the problem of understanding the structure of arithmetic progressions $x\Gamma$, $\alpha x\Gamma$, $\alpha^2 x\Gamma$, $\alpha^3 x\Gamma$ on the nilsystem, which can be handled by algebraic arguments, for instance using the machinery of Hall–Petresco sequences [44].

The ergodic methods, while non-elementary and non-quantitative (though see [54]), have proven to be the most powerful and flexible approach to Szemerédi's theorem, leading to many generalizations and refinements. However, it seems that a purely "soft" ergodic approach is not quite capable by itself of extending to the primes as in Theorem 1.2, though it comes tantalizingly close. In particular, one can use Theorem 2.1 and a variant of the Furstenberg correspondence principle to establish Theorem 1.2 when the set of primes P is replaced by a random subset $\tilde{P}$ of the positive integers, with $n \in \tilde{P}$ with independent probability $1/\log n$ for $n > 1$; see [60]. Roughly speaking, if A is a subset of $\tilde{P}$, the idea is to construct an abstract measure-preserving system generated by a set $\tilde{A}$, in which $\mu(T^{n_1}\tilde{A} \cap \cdots \cap T^{n_k}\tilde{A})$ is the normalized density of $(A+n_1) \cap \cdots \cap (A+n_k)$ for any $n_1, \ldots, n_k$. Unfortunately, this approach requires the ambient space $\tilde{P}$ to be extremely pseudorandom and does not seem to extend easily to the primes.

3. Fourier analysis

We now turn to a more quantitative approach to Szemerédi's theorem, based primarily on Fourier analysis and arithmetic combinatorics. Here, one analyzes a set of integers A finitarily, truncating to a finite setting such as the discrete integral $\{1, \ldots, N\}$ or the cyclic group $\mathbb{Z}/N\mathbb{Z}$, and then testing the correlations of A with linear phases such as $n \mapsto e^{2\pi ikn/N}$, quadratic phases $n \mapsto e^{2\pi ikn^2/N}$, or similar objects. This approach has lead to the best known bounds on Szemerédi's theorem, though it has not yet been able to handle many of the generalizations of this theorem that can be treated by ergodic or graph-theoretic methods. In analogy with the ergodic arguments, the $k = 3$ case of Szemerédi's theorem can be handled by linear Fourier analysis (as was done by Roth [50]), while the $k = 4$ case requires quadratic Fourier analysis (as was done by Gowers [19]), and so forth for higher order k (see [20]). The Fourier analytic approach seems to be closely related to the theory of the "hard" characteristic

factors discovered in the ergodic theory arguments, although the precise nature of this relationship is still being understood.

It is convenient to work in a cyclic group $\mathbb{Z}/N\mathbb{Z}$ of prime order. It can be shown via averaging arguments (see [63]) that Szemerédi's theorem is equivalent to the following quantitative version:

Theorem 3.1 (Szemerédi's theorem, quantitative version). *Let $N > 1$ be a large prime, let $k \geq 3$, and let $0 < \delta < 1$. Let $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ be a function with $0 \leq f(x) \leq 1$ for all $x \in \mathbb{Z}/N\mathbb{Z}$ and $\mathbf{E}_{x\in\mathbb{Z}/N\mathbb{Z}} f(x) \geq \delta$. Then we have*

$$\mathbf{E}_{x,r\in\mathbb{Z}/N\mathbb{Z}} f(x)T^r f(x)\dots T^{(k-1)r} f(x) \geq c(k,\delta)$$

for some $c(k,\delta) > 0$ depending only on k and δ, where $T^r f(x) := f(x+r)$ is the shift operator on $\mathbb{Z}/N\mathbb{Z}$.

We remark that the Fourier-analytic arguments in Gowers [20] give the best known lower bounds on $c(k,\delta)$, namely $c(k,\delta) > 2^{-2^{1/\delta^{c_k}}}$ where $c_k := 2^{2^{k+9}}$. In the $k = 3$ case it is known that $c(3,\delta) \geq \delta^{C/\delta^2}$ for some absolute constant C, see [5]. A conjecture of Erdős and Turán [8] is roughly equivalent to asserting that $c(k,\delta) > e^{-C_k/\delta}$ for some C_k. In the converse direction, an example of Behrend shows that $c(3,\delta)$ cannot exceed $e^{c\log^2(1/\delta)}$ for some small absolute constant c, with similar results for higher values of k; in particular, $c(k,\delta)$ cannot be as large as any fixed power of δ. This already rules out a number of elementary approaches to Szemerédi's theorem and suggests that any proof must involve some sort of iterative argument.

Let us first describe (in more "modern" language) Roth's original proof [50] of Szemerédi's theorem in the $k = 3$ case. We need to establish a bound of the form

$$\mathbf{E}_{x,r\in\mathbb{Z}/N\mathbb{Z}} f(x)T^r f(x)T^{2r} f(x) \geq c(3,\delta) > 0 \tag{5}$$

when f takes values between 0 and 1 and has mean at least δ. As in the ergodic argument, we first look for a notion of pseudorandomness which will ensure that the average in (5) is negligible. It is convenient to introduce the *Gowers $U^2(\mathbb{Z}/N\mathbb{Z})$ uniformity norm* by the formula

$$\|f\|^4_{U^2(\mathbb{Z}/N\mathbb{Z})} := \mathbf{E}_{n\in\mathbb{Z}/N\mathbb{Z}}|\mathbf{E}_{x\in\mathbb{Z}/N\mathbb{Z}} T^n f(x)\,\overline{f(x)}|^2,$$

and informally refer to f as *linearly pseudorandom* (or *linearly Gowers-uniform*) if its U^2 norm is small; compare this with (2). The U^2 norm is indeed a norm; this can be verified either by several applications of the Cauchy–Schwarz inequality, or via the Fourier identity

$$\|f\|^4_{U^2(\mathbb{Z}/N\mathbb{Z})} = \sum_{\xi\in\mathbb{Z}/N\mathbb{Z}} |\hat{f}(\xi)|^4, \tag{6}$$

where $\hat{f}(\xi) := \mathbf{E}_{x\in\mathbb{Z}/N\mathbb{Z}} f(x)e^{-2\pi i x\xi/N}$ is the usual Fourier transform. Some further applications of Cauchy–Schwarz (or Plancherel's theorem and Hölder's inequality)

yields the generalized von Neumann theorem

$$|\mathbf{E}_{x,r\in\mathbb{Z}/N\mathbb{Z}} f_0(x)T^r f_1(x)T^{2r} f_2(x)| \leq \min_{j=0,1,2} \|f_j\|_{U^2(\mathbb{Z}/N\mathbb{Z})} \tag{7}$$

whenever f_0, f_1, f_2 are bounded in magnitude by 1. Thus, as before, linearly pseudorandom functions give a small contribution to the average in (5), though now that we are in a finitary setting the contribution does not vanish completely.

The next step is to establish a dichotomy between linear pseudorandomness and some sort of usable structure. From (6) and Plancherel's theorem we easily obtain the following analogue of Lemma 2.2:

Lemma 3.2 (Dichotomy between randomness and structure). *Suppose that the function $f: \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ is bounded in magnitude by 1 with $\|f\|_{U^2(\mathbb{Z}/N\mathbb{Z})} \geq \eta$ for some $0 < \eta < 1$. Then there exists a linear phase function $\phi: \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}/\mathbb{Z}$ (thus $\phi(x) = \xi x/N + c$ for some $\xi \in \mathbb{Z}/N\mathbb{Z}$ and $c \in \mathbb{R}/\mathbb{Z}$) such that $|\mathbf{E}_{x\in\mathbb{Z}/N\mathbb{Z}} f(x)e^{-2\pi i\phi(x)}| \geq \eta^2$.*

The next step is to iterate this lemma to obtain a suitable structure theorem. There are two slightly different ways to do this. Firstly there is the original *density increment argument* approach of Roth [50], which we sketch as follows. It is convenient to work on a discrete interval $[1, N/3]$, which we identify with a subset of $\mathbb{Z}/N\mathbb{Z}$ in the obvious manner. Let $f: [1, N/3] \to \mathbb{R}$ be a non-negative function bounded in magnitude by 1, and let η be a parameter to be chosen later. If $f - \mathbf{E}_{1\leq x\leq N/3} f(x)$ is not linearly pseudorandom, in the sense that $\|f - \mathbf{E}_{1\leq x\leq N/3} f(x)\|_{U^2(\mathbb{Z}/N\mathbb{Z})} \geq \eta$, then we apply Lemma 3.2 to obtain a correlation with a linear phase ϕ. An easy application of the Dirichlet approximation theorem then shows that one can partition $[1, N/3]$ into arithmetic progressions (of length roughly $\eta^2\sqrt{N}$) on which ϕ is essentially constant (fluctuating by at most $\eta^2/100$, say). A pigeonhole argument (exploiting the fact that $f - \mathbf{E}_{1\leq x\leq N/3} f(x)$ has mean zero) then shows that on one of these progressions, say P, f has significantly higher density than on the average, in the sense that $\mathbf{E}_{x\in P} f(x) \geq \mathbf{E}_{x\in\mathbb{Z}/N\mathbb{Z}} f(x) + \eta^2/100$. One can then apply an affine transformation to convert this progression P into another discrete interval $\{1, \ldots, N'/3\}$, where N' is essentially the square root of N. One then iterates this argument until linear pseudorandomness is obtained (using the fact that the density of f cannot increase beyond 1), and one eventually obtains

Theorem 3.3 (Local structure theorem). *Let $f: [1, N/3] \to \mathbb{R}$ be a non-negative function bounded by 1, and let $\eta > 0$. Then there exists a progression P in $[1, N/3]$ of length at least $c(\eta)N^{c(\eta)}$ for some $c(\eta) > 0$, on which we have the splitting $f = f_{U^\perp} + f_U$, where $f_U^\perp := \mathbf{E}_{x\in P} f(x) \geq \mathbf{E}_{1\leq x\leq N/3} f(x)$ is the mean of f on P, and f_U is linearly pseudorandom in the sense that*

$$\|f_U\|_{U^2(\mathbb{Z}/M\mathbb{Z})} \leq \eta$$

where we identify P with a subset of a cyclic group $\mathbb{Z}/M\mathbb{Z}$ of cardinality $M \approx 3|P|$ in the usual manner.

More informally, any function will contain an arithmetic progression P of significant size on which f can be decomposed into a non-trivial structured component $f_{U^\perp}$ and a pseudorandom component f_U. In the language of the introduction, it is essentially saying that any dense set A of integers will contain components which are dense pseudorandom subsets of long progressions. Once one has this theorem, it is an easy matter to establish Szemerédi's theorem in the $k = 3$ case. Indeed, if $A \subseteq \mathbb{Z}$ has upper density greater than δ, then we can find arbitrarily large primes N such that $|A \cap [1, N/3]| \geq \delta N/3$. Applying Theorem 3.3 with $\eta := \delta^3/100$, and f equal to the indicator function of $A \cap [1, N/3]$, we can find a progression P in $\{1, \ldots, N/3\}$ of length at least $c(\delta)N^{c(\delta)}$ on which $\mathbf{E}_{x\in P} f(x) \geq \delta$ and $f - \mathbf{E}_{x\in P} f(x)$ is linearly pseudorandom in the sense of Theorem 3.3. It is then an easy matter to apply the generalized von Neumann theorem to show that $A \cap P$ contains many arithmetic progressions of length three (in fact it contains $\gg \delta^3|P|^3$ such progressions). Letting N (and hence $|P|$) tend to infinity we obtain Szemerédi's theorem in the $k = 3$ case. An averaging argument of Varnavides [63] then yields the more quantitative version in Theorem 3.1 (but with a moderately bad bound for $c(3, \delta)$, namely $c(3, \delta) = 2^{-2^{C/\delta^C}}$ for some absolute constant C).

A more refined structure theorem was given in [23] (see also [35]), which was termed an "arithmetic regularity lemma" in analogy with the Szemerédi regularity lemma which we discuss in the next section. That theorem has similar hypotheses to Theorem 3.3, but instead of constructing a single progression on P on which one has pseudorandomness, one partitions $[1, N/3]$ into *many* long progressions[4], where on most of which the function f becomes linearly pseudorandom (after subtracting the mean). A related structure theorem (with a more "ergodic" perspective) was also given in [56]. Here we give an alternate approach based on Fourier expansion and the pigeonhole principle. Observe that for any $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ and any threshold λ we have the Fourier decomposition $f = f_{U^\perp} + f_U$, where the "structured" component $f_{U^\perp} := \sum_{\xi:|\hat{f}(\xi)|\geq\lambda} \hat{f}(\xi)e^{2\pi ix\xi/N}$ contains all the significant Fourier coefficients, and the "pseudorandom" component $f_U := \sum_{\xi:|\hat{f}(\xi)|\leq\lambda} \hat{f}(\xi)e^{2\pi ix\xi/N}$ contains all the small Fourier coefficients. Using Plancherel's theorem one can easily establish

Theorem 3.4 (Weak structure theorem). *Let $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ be a function bounded in magnitude by* 1*, and let $0 < \lambda < 1$. Then we can split $f = f_{U^\perp} + f_U$, where $f_{U^\perp}$ is the linear combination of at most $O(1/\lambda^2)$ linear phase functions $x \mapsto e^{2\pi ix\xi/N}$, and f_U is linearly pseudorandom in the sense that $\|f_U\|_{U^2(\mathbb{Z}/N\mathbb{Z})} \leq \lambda$.*

This theorem asserts that an arbitrary bounded function only has a bounded amount of significant linear Fourier-analytic structure; after removing this bounded amount of structure, the remainder is linearly pseudorandom.

This theorem, while simple to state and prove, has two weaknesses which make it unsuitable for such tasks as counting progressions of length three. Firstly, even

[4]Actually, for technical reasons it is more efficient to replace the notion of an arithmetic progression by a slightly different object known as a *Bohr set*; see [23], [35] for details.

though f is bounded by 1, the components $f_{U^\perp}$, f_U need not be. Related to this, if f is non-negative, there is no reason why $f_{U^\perp}$ should be non-negative also. Secondly, the pseudorandomness control on f_U is not very good when compared against the complexity of $f_{U^\perp}$ (i.e. the number of linear exponentials needed to describe $f_{U^\perp}$). In practice, this means that any control one obtains on the structured component of f will be dominated by the errors one has to concede from the pseudorandom component. Fortunately, both of these defects can be repaired, the former by a Fejér summation argument, and the latter by a pigeonhole argument (which introduces a second error term f_S, which is small in L^2 norm). More precisely, we have

Theorem 3.5 (Strong structure theorem). *Let $f: \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ be a non-negative function bounded by* 1*, and let $0 < \varepsilon < 1$. Let $F: \mathbb{N} \to \mathbb{N}$ be an arbitrary increasing function (e.g. $F(n) = 2^{2^n}$). Then there exists an integer $T = O_{F,\varepsilon}(1)$ and a decomposition $f = f_{U^\perp} + f_S + f_U$, where $f_{U^\perp}$ is the linear combination of at most T linear phase functions, f_U is linearly pseudorandom in the sense that $\|f_U\|_{U^2(\mathbb{Z}/N\mathbb{Z})} = O(1/F(T))$, and f_S is small in the sense that $\|f_S\|_{L^2(\mathbb{Z}/N\mathbb{Z})} := (\mathbf{E}_{n\in\mathbb{Z}/N\mathbb{Z}}|f_S(n)|^2)^{1/2} = O(\varepsilon)$. Furthermore, $f_{U^\perp}$, f_U are bounded in magnitude by* 1*. Also, $f_{U^\perp}$ and $f_{U^\perp} + f_S$ are non-negative with the same mean as f.*

This theorem can be proven by adapting arguments from [26], [35], or [56]; we omit the details. Note that we have the freedom to set the growth function F arbitrarily fast in the above proposition; this corresponds roughly speaking to the fact that in the ergodic counterpart to this structure theorem (Proposition 2.4) the pseudorandom error f_U has asymptotically *vanishing* Gowers U^2 norm. One can view $f_{U^\perp}$ as a "coarse" Fourier approximation to f, and $f_{U^\perp} + f_S$ as a "fine" Fourier approximation to f; this perspective links this proposition with the graph regularity lemmas that we discuss in the next section.

Theorem 3.5 can be used to deduce the structure theorems in [23], [56], [35], while a closely related result was also established in [4]. It can also be used to directly derive the $k = 3$ case of Theorem 3.1, as follows. Let f be as in that proposition, and let $\varepsilon := \delta^3/100$. We apply Theorem 3.5 to decompose $f = f_{U^\perp} + f_S + f_U$. Because $f_{U^\perp}$ has only T Fourier exponentials, it is easy to see that $f_{U^\perp}$ is almost periodic, in the sense that $\|T^n f_{U^\perp} - f_{U^\perp}\|_{L^2(\mathbb{Z}/N\mathbb{Z})} \leq \varepsilon$ for at least $\sigma(\varepsilon, T)N$ values of $n \in \mathbb{Z}/N\mathbb{Z}$, for some $\sigma(\varepsilon, T) > 0$. For such values of n, one can easily verify that

$$\mathbf{E}_{x\in\mathbb{Z}/N\mathbb{Z}} f_{U^\perp}(x) T^n f_{U^\perp}(x) T^{2n} f_{U^\perp}(x) \geq \delta^3/2.$$

Because f_S is small, we can also deduce that

$$\mathbf{E}_{x\in\mathbb{Z}/N\mathbb{Z}} (f_{U^\perp} + f_S)(x) T^n (f_{U^\perp} + f_S)(x) T^{2n} (f_{U^\perp} + f_S)(x) \geq \delta^3/4$$

for these values of n. Averaging in n (and taking advantage of the non-negativity of $f_{U^\perp} + f_S$) we conclude that

$$\mathbf{E}_{x,n\in\mathbb{Z}/N\mathbb{Z}} (f_{U^\perp} + f_S)(x) T^n (f_{U^\perp} + f_S)(x) T^{2n} (f_{U^\perp} + f_S)(x) \geq \delta^3 \sigma(\varepsilon, T)/4.$$

Adding in the pseudorandom error f_U using the generalized von Neumann theorem (7), we conclude that

$$\mathbf{E}_{x,n\in\mathbb{Z}/N\mathbb{Z}} f(x)T^n f(x)T^{2n} f(x) \geq \delta^3\sigma(\varepsilon, T)/4 - O(1/F(T)).$$

If we choose F to be sufficiently rapidly growing depending on δ and ε, we can absorb the error term in the main term and conclude that

$$\mathbf{E}_{x,n\in\mathbb{Z}/N\mathbb{Z}} f(x)T^n f(x)T^{2n} f(x) \geq \delta^3\sigma(\varepsilon, T)/8.$$

Since $T = O_{F,\varepsilon}(1) = O_\delta(1)$, we obtain the $k = 3$ case of Theorem 3.1 as desired.

Roth's original Fourier-analytic argument was published in 1953. But the extension of this Fourier argument to the $k > 3$ case was not achieved until the work of Gowers [19], [20] in 1998. For simplicity we once again restrict attention to the $k = 4$ case, where the theory is more complete. Our objective is to show

$$\mathbf{E}_{x,r\in\mathbb{Z}/N\mathbb{Z}} f(x)T^r f(x)T^{2r} f(x)T^{3r} f(x) \geq c(4,\delta) > 0 \tag{8}$$

whenever f is non-negative, bounded by 1, and has mean at least δ. There are some significant differences between this case and the $k = 3$ case (5). Firstly, linear pseudorandomness is not enough to guarantee that a contribution to (8) is negligible: for instance, if $f(x) := e^{2\pi i\xi x^2/N}$, then

$$\mathbf{E}_{x,r\in\mathbb{Z}/N\mathbb{Z}} f(x)T^r \bar{f}^3(x)T^{2r} f^3(x)T^{3r} \bar{f}(x) = 1$$

despite f being very linearly pseudorandom (the U^2 norm of f is $N^{-1/4}$); compare this example with (4). One must now utilize some sort of "quadratic Fourier analysis" in order to capture the correct concept of pseudorandomness and structure. Secondly, the Fourier-analytic arguments must now be supplemented by some results from arithmetic combinatorics (notably the Balog–Szemerédi theorem, and results related to Freiman's inverse sumset theorem) in order to obtain a usable notion of quadratic structure. Finally, as in the ergodic case, one cannot rely purely on quadratic phase functions such as $e^{2\pi i(\xi x^2+\eta x)/N}$ to generate all the relevant structured objects, and must also consider generalized quadratic objects such as locally quadratic phase functions, 2-step nilsequences (see below), or bracket quadratic phases such as $e^{2\pi i\lfloor\sqrt{2}n\rfloor\sqrt{3}n}$.

Let us now briefly sketch how the theory works in the $k = 4$ case. The correct notion of pseudorandomness is now given by the *Gowers U^3 uniformity norm*, defined by

$$\|f\|^8_{U^3(\mathbb{Z}/N\mathbb{Z})} := \mathbf{E}_{n\in\mathbb{Z}/N\mathbb{Z}} \|T^n f\bar{f}\|^4_{U^2(\mathbb{Z}/N\mathbb{Z})}.$$

This norm measures the extent to which f behaves quadratically; for instance, if $f = e^{2\pi i P(x)/N}$ for some polynomial P of degree k in the finite field $\mathbb{Z}/N\mathbb{Z}$, then one can verify that $\|f\|_{U^3(\mathbb{Z}/N\mathbb{Z})} = 1$ if P has degree at most 2, but (using the Weil

estimates) we have $\|f\|_{U^3(\mathbb{Z}/N\mathbb{Z})} = O_k(N^{-1/16})$ if P has degree $k > 2$. Repeated application of Cauchy–Schwarz then yields the generalized von Neumann theorem

$$|\mathbf{E}_{x,r\in\mathbb{Z}/N\mathbb{Z}} f_0(x)T^r f_1(x)T^{2r} f_2(x)T^{3r} f_3(x)| \leq \min_{0\leq j\leq 3} \|f_j\|_{U^3(\mathbb{Z}/N\mathbb{Z})} \qquad (9)$$

whenever f_0, f_1, f_2, f_3 are bounded in magnitude by 1. The next step is to establish a dichotomy between quadratic structure and quadratic pseudorandomness in the spirit of Lemma 3.2. In the original work of Gowers [19], it was shown that a function which was not quadratically pseudorandom had local correlation with quadratic phases on medium-length arithmetic progressions. This result (when combined with the density increment argument of Roth) was already enough to prove (8) with a reasonable bound on $c(4, \delta)$ (basically of the form $1/\exp(\exp(\delta^{-C}))$); see [19], [20]. Building upon this work, a stronger dichotomy, similar in spirit to Lemma 2.5, was established in [29]. Here, a number of essentially equivalent formulations of quadratic structure were established, but the easiest to state (and the one which generalizes most easily to higher k) is that of a *(basic)* 2*-step nilsequence*, which can be viewed as a notion of "quadratic almost periodicity" for sequences. More precisely, a 2-step nilsequence a sequence of the form $n \mapsto F(T^n x\Gamma)$, where F is a Lipschitz function on a 2-step nilmanifold G/Γ, $x\Gamma$ is a point in this nilmanifold, and T is a shift operator $T: x\Gamma \mapsto \alpha x\Gamma$ for some fixed group element $\alpha \in G$. We remark that quadratic phase sequences such as $n \mapsto e^{2\pi i\alpha n^2}$ are examples of 2-step nilsequences, and generalized quadratics such as $n \mapsto e^{2\pi i\lfloor\sqrt{2}n\rfloor\sqrt{3}n}$ can also be written (outside of sets of arbitrarily small density) as 2-step nilsequences.

Lemma 3.6 (Dichotomy between randomness and structure [29]). *Suppose that* $f: \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ *is bounded in magnitude by* 1 *with* $\|f\|_{U^3(\mathbb{Z}/N\mathbb{Z})} \geq \eta$ *for some* $0 < \eta < 1$. *Then there exists a* 2*-step nilsequence* $n \mapsto F(T^n x\Gamma)$, *where* G/Γ *is a nilmanifold of dimension* $O_\eta(1)$, *and* F *is a bounded Lipschitz function* G/Γ *with Lipschitz constant* $O_\eta(1)$, *such that* $|\mathbf{E}_{1\leq x\leq N} f(x)\overline{F(T^n x\Gamma)}| \geq c(\eta)$ *for some* $c(\eta) > 1$. (*We identify the integers from* 1 *to* N *with* $\mathbb{Z}/N\mathbb{Z}$ *in the usual manner.*)

In fact the nilmanifold G/Γ constructed in [29] is of a very explicit form, being the direct sum of at most $O_\eta(1)$ circles (which are one-dimensional), skew shifts (which are two-dimensional), and Heisenberg nilmanifolds (which are three-dimensional). The dimension $O_\eta(1)$ is in fact known to be polynomial in η, but the best bounds for $c(\eta)$ are currently only exponential in nature. See [29] for further details and discussion.

The proof of Lemma 3.6 is rather lengthy but can be summarized as follows. If f has large U^3 norm, then by definition $T^n f\bar{f}$ has large U^2 norm for many n. Applying Lemma 3.2, this shows that for many n, $T^n f\bar{f}$ correlates with a linear phase function of some frequency $\xi(n)$ (which can be viewed as a kind of "derivative" of the phase of f in the "direction" n). Some manipulations involving the Cauchy–Schwarz inequality then show that $\xi(n)$ contains some additive structure (in that there are many quadruples

n_1, n_2, n_3, n_4 with $n_1 + n_2 = n_3 + n_4$ and $\xi(n_1) + \xi(n_2) = \xi(n_3) + \xi(n_4)$). Methods from additive combinatorics (notably the Balog–Szemerédi(–Gowers) theorem and Freiman's theorem, see e.g. [61]) are then used to "linearize" ξ, in the sense that $\xi(n)$ agrees with a (generalized) linear function of n on a large (generalized) arithmetic progression. One then "integrates" this fact to conclude that f itself correlates with a certain "anti-derivative" of $\xi(n)$, which is a (generalized) quadratic function on this progression. This in turn can be approximated by a 2-step nilsequence. For full details, see [29].

Thus, quadratic nilsequences are the only obstruction to a function being quadratically pseudorandom. This can be iterated to obtain structural results. The following "weak" structural theorem is already quite useful:

Theorem 3.7 (Weak structure theorem [35]). *Let $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{C}$ be a function bounded in magnitude by* 1*, and let $0 < \lambda < 1$. Then we can split $f = f_{U^\perp} + f_U$, where $f_{U^\perp}$ is a* 2*-step nilsequence given by a nilmanifold of dimension $O_\lambda(1)$ and by a bounded Lipschitz function F with Lipschitz constant $O_\lambda(1)$, and f_U is quadratically pseudorandom in the sense that $\|f_U\|_{U^3(\mathbb{Z}/N\mathbb{Z})} \leq \lambda$. Furthermore, $f_{U^\perp}$ is non-negative, bounded by* 1*, and has the same mean as f.*

This is an analogue of Theorem 3.4, and asserts that any bounded function has only a bounded amount of quadratic structure, with the function becoming quadratically pseudorandom once this structure is subtracted. It cannot be proven in quite the same way as in Theorem 3.4, because we have no "quadratic Fourier inversion formula" that decomposes a function neatly into quadratic components (the problem being that there are so many quadratic objects that such a formula is necessarily overdetermined). However, one can proceed by a finitary analogue of the ergodic theory approach, known as an "energy increment argument". In the ergodic setting, one uses all the quadratic objects to create a σ-algebra $\mathcal{Z}_2$, and sets $f_{U^\perp}$ to be the conditional expectation of f with respect to that σ-algebra. In the finitary setting, it turns out to be too expensive to try to use *all* the 2-step nilsequences to create a σ-algebra. However, by adopting a more adaptive approach, selecting only those 2-step nilsequences which have some significant correlation with f (or some component of f), one can obtain the above theorem; we omit the details.

It is likely that quantitative versions of this structure theorem will improve the known bounds on Szemerédi's theorem in the $k = 4$ case; see [32], [33], [34]. A closely related version of this argument was also essential in establishing Theorem 1.2, see Section 5 below.

4. Graph theory

We now turn to the third major line of attack to Szemerédi's theorem, based on graph theory (and hypergraph theory), and which is perhaps the purest embodiment of the strategy of exploiting the dichotomy between randomness and structure. For graphs,

the relevant structure theorem is the *Szemerédi regularity lemma*, which was developed in [53] in the original proof of Szemerédi's theorem, and has since proven to have many further applications in graph theory and computer science; see [41] for a survey. More recently, the analogous regularity lemma for hypergraphs have been developed in [21], [46], [47], [48], [49], [58]. Roughly speaking, these very useful lemmas assert that any graph (binary relation) or hypergraph (higher order relation), no matter how complex, can be modelled effectively as a pseudorandom sub(hyper)graph of a finite complexity (hyper)graph. Returning to the setting of the introduction, the graph regularity lemma would assert that there exists a colouring of the integers into finitely many colours such that relations such as $x - y \in A$ can be viewed approximately as pseudorandom relations, with the "probability" of the event $x - y \in A$ depending only on the colour of x and y.

The strategy of the graph theory approach is to abstract away the arithmetic structure in Szemerédi's theorem, converting the problem to one of finding solutions to an abstract set of equations, which can be modeled by graphs or hypergraphs. As before, we first illustrate this with the simple case of the $k = 3$ case of Szemerédi's theorem, which we will take in the form of Theorem 3.1. For simplicity we specialize to the case when f is the indicator function of a set A (which thus has density at least δ in $\mathbb{Z}/N\mathbb{Z}$); it is easy to see (e.g. by probabilistic arguments) that this special case in fact implies the general case. The key observation is that the problem of locating an arithmetic progression of length three can be recast as the problem of solving three constraints in three unknowns, where each constraint only involves two of the unknowns. Specifically, if $x, y, z \in \mathbb{Z}/N\mathbb{Z}$ solve the system of constraints

$$\begin{array}{cccc} & y & +2z & \in A \\ -x & & +z & \in A \\ -2x & -y & & \in A \end{array} \tag{10}$$

then $y + 2z, -x + z, -2x - y$ is an arithmetic progression of length three in A. Conversely, each such progression comes from exactly N solutions to (10). Thus, it will suffice to show that there are at least $c(3, \delta)N^3$ solutions to (10). Note that we already can construct at least δN^2 "trivial solutions" to (10), in which $y + 2z = -x + z = -2x + y$ is an element of A. Furthermore, these trivial solutions (x, y, z) are "edge-disjoint" in the sense that no two of these solutions share more than one value in common (i.e. if (x, y, z) and (x', y', z') are distinct trivial solutions then at most one of $x = x'$, $y = y'$, $z = z'$ are true). It turns out that these trivial solutions automatically generate a large number of non-trivial solutions to (10) – without using any further arithmetic structure present in these constraints. Indeed, the claim now follows from the following graph-theoretical statement.

Lemma 4.1 (Triangle removal lemma [51]). *For every $0 < \delta < 1$ there exists $0 < \sigma < 1$ with the following property. Let $G = (V, E)$ be an (undirected) graph with $|V| = N$ vertices which contains fewer than σN^3 triangles. Then it is possible to remove $O(\delta N^2)$ edges from G to create a graph G' which contains no triangles whatsoever.*

To see how the triangle removal lemma implies the claim, consider a vertex set V which consists of three copies V_1, V_2, V_3 of $\mathbb{Z}/N\mathbb{Z}$ (so $|V| = 3N$), and consider the tripartite graph $G = (V, E)$ whose edges are of the form

$$E = \{(y, z) \in V_2 \times V_3 : y + 2z \in A\} \cup \{(x, z) \in V_1 \times V_3 : -x + z \in A\} \\ \cup \{(x, y) \in V_1 \times V_2 : -2x - y \in A\}.$$

One can think of G as a variant of the Cayley graph for A. Observe that solutions to (10) are in one-to-one correspondence with triangles in G. Furthermore, the δN^2 trivial solutions to (10) correspond to δN^2 edge-disjoint triangles in G. Thus to delete all the triangles one needs to remove at least δN^2 edges. Applying Lemma 4.1 in the contrapositive (adjusting N, δ, σ by constants such as 3 if necessary), we see that G contains at least σN^3 triangles for some $\sigma = \sigma(\delta) > 0$, and the claim follows.

The only known proof of the triangle removal lemma proceeds by a structure theorem for graphs known as the *Szemerédi regularity lemma*. In order to emphasize the similarities between this approach and the previously discussed approaches, we shall not use the standard formulation of this lemma, but instead use a more recent formulation from [57], [58] (see also [1], [45]), which replaces graphs with functions, and then obtains a structure theorem decomposing such functions into a structured (finite complexity) component, a small component, and a pseudorandom (regular) component. More precisely, we work with functions $f : V \times V \to \mathbb{R}$; this can be thought of as a weighted, directed generalization of a graph on V in which every edge (x, y) is assigned a real-valued weight $f(x, y)$. The first step is to define a notion of pseudorandomness. For graphs, this concept is well understood. There are many equivalent formulations of this concept (see [6]), but we shall adopt one particularly close to the analogous concepts in previous sections, by introducing the *Gowers* $\square^2$ *cube norm* as

$$\|f\|_{\square^2}^4 := \mathbf{E}_{x,y,x',y' \in V} f(x, y) f(x, y') f(x', y) f(x', y');$$

when f is the incidence function of a graph, the right-hand side essentially counts the number of 4-cycles in that graph. Again, one can use the Cauchy–Schwarz inequality to establish that the $\square^2$ norm is indeed a norm; alternatively, one can use spectral theory and observe that the $\square^2$ norm is essentially the Schatten-von Neumann p-norm of f with $p = 4$. We refer to f as *pseudorandom* if its $\square^2$ norm is small. By two applications of Cauchy–Schwarz we have the generalized von Neumann inequality

$$|\mathbf{E}_{x,y,z \in V} f(x, y) g(y, z) h(z, x)| \leq \min(\|f\|_{\square^2}, \|g\|_{\square^2}, \|h\|_{\square^2}) \tag{11}$$

whenever f, g, h are bounded in magnitude by 1 (note that this generalizes (5)).

The next step, as before, is to establish a dichotomy between pseudorandomness and structure. The analogue of Lemma 2.2 or Lemma 3.2 is

Lemma 4.2 (Dichotomy between randomness and structure). *Suppose that* $f : V \times V \to \mathbb{R}$ *is bounded in magnitude by* 1 *with* $\|f\|_{\square^2(\mathbb{Z}/N\mathbb{Z})} \geq \eta$ *for some* $0 < \eta < 1$.

Then there exists sets $A, B \subset V$ *such that* $|\mathbf{E}_{x,y\in V} f(x,y)1_A(x)1_B(y)| \geq \eta^4/4$. *Here* $1_A(x)$ *denotes the indicator function of* A (*thus* $1_A(x) = 1$ *if* $x \in A$ *and* $1_A(x) = 0$ *otherwise*).

This lemma follows from an easy application of the pigeonhole principle and is omitted. One can iterate it (by an energy increment argument, as in Theorem 3.7) to obtain a weak version of the Szemerédi regularity lemma:

Theorem 4.3 (Weak structure theorem [10]). *Let* $f : V \times V \to \mathbb{R}$ *be a non-negative function bounded by* 1, *and let* $\varepsilon > 0$. *Then we can decompose* $f = f_{U^\perp} + f_U$, *where* $f_{U^\perp} = \mathbf{E}(f|\mathcal{Z}\otimes\mathcal{Z})$, $\mathcal{Z}$ *is a* σ*-algebra of* V *generated by at most* $2/\varepsilon$ *sets, and* $\|f_U\|_{\square^2} \leq \varepsilon$.

As with Theorem 3.4, the above theorem is too weak to be of much use, because the control one has on the pseudorandomness of f_U is fairly poor compared to the control on the complexity of $f_{U^\perp}$. The following strong version of the regularity lemma is far more useful (compare with Theorem 3.5):

Theorem 4.4 (Strong structure theorem [57]). *Let* $f : V \times V \to \mathbb{R}$ *be a non-negative function bounded by* 1, *and let* $\varepsilon > 0$. *Let* $F : \mathbb{N} \to \mathbb{N}$ *be an arbitrary increasing function* (*e.g.* $F(n) = 2^{2^n}$). *Then there exists an integer* $T = O_{F,\varepsilon}(1)$ *and a decomposition* $f = f_{U^\perp} + f_S + f_U$, *where* $f_{U^\perp} = \mathbf{E}(f|\mathcal{Z}\otimes\mathcal{Z})$, $\mathcal{Z}$ *is generated by at most* T *sets in* V, f_U *is pseudorandom in the sense that* $\|f_U\|_{\square^2} = O(1/F(T))$, *and* f_S *is small in the sense that* $\|f_S\|_{L^2(V\times V)} := (\mathbf{E}_{x,y\in V}|f_S(x,y)|^2)^{1/2} = O(\varepsilon)$. *Furthermore,* $f_{U^\perp}$, f_U *are bounded in magnitude by* 1. *Also,* $f_{U^\perp}$ *and* $f_{U^\perp} + f_S$ *are non-negative and bounded by* 1.

One can view $f_{U^\perp}$ as a "coarse" approximation to f, as it is measurable with respect to a fairly low-complexity σ-algebra, and $f_{U^\perp} + f_S = \mathbf{E}(f|\mathcal{Z}^{(n')}\otimes\mathcal{Z}^{(n')})$ as a "fine" approximation to f, which is considerably more complex but is also a far better approximation to f, in fact the accuracy of the fine approximation exceeds the complexity of the coarse approximation by any specified growth function F. Also the difference between the coarse and fine approximations is controlled by an arbitrarily small constant ε.

Theorem 4.4 already easily implies the Szemerédi regularity lemma in its traditional formulation; see [57]. It also implies Lemma 4.1, similar to how Theorem 3.5 implies the $k = 3$ version of Szemerédi's theorem; we omit the standard details.

As in the other two approaches, the above arguments extend (with some additional difficulties) to higher values of k. Again we restrict attention to the $k = 4$ case for simplicity. To locate a progression of length four in a set $A \subset \mathbb{Z}/N\mathbb{Z}$ is now equivalent to solving the system of constraints

$$\begin{array}{cccc} & y & +2z & +3w & \in A \\ -x & & +z & +2w & \in A \\ -2x & -y & & +w & \in A \\ -3x & -2y & -z & & \in A. \end{array} \tag{12}$$

This in turn follows from a hypergraph analogue of the triangle removal lemma. Define a 3-*uniform hypergraph* to be a pair $H = (V, E)$ where V is a finite set of vertices and E is a finite set of unordered triplets (x, y, z) in V, which we refer to as the *edges* of H. Define a *tetrahedron* in H to be a quadruple (x, y, z, w) of vertices such that all four triplets (x, y, z), (y, z, w), (z, w, x), (w, x, y) are edges of H.

Lemma 4.5 (Tetrahedron removal lemma [9]). *For every $0 < \delta < 1$ there exists $0 < \sigma < 1$ with the following property. Let $H = (V, E)$ be a 3-uniform hypergraph graph with $|V| = N$ vertices which contains fewer than σN^4 tetrahedra. Then it is possible to remove $O(\delta N^3)$ edges from H to create a hypergraph H' which contains no tetrahedra whatsoever.*

Letting f be the indicator function of H, we now have a situation where

$$\mathbf{E}_{x,y,z,w \in V} f(x, y, z) f(y, z, w) f(z, w, x) f(w, x, y) \leq \sigma$$

and we need to remove some small components from f so that this average now vanishes completely. Again, the key step here is to obtain a structure theorem that decomposes f into structured parts, small errors, and pseudorandom errors. The notion of pseudorandomness is now captured by the Gowers $\square^3$ cube norm, defined by

$$\|f\|_{\square^3}^8 := \mathbf{E}_{x,y,z,x',y',z' \in V} f(x, y, z) \dots f(x', y', z')$$

where the product is over the eight values of f with first co-ordinate x or x', second co-ordinate y or y', and third co-ordinate z or z'. In the case when f is the indicator function of a hypergraph H, this norm essentially counts the number of octahedra present in H. One can obtain a strong structure theorem analogous to Theorem 4.4, but with one significant difference. In Theorem 4.4, the structured component $f_{U^\perp}(x, y)$ can be broken up into a small number of components which are of the form $1_A(x)1_B(y)$. In the 3-uniform hypergraph analogue of Theorem 4.4, the structured component $f_{U^\perp}(x, y, z)$ will be broken up into a small number of components of the form $1_A(x, y)1_B(y, z)1_C(z, x)$. It turns out that in order to conclude the proof of Lemma 4.5, this structural decomposition is not sufficient by itself; one must also turn to the functions $1_A(x, y)$, $1_B(y, z)$, $1_C(z, x)$ generated by this structure theorem and decompose them further, essentially by invoking Theorem 4.4. This leads to some technical complications in the argument, although this approach to Szemerédi's theorem is still the most elementary and self-contained. See [21], [46], [47], [48], [49], [58] for details.

5. The primes

Having surveyed the three major approaches to Szemerédi's theorem, we now turn to the question of counting progressions in the primes (or in dense subsets of the primes). The major new difficulty here, of course, is that the primes have asymptotically zero

density rather than positive density, and even the most recent quantitative bounds on Szemerédi's theorem (see the discussion after Theorem 3.1) are not strong enough by themselves to overcome the "thinness" of the primes. However, it turns out that the primes (and functions supported on the primes) are still within the range of applicability of structure theorems. For instance, to oversimplify dramatically, the structure theorem in [27] essentially[5] represents the primes (or any dense subset of the primes) as a (sparse) pseudorandom subset of a set of positive density. Since sets of positive density already contain many progressions thanks to Szemerédi's theorem, it turns out that enough of these progressions survive when passing to a pseudorandom subset that one can conclude Theorem 3.1.

Interestingly, Theorem 1.2 can be tackled by (quantitative) ergodic methods, by Fourier-analytic methods, and by graph-theoretic methods, with the three approaches leading to slightly different results. For instance, the establishment of infinitely many progressions of length three in the primes by van der Corput [62] was Fourier-analytic, as was the corresponding statement for dense subsets of the primes (i.e. the $k = 3$ case of Theorem 1.2), proven 76 years later by Green [22]. The argument in [27] which proves Theorem 1.2 in full combines ideas from all three approaches, but is closest in spirit to the ergodic approach, albeit set in the finitary context of a cyclic group $\mathbb{Z}/N\mathbb{Z}$ rather than on an infinitary measure space. The argument in [59], which shows that the Gaussian primes (or any dense subset thereof) contains infinitely many constellations of any prescribed shape, and can be viewed as a two-dimensional analogue of Theorem 1.2, was proven via the (hyper)graph-theoretical approach. Finally, a more recent argument in [30], [31], in which precise asymptotics for the number of progressions of length four in the primes are obtained, as well as a "quadratic pseudorandomness" estimate on a renormalized counting function for the primes, proceeds by returning back to the original Fourier-analytic approach, but now using quadratic Fourier-analytic tools (Lemma 3.6 and Theorem 3.7) rather than linear ones.

As mentioned in the introduction, these results are discussed in other surveys [42], [25], [55], [56], [37], and we will only sketch some highlights here. In all the results, the strategy is to try to isolate the "structured" component of the primes from the "pseudorandom" component. There is some obvious structure present in the primes; for instance, they are almost all odd, they are almost all coprime to three, and so forth. This obvious structure can be normalized away fairly easily. For instance, to remove the bias the primes have towards being odd, one can replace the primes $P = \{2, 3, 5, \dots\}$ with the renormalized set $P_{2,1} := \{n : 2n+1 \text{ prime}\} = \{1, 2, 3, 5, \dots\}$. Each arithmetic progression in $P_{2,1}$ clearly induces a corresponding progression in P,

[5] This is a gross oversimplification. The precise statement is that after eliminating obvious irregularities in the primes caused by small residue classes, and excluding a small and technical exceptional set, a normalized counting function on the primes can be decomposed as a bounded function (which is thus spread out over a set of positive density), plus a pseudorandom error. Ignoring the initial elimination of obvious irregularities and the exceptional set, and pretending the bounded function was the indicator function of a positive density set A, one recovers the interpretation of the primes as a sparse pseudorandom subset of A.

but the set $P_{2,1}$ has no bias modulo 2. More generally, to reduce all the bias present in residue classes mod p for all $p < w$ (where w is a medium-sized parameter to be chosen later), one can work with a set $P_{W,b} := \{n : Wn + b \text{ prime}\}$, where W is the product of all the primes less than w and $1 \leq b < W$ is a number coprime to W. This "W-trick" allows for some technical simplifications.

Next, it is convenient not to work with the primes as a set, but rather as a renormalized counting function. One convenient choice is the von Mangoldt function $\Lambda(n)$, defined as $\log p$ if n is a power of a prime p and 0 otherwise. Actually, because of the W-trick, it is better to consider a renormalized von Mangoldt function such as $\Lambda_{W,b}(n) := \frac{W}{\phi(W)}\Lambda(Wn + b)$, where $\phi(W)$ is the Euler totient function of W. The prime number theorem in arithmetic progressions asserts that the asymptotic average value of $\Lambda_{W,b}(n)$ is equal to 1. To establish progressions of length k in the primes, it suffices to obtain a nontrivial lower bound for the asymptotic value of the average

$$\mathbf{E}_{1\leq n,r\leq N}\Lambda_{W,b}(n)\Lambda_{W,b}(n+r)\dots\Lambda_{W,b}(n+(k-1)r). \tag{13}$$

In fact this quantity is conjectured to asymptotically equal 1 as $W, N \to \infty$, with W growing much slower than N (a special case of the Hardy–Littlewood prime tuples conjecture); the intuition is that by removing all the bias present in the small residue classes, we have eliminated all the "obvious" structure in the primes, and the renormalized function $\Lambda_{W,b}$ should now fluctuate pseudorandomly around its mean value 1. However, this conjecture has only been verified in the cases $k = 3, 4$ (leading to an asymptotic count for the number of progressions of primes of length k less than a large number N); for the cases $k > 4$ we only have a lower bound of $c(k)$ for some small $c(k) > 0$.

Let us cheat slightly by pretending that $\Lambda_{W,b}$ is a function on the cyclic group $\mathbb{Z}/N\mathbb{Z}$ rather than on the integers $\mathbb{Z}$; there are some minor technical truncation issues that need to be addressed to pass from one to the other but we shall ignore them here. In order to show that (13) is close to 1, an obvious way to proceed would be to establish some kind of pseudorandomness control on the deviation $\Lambda_{W,b} - 1$ from the mean, and then some sort of generalized von Neumann theorem to show that this deviation is negligible. Based on the experience with Szemerédi's theorem, one would expect linear pseudorandomness to be the correct notion for $k = 3$, quadratic pseudorandomness for $k = 4$, and so forth. In the $k = 3$ case it is indeed a standard computation (using Vinogradov's method, or a modern variant of that method such as the one based on Vaughan's identity) to show that $\Lambda_{W,b} - 1$ is has small Fourier coefficients, which is a reasonable proxy for linear pseudorandomness; the point being that the W-trick has eliminated all the "major arcs" which would otherwise destroy the pseudorandomness. It then remains to obtain a generalized von Neumann theorem, similar to (7). In preceding sections, one was working with functions that were bounded (and hence square integrable), and one could obtain these theorems easily from Plancherel's theorem. In the current setting, the L^2 estimates on $\Lambda_{W,b}$ are unfavourable, and what one needs instead is some sort of l^p bound on the Fourier coefficients of $\Lambda_{W,b}$ for some $2 < p < 3$. This can be done by a more careful

application of Vinogradov's method, but can also be achieved using harmonic analysis methods arising from restriction theory; see [22], [28]. The key new insight here is that while the Fourier coefficients of $\Lambda_{W,b}$ are difficult to understand directly, one can *majorize* $\Lambda_{W,b}$ pointwise by (a constant multiple of) a much better behaved function ν of comparable size, whose Fourier coefficients are much easier to obtain bounds for (indeed ν is essentially linearly pseudorandom once one subtracts off its mean, which is essentially 1). This "enveloping sieve" ν is essentially the Selberg upper bound sieve, and can be viewed as a "smoothed out" version[6] of $\Lambda_{W,b}$. Restriction theory (related to the method of the large sieve) is then used to pass from Fourier control of ν to Fourier control of $\Lambda_{W,b}$.

A similar idea was used in [22], [28] to establish the $k = 3$ case of Theorem 1.2; we sketch the argument from [28] here as follows. The main objective is to establish a lower bound for expressions such as

$$\mathbf{E}_{x,r\in\mathbb{Z}/N\mathbb{Z}}\Lambda_{W,b}1_A(x)\Lambda_{W,b}1_A(x+r)\Lambda_{W,b}1_A(x+2r) \tag{14}$$

for large sets A. Restriction theory still allows us to obtain good l^p upper bound for the Fourier coefficients of $\Lambda_{W,b}1_A$. This functions as a substitute for Plancherel's theorem (which is not favourable here), and one can now obtain structure theorems such as Theorem 3.4 (and with some more effort, Theorem 3.5). This decomposes $\Lambda_{W,b}1_A$ into some structured component $f_{U^\perp}$ and a linearly pseudorandom component f_U. The generalized von Neumann theorem lets us dispose the contribution of f_U to (14), so let us focus on $f_{U^\perp}$. One can try to use the complexity bound on $f_{U^\perp}$ (controlling the number of linear phases that comprise $f_{U^\perp}$) to get some lower bound here, but this would require developing a strong structure theorem analogous to Theorem 3.5. It turns out that one can argue more cheaply, using a weaker structure theorem analogous to Theorem 3.4. The key observation is that because $\Lambda_{W,b}1_A$ is dominated (up to a constant) by the enveloping sieve ν, the structured component of $\Lambda_{W,b}1_A$ (which is essentially a convolution of $\Lambda_{W,b}1_A$ with a Fejér-like kernel) is pointwise dominated (up to a constant) by a corresponding structured component of ν. But since ν is linearly pseudorandom after subtracting off its mean, the structured component of ν turns out to essentially be just the mean of ν, which is bounded. We conclude that $f_{U^\perp}$ is bounded, at which point one can just apply Szemerédi's theorem (Theorem 3.1) directly to obtain a good lower bound on this contribution to (14), and one can now conclude the $k = 3$ case of Theorem 1.2.

The proof of Theorem 1.2 for general k in [27] follows the same general strategy, but it is convenient to abandon the Fourier framework (which becomes quite complicated for $k > 3$) and instead take an approach which borrows ingredients from all three approaches, especially the ergodic theory approach. From the Fourier approach one borrows the Gowers uniformity norms $U^{k-2}(\mathbb{Z}/N\mathbb{Z})$, which are a convenient way to define the appropriate notion of pseudorandomness for counting progressions of

[6]What is essentially happening here is that we are viewing the primes not as a zero density subset of the integers, but as a positive density subset of a set of "almost primes" which can be controlled efficiently via sieve theory.

length k. One still needs an enveloping sieve ν, but instead of using a Selberg-type sieve that enjoys good Fourier coefficient control, it turns out to be more convenient to use an enveloping sieve[7] of Goldston and Yıldırım [15], [16], [17] which has good control on k-point correlations (indeed, it behaves pseudorandomly after subtracting off its mean, which is essentially 1).

The next step is a generalized von Neumann theorem to show that the contribution of pseudorandom functions are negligible. The fact that the functions involved are no longer bounded by 1, but are instead dominated by ν, makes this theorem somewhat trickier to establish, however it can still be achieved by a number of applications of the Cauchy–Schwarz and taking advantage of the pseudorandomness properties of $\nu - 1$. This type of argument is inspired by certain "sparse counting lemmas" arising from the hypergraph approach, particuarly from [21].

The main step, as in previous sections, is a structure theorem which decomposes $\Lambda_{W,b}$ (or $\Lambda_{W,b} 1_A$) into a structured component and a pseudorandom component. In principle one could use higher order Fourier analysis (or the precise characteristic factors achieved in [39], [64] to obtain this decomposition, but this looks rather difficult technically, though progress has been made in the $k = 4$ case. Fortunately, there is a "softer" approach in which one defines structure purely by duality; to oversimplify substantially, one defines a function to be structured if it is approximately orthogonal to all pseudorandom functions. One can then obtain a soft structural theorem in which the structural component is essentially a conditional expectation of the original function to a certain σ-algebra generated by certain special structured functions which are called "dual functions" in [27]. This σ-algebra (the finitary analogue of a characteristic factor) is not too tractable to work with, but somewhat miraculously, one can utilize the pseudorandomness properties of ν and a large number of applications of the Cauchy–Schwarz inequality to show that the conditional expectation of ν with respect to this σ-algebra remains bounded (outside of a small exceptional set, which turns out to have a negligible impact). Since $\Lambda_{W,b} 1_A$ is pointwise dominated by a constant multiple of ν, the structured component of $\Lambda_{W,b} 1_A$ is similarly bounded and can thus be controlled using Szemerédi's theorem. Combining this with the generalized von Neumann theorem to handle the pseudorandom component, one obtains Theorem 1.2. The result for the Gaussian prime constellations is similar, but uses the Gowers cube norms $\square^{k-2}$ instead of the uniformity norms, and replaces Szemerédi's theorem by a hypergraph removal lemma similar to Lemma 4.1 and Lemma 4.5; see [58], [59].

The arguments used to prove Theorem 1.2 give a lower bound for the expression (13), but do not compute its asymptotic value (which should be 1). As mentioned earlier, for $k = 3$ this can be achieved by the circle method. More recently, the $k = 4$ case has been carried out in [30], [31]; the same method in fact allows one to asymptotically count the number of solutions to any two linear homogeneous equations in four prime unknowns. The key point is to show that $\Lambda_{W,b} - 1$ is quadratically pseudorandom, as the generalized von Neumann theorem will then allow one to con-

[7] A related enveloping sieve was also used in the recent establishment of narrow gaps in the primes [18].

trol (13) satisfactorily. It turns out that a variant of Lemma 3.6 applies here, and reduces matters to showing that $\Lambda_{W,b} - 1$ does not correlate significantly with any 2-step nilsequences. This task is attackable by Vinogradov's method, although it is rather lengthy and it turns out to be simpler to first replace $\Lambda_{W,b} - 1$ with the closely related Möbius function.

References

[1] Alon, N., Shapira, A., A characterization of the (natural) graph properties testable with one-sided error. In *46th Symposium on Foundations of Computer Science*, IEEE Computer Soc. Press, Los Alamitos, CA, 2005, 429–438.

[2] Behrend, F. A., On sets of integers which contain no three terms in arithmetic progression. *Proc. Nat. Acad. Sci.* **32** (1946), 331–332.

[3] Bergelson, V., Host, B., Kra, B., Multiple recurrence and nilsequences. *Invent. Math.* **160** (2) (2005), 261–303.

[4] Bourgain, J., A Szemerédi type theorem for sets of positive density in $\mathbb{R}^k$. *Israel J. Math.* **54** (3) (1986), 307–316.

[5] Bourgain, J., On triples in arithmetic progression. *Geom. Funct. Anal.* **9** (1999), 968–984.

[6] Chung, F., Graham, R., Wilson, R. M., Quasi-random graphs. *Combinatorica* **9** (1989), 345–362.

[7] Conze, J. P., Lesigne, E., Sur un théorème ergodique pour les mesures diagonales. In *Probabilités*, Publ. Inst. Rech. Math. Rennes, 1987-1, Univ. Rennes I, Rennes 1988, 1–31.

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

[9] Frankl, P., Rödl, V., Extremal problems on set systems. *Random Structures Algorithms* **20** (2) (2002), 131–164.

[10] Frieze, A., Kannan, R., Quick approximation to matrices and applications. *Combinatorica* **19** (2) (1999), 175–220.

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

[12] Furstenberg, H., *Recurrence in Ergodic theory and Combinatorial Number Theory*. Princeton University Press, Princeton, NJ, 1981.

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

[14] Furstenberg, H., Weiss, B., 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. 5, Walter de Gruyter, Berlin 1996, 193–227.

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

[16] Goldston, D., Yıldırım, C. Y., Higher correlations of divisor sums related to primes. III: k-correlations. Preprint.

[17] Goldston, D., Yıldırım, C. Y., Small gaps between primes, I. Preprint.

[18] Goldston, D., Motohashi, Y., Pintz, J., Yıldırım, C.Y., Small gaps between primes exist. *Proc. Japan Acad. Ser. A Math. Sci.* **82** (4) (2006), 61–65.

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

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

[21] Gowers, T., Hypergraph regularity and the multidimensional Szemerédi theorem. Preprint.

[22] Green, B. J., Roth's theorem in the primes. *Ann. of Math.* **161** (3) (2005), 1609–1636.

[23] Green, B. J., A Szemerédi-type regularity lemma in abelian groups. *Geom. Funct. Anal.* **15** (2) (2005), 340–376.

[24] Green, B. J., Finite field models in additive combinatorics. In *Surveys in Combinatorics*, London Math. Soc. Lecture Note Ser. 327, Cambridge University Press, Cambridge 2005, 1–27.

[25] Green, B. J., Long arithmetic progressions of primes. Preprint.

[26] Green, B. J., Konyagin, S., On the Littlewood problem modulo a prime. *Canad. J. Math.*, to appear.

[27] Green, B. J., Tao, T., The primes contain arbitrarily long arithmetic progressions. *Ann. of Math.*, to appear.

[28] Green, B. J., Tao, T., Restriction theory of Selberg's sieve, with applications. *J. Théor. Nombres Bordeaux* **18** (2006), 147–182.

[29] Green, B. J., Tao, T., An inverse theorem for the Gowers U^3 norm. *Proc. Edinburgh Math. Soc.*, to appear.

[30] Green, B. J., Tao, T., Quadratic uniformity of the Möbius function. Preprint.

[31] Green, B. J., Tao, T., Two linear equations in four prime unknowns. Preprint.

[32] Green, B. J., Tao, T., New bounds for Szemerédi's theorem, I: Progressions of length 4 in finite field geometries. Preprint.

[33] Green, B. J., Tao, T., New bounds for Szemerédi's theorem, II: A new bound for $r_4(N)$. In preparation.

[34] Green, B. J., Tao, T., New bounds for Szemerédi's theorem, III: A polylog bound for $r_4(N)$. In preparation.

[35] Green, B. J., Tao, T., On arithmetic regularity lemmas. In preparation.

[36] Hardy, G. H., Littlewood, J. E., Some problems of "partitio numerorum"; III: On the expression of a number as a sum of primes. *Acta Math.* **44** (1923), 1–70.

[37] Host, B., Progressions arithmetiques dans les nombres premiers (d'apres B. Green and T. Tao). *Seminaire Bourbaki*, Mars 2005, 57eme annee, 2004-2005, no. 944.

[38] Host, B., Kra, B., Convergence of Conze-Lesigne averages. *Ergodic Theory Dynam. Systems* **21** (2) (2001), 493–509.

[39] Host, B., Kra, B., Non-conventional ergodic averages and nilmanifolds. *Ann. of Math.* **161** (1) (2005) 397–488.

[40] Komlós, J., A generalization of a problem of Steinhaus. *Acta Math. Hungar.* **18** (1967), 217–229.

[41] Komlós, J., Simonovits, M., Szemerédi's regularity lemma and its applications in graph theory. In *Combinatorics*, Paul Erdös is eighty, Vol. 2 (Keszthely, 1993), Bolyai Soc. Math. Stud. 2, János Bolyai Math. Soc., Budapest 1996, 295–352.

[42] Kra, B., The Green-Tao Theorem on arithmetic progressions in the primes: an ergodic point of view. *Bull. Amer. Math. Soc.* (*N.S.*) **43** (1) (2006), 3–23.

[43] Lacey, M., Thiele, C. L^p estimates on the bilinear Hilbert transform for $2 < p < \infty$. *Ann. of Math.* **146** (1997), 693–724.

[44] Leibman, A., Polynomial sequences in groups. *J. Algebra* **201** (1998), 189–206.

[45] L. Lovász, Szegedy, B., Szemerédi's lemma for the analyst. Preprint.

[46] Nagle, B., Rödl, V., Schacht, M., The counting lemma for regular k-uniform hypergraphs. *Random Structures Algorithms* **28** (2) (2006), 113–179.

[47] Rödl, V., Schacht, M., Regular partitions of hypergraphs. *Combin. Probab. Comput.*, to appear.

[48] Rödl, V., Skokan, J., Regularity lemma for uniform hypergraphs. *Random Structures Algorithms* **25** (1) (2004), 1–42.

[49] Rödl, V., Skokan, J., Applications of the regularity lemma for uniform hypergraphs. *Random Structures Algorithms* **28** (2) (2006), 180–194.

[50] Roth, K.F., On certain sets of integers. *J. London Math. Soc.* **28** (1953), 245–252.

[51] Ruzsa, I., Szemerédi E., Triple systems with no six points carrying three triangles. *Colloq. Math. Soc. J. Bolyai* **18** (1978), 939–945.

[52] Stein, E. *Harmonic Analysis: Real Variable Methods, Orthogonality, and Oscillatory Integrals*. Princeton University Press, Princeton, NJ, 1993.

[53] Szemerédi, E., On sets of integers containing no k elements in arithmetic progression. *Acta Arith.* **27** (1975), 299–345.

[54] Tao, T., A quantitative ergodic theory proof of Szemerédi's theorem. *Electron. J. Combin.*, to appear.

[55] Tao, T., Obstructions to uniformity, and arithmetic patterns in the primes. *Quarterly J. Pure Appl. Math.* **2** (2006), 199–217.

[56] Tao, T., Arithmetic progressions in the primes. *Collect. Math.* (2006), Vol. Extra., 37–88.

[57] Tao, T., Szemerédi's regularity lemma revisited. *Contrib. Discrete Math.* **1** (1) (2006), 8–28.

[58] Tao, T., A variant of the hypergraph removal lemma. *J. Combin. Theory Ser. A* **113** (7), 1257–1280.

[59] Tao, T., The Gaussian primes contain arbitrarily shaped constellations. *J. Anal. Math.* **99** (2006), 109–176.

[60] Tao, T., An ergodic transference theorem. Unpublished.

[61] Tao, T., Vu, V., Additive Combinatorics. Book in preparation, Cambridge University Press.

[62] van der Corput, J. G., Über Summen von Primzahlen und Primzahlquadraten. *Math. Ann.* **116** (1939), 1–50.

[63] Varnavides, P., On certain sets of positive density. *J. London Math. Soc.* **34** (1959) 358–360.

[64] Ziegler, T., Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.* **20** (2007), 53–97.

Department of Mathematics, UCLA, Los Angeles CA 90095, U.S.A.

E-mail: tao@math.ucla.edu