THE GREEN-TAO THEOREM ON ARITHMETIC PROGRESSIONS IN THE PRIMES: AN ERGODIC POINT OF VIEW

BRYNA KRA

Abstract. A long standing and almost folkloric conjecture is that the primes contain arbitrarily long arithmetic progressions. Until recently, the only progress on this conjecture was due to van der Corput, who showed in 1939 that there are infinitely many triples of primes in arithmetic progression. In an amazing fusion of methods from analytic number theory and ergodic theory, Ben Green and Terence Tao showed that for any positive integer k, there exist infinitely many arithmetic progressions of length k consisting only of prime numbers. This is an introduction to some of the ideas in the proof, concentrating on the connections to ergodic theory.

1. Background

For hundreds of years, mathematicians have made conjectures about patterns in the primes: one of the simplest to state is that the primes contain arbitrarily long arithmetic progressions. It is not clear exactly when this conjecture was first formalized, but as early as 1770 Lagrange and Waring studied the problem of how large the common difference of an arithmetic progression of k primes must be. A natural extension of this question is to ask if the prime numbers contain arbitrarily long arithmetic progressions.

Support for a positive answer to this question is provided by the following simple heuristic. The Prime Number Theorem states that the number of prime numbers less than the integer N is asymptotically $N/\log N$. It follows that the density of primes around a positive large $x \in \mathbb{R}$ is about $1/\log x$. Thus if we model the sequence of prime numbers in $\{1, \dots, N\}$ by choosing integers at random, with an integer in $\{1, \dots, N\}$ being chosen with probability $1/\log N$, then there ought to be approximately $N^2/\log^k N$ progressions of length k consisting of prime numbers less than or equal to N. Unfortunately, one cannot

This article is an expanded version of a lecture presented January 7, 2005 at the AMS Special Session on Current Events, Joint Mathematics Meeting, Atlanta, GA. The author was partially supported by NSF grant DMS-0244994.

hope to use this sort of argument to show that the primes contain any particular pattern, since the primes are far from being randomly distributed: 2 is the only even prime, 3 is the only prime congruent to $0 \mod 3$, and so on.

In 1923, Hardy and Littlewood [18] made a very general conjecture (the *k-tuple conjecture*) about patterns and their distribution in the primes: if $a_1, \ldots, a_k$ and $b_1, \ldots, b_k$ are nonnegative integers such that $P(x) = \prod_{i=1}^{k}(a_ix + b_i)$ is not identically 0 modulo any prime p, then there are infinitely many integers n such that $\{a_in + b_i \colon 1 \leq i \leq k\}$ consists only of primes. This conjecture includes the twin prime conjecture (there exist infinitely many primes p such that $p + 2$ is also prime) as a special case and it trivially implies that the primes contain arbitrarily long arithmetic progressions. Moreover it implies that the number of k-term arithmetic progressions in the primes bounded by N is asymptotically $c_kN^2/\log^k N$ for a certain explicit value of c_k.

There are numerous related conjectures about the existence of arithmetic progressions in certain subsets of the integers. For example, the famous conjecture of Erdös and Turán [6] states that if $A = \{a_1 < a_2 < \ldots\}$ is an infinite sequence of integers with $\sum_i 1/a_i = \infty$, then A contains arbitrarily long arithmetic progressions. A corollary would be that the primes contain arbitrarily long arithmetic progressions.

The first major progress on arithmetic progressions in the primes was made by van der Corput [30], who proved in 1939 that the primes contain infinitely many arithmetic progressions of length 3. Further progress was only made in 1981, when Heath-Brown [19] showed that there are infinitely many arithmetic progressions of length 4 consisting of three primes and an almost prime, meaning either a prime or a product of two primes. In a slightly different direction are the elegant results of Balog ([1], [2]) on patterns in the primes. For example, he shows that for any positive integer k, there exist infinitely many k-tuples of distinct primes $p_1 < p_2 < \ldots < p_k$ such that $(p_i + p_j)/2$ is prime for all $i, j \in \{1, \ldots, k\}$. For $k = 2$ this implies, in particular, that the primes contain infinitely many arithmetic progressions of length 3.

Computational mathematicians have also worked on the problem of finding long arithmetic progressions in the primes. In 1995, Moran, Pritchard and Thyssen [25] found a progression of length 22 in the primes. This record was finally broken in 2004, when Frind, Jobling and Underwood [7] found a progression of length 23 starting with the prime 56211383760397 and with common difference 44546738095860.

In 2004, Ben Green and Terence Tao announced a major breakthrough, with a proof of the general case:

Theorem 1.1 (Green and Tao [15]). *For every integer $k \geq 1$, the prime numbers contain an arithmetic progression of length k.*

They [16] also extract a bound on how far out in the primes one must go in order to guarantee finding an arithmetic progression of length k, showing that there is a k-term arithmetic progression of primes all of whose entries are bounded by

$$2^{2^{2^{2^{2^{2^{2^{2^{(100k)}}}}}}}} .$$

This bound is considered far from optimal; standard heuristics in number theory, plus a little calculation, lead to the conjecture [13] that there is an arithmetic progression of length k in the primes all of whose entries are bounded by $k! + 1$.

Green and Tao prove a stronger statement than that given in Theorem 1.1. They show that not only do the primes contain arbitrarily long arithmetic progressions, but so does any sufficiently dense subset of the primes:

Theorem 1.2 (Green and Tao [15]). *If A is a subset of prime numbers with*

$$\limsup_{N\to\infty} \frac{1}{\pi(N)} |A \cap \{1, \ldots, N\}| > 0 ,$$

where $\pi(N)$ is the number of primes in $\{1, \ldots, N\}$, then for every integer $k \geq 1$, A contains an arithmetic progression of length k.

For $k = 3$, this was proved by Green [14].

The theorem of Green and Tao is a beautiful result answering an old conjecture that has attracted much work. Perhaps even more impressive is the fusion of methods and results from number theory, ergodic theory, harmonic analysis, discrete geometry, and combinatorics used in its proof. The starting point for Green and Tao's proof is the celebrated theorem of Szemerédi [27]: a set of integers with positive upper density[1] contains arbitrarily long arithmetic progressions. One of the main ideas is to generalize this, showing that a dense subset of a sufficiently *pseudorandom* collection (see Section 7 for the precise definition) of the integers contains arbitrarily long arithmetic progressions. There are three major ingredients in the proof. The first is Szemerédi's Theorem itself. Since the primes do not have positive upper density,

[1] The *upper density* $d^*(A)$ of a subset A of the integers is defined to be

$$d^*(A) := \limsup_{N\to\infty} |A \cap \{1, \ldots, N\}|/N .$$

Szemerédi's Theorem can not be directly applied and the second major ingredient in Green and Tao's proof is a certain transference principle that allows one to use Szemerédi's Theorem in a more general setting. The last major ingredient is the use of specific properties of the primes and their distribution, based on recent work of Goldston and Yildirim [17], showing that this generalized Szemerédi Theorem applies to the primes.

It is impossible to give a complete proof of their theorem in this limited space, nor even to do justice to the main ideas. Our goal is to outline the main ingredients and focus on the relation between their work and recent parallel advances in ergodic theory. The interaction between combinatorial number theory and ergodic theory began with Furstenberg's proof [8] of Szemerédi's Theorem (see Section 3) and has led to many new results. Until the present, this interaction has mainly taken the form of using ergodic theory to prove statements in combinatorial number theory, such as Szemerédi's Theorem, its generalizations (including a multidimensional version [9] and a polynomial version [4]), and the density Hales-Jewett Theorem [10]. Green and Tao's work opens a new chapter in this interaction, with ergodic theoretic *proof techniques* being adapted for use in a number theoretic setting.

Acknowledgments: I appreciate the many helpful comments I received while preparing this manuscript, including those from Keith Burns, Nikos Frantzikinakis, Bernard Host, Margaret Symington, Terence Tao, and Mate Wierdl, and especially those from Andrew Granville.

2. Szemerédi's Theorem

Substituting the set of all integers for the set of primes in Theorem 1.2, one obtains Szemerédi's Theorem. We state an equivalent finite version of this theorem:

Theorem 2.1 (**Finite Szemeréd**i [27]). *Let $0 < \delta \leq 1$ be a real number and let $k \geq 1$ be an integer. There exists $N_0(\delta, k)$ such that if $N > N_0(\delta, k)$ and $A \subset \{1, \ldots, N\}$ with $|A| \geq \delta N$, then A contains an arithmetic progression of length k.*

It is clear that this version implies the first version of Szemerédi's Theorem, and an easy argument gives the converse implication.

Szemerédi's [27] original proof in 1975 was combinatorial in nature. Shortly thereafter, Furstenberg developed the surprising relation between combinatorics and ergodic theory, proving Szemerédi's Theorem via a multiple recurrence theorem (see Section 3). More recently, Gowers [12] gave a new proof of Szemerédi's Theorem using harmonic analysis, vastly improving the known bounds for $N_0(\delta, k)$ in the finite version. Although the various proofs, Szemerédi's, Furstenberg's, and Gowers', seem to use very different methods, they have several features in common. In each, a key idea is the dichotomy in the underlying space (whether it be a subset of the integers, a measure space, or the finite group $\mathbb{Z}/N\mathbb{Z}$) between randomness and structure. One then has to analyze the structured part of the space to understand the intersection of a set with itself along arithmetic progressions. We start by further discussing Furstenberg's proof of Szemerédi's Theorem using ergodic theory.

3. Szemerédi's Theorem and ergodic theory

Furstenberg proved the multiple ergodic theorem:

Theorem 3.1 (**Multiple Recurrence** [8])**.** *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving probability system*[2] *and let $k \geq 1$ be an integer. For any set $E \in \mathcal{X}$ with $\mu(E) > 0$,*

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\left(E \cap T^{-n}E \cap T^{-2n}E \cap \ldots \cap T^{-(k-1)n}E\right) > 0. \tag{3.1}$$

An obvious corollary is:

Corollary 3.2. *Let $(X, \mathcal{X}, \mu, T)$ be a measure preserving probability system and let $k \geq 1$ be an integer. For any set $E \in \mathcal{X}$ with $\mu(E) > 0$, there exists an integer $n \geq 1$ such that*

$$\mu\left(E \cap T^{-n}E \cap T^{-2n}E \cap \ldots \cap T^{-(k-1)n}E\right) > 0\ .$$

Furstenberg then made the beautiful connection to combinatorics, showing that regularity properties of integers with positive upper density correspond to multiple recurrence results:

Theorem 3.3 (**Correspondence Principle** [8])**.** *Assume that A is a subset of integers with positive upper density. There exist a measure preserving probability system $(X, \mathcal{X}, \mu, T)$ and a measurable set $E \in \mathcal{X}$*

[2] By a *measure preserving probability system*, we mean a quadruple $(X, \mathcal{X}, \mu, T)$, where X is a set, $\mathcal{X}$ denotes a σ-algebra on X (meaning an algebra $\mathcal{X}$ of subsets of X that is closed under countable unions), μ is a probability measure on $(X, \mathcal{X})$, and $T\colon X \to X$ is a measurable map such that $\mu(A) = \mu(T^{-1}A)$ for all $A \in \mathcal{X}$. Usually, we assume that X is a metrizable compact set and $\mathcal{X}$ is its *Borel σ-algebra*, meaning the σ-algebra generated by the open sets. In particular, for the spaces under consideration, $L^2(X, \mathcal{X}, \mu)$ is separable. We always denote the σ-algebra by the calligraphic version of the letter used for the space.

with $\mu(E) = d^*(A)$ *such that for all integers* $k \geq 1$ *and all integers* $m_1, \ldots, m_{k-1} \geq 1$*,*

$$d^*\big(A\cap(A+m_1)\cap\ldots\cap(A+m_{k-1})\big) \geq \mu\big(E\cap T^{-m_1}E\cap\ldots\cap T^{-m_{k-1}}E\big)\ .$$

Taking $m_1 = n, m_2 = 2n, \ldots, m_{k-1} = (k-1)n$, Szemerédi's Theorem follows from Corollary 3.2.

Furstenberg's proof relies on a compactness argument, making it difficult to extract any explicit bounds in the finite version of Szemerédi's Theorem. On the other hand, Theorem 3.1 and its proof gave rise to a new area in ergodic theory, called "ergodic Ramsey theory." Ergodic theoretic proofs have lead to many other results in combinatorics, such as the multidimensional Szemerédi Theorem [9] and the polynomial Szemerédi Theorem [4], and some of these generalizations have yet to be attained by other methods. More recent developments in ergodic Ramsey theory closely parallel ideas in Green and Tao's work; we return to this in Section 5.

To prove Theorem 3.1, Furstenberg showed that in any measure preserving system, one of two distinct phenomena occurs to make the measure of this intersection positive. The first is weak mixing,[3] when for any set E, $\mu(E\cap T^{-n}E)$ is approximately $\mu(E)^2$ for most choices of the integer n. Then it can be shown that

$$\mu(E\cap T^{-n}E\cap T^{-2n}E\cap\ldots\cap T^{-(k-1)n}E)$$

is approximately $\mu(E)^k$ for most choices of n, which is clearly positive when E is a set of positive measure. The opposite situation is rigidity, when for appropriately chosen n, T^n is very close to the identity. Then $T^{jn}E$ is very close to E and

$$\mu(E\cap T^{-n}E\cap T^{-2n}E\cap\ldots\cap T^{-(k-1)n}E)$$

is very close to $\mu(E)$, again giving positive intersection for a set E of positive measure. One then has to show that the average along arithmetic progressions for any function can be decomposed into two pieces, one which exhibits a generalized weak mixing property and another that exhibits a generalized rigidity property. One of the difficulties lies in proving a structure theorem for the latter situation, showing that

[3] The system $(X, \mathcal{X}, \mu, T)$ is *weak mixing* if for all $A, B \in \mathcal{X}$,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}\left|\mu\left(T^{-n}A\cap B\right)-\mu(A)\mu(B)\right| = 0\ .$$

this portion of the system can be reduced to a finite series of compact extensions of a one point system (a *Furstenberg tower*) and then proving a recurrence statement for this tower.

4. Gowers norms in combinatorics

In his proof of Szemerédi's Theorem, Gowers [12] defined certain norms, now referred to as *Gowers (uniformity) norms*, that capture behavior along arithmetic progressions. We start with a description of this key idea, explaining it in the combinatorial setup in this section and in the ergodic version in the next section. To define the norms, we introduce some notation.

For a positive integer N, let $\mathbb{Z}_N := \mathbb{Z}/N\mathbb{Z}$. If $f\colon \mathbb{Z}_N \to \mathbb{C}$ is a function, let $\mathbb{E}\left(f(x) \,|\, x \in \mathbb{Z}_N\right)$ denote the average value of f on $\mathbb{Z}_N$:

$$\mathbb{E}\left(f(x) \,|\, x \in \mathbb{Z}_N\right) = \frac{1}{N} \sum_{x \in \mathbb{Z}_N} f(x) \ .$$

We also use a higher dimensional version of the expectation. For example, by $\mathbb{E}(f(x,y) \,|\, x, y \in \mathbb{Z}_N)$, we mean iteration of the one variable expectation:

$$\mathbb{E}\big(\mathbb{E}(f(x,y) \,|\, x \in \mathbb{Z}_N) \,|\, y \in \mathbb{Z}_N\big) \ .$$

Taking f to be the indicator function of a set E whose average on $\mathbb{Z}_N$ is at least δ, in this new terminology Szemerédi's Theorem becomes:

Theorem 4.1 (Reformulated Szemerédi). *Let $0 < \delta \leq 1$ be a real number and let $k \geq 2$ be an integer. If N is sufficiently large and $f\colon \mathbb{Z}_N \to \mathbb{R}$ is a function with $0 \leq f(x) \leq 1$ for all $x \in \mathbb{Z}_N$ and $\mathbb{E}\left(f(x) \,|\, x \in \mathbb{Z}_N\right) \geq \delta$, then*

$$\mathbb{E}\big(f(x)f(x+r)\ldots f(x+(k-1)r) \,|\, x, r \in \mathbb{Z}_N\big) \geq c(k,\delta) \tag{4.1}$$

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

At first glance, this appears to be a stronger version than the original statement of Szemerédi's Theorem, showing not only the existence of a single arithmetic progression but of some positive multiple of N^2 progressions. However, using some combinatorial trickery one can quickly show that the two versions are equivalent.

The average of Equation 4.1 is similar to the average in Equation 3.1, with the former being over $\mathbb{Z}_N$ and the latter over $\mathbb{Z}$. These averages along arithmetic progressions are controlled by certain norms and we now make this idea precise. The definition of the norms is motivated by a variation on the classic van der Corput difference theorem:

Lemma 4.2 (**van der Corput Lemma for $\mathbb{Z}_N$**)**.** *If $f\colon \mathbb{Z}_N \to \mathbb{C}$ is a function, then*

$$|\mathbb{E}(f(x) \,|\, x \in \mathbb{Z}_N)|^2 = \mathbb{E}(\overline{f(x)} f(x+h) \,|\, x, h \in \mathbb{Z}_N) \ .$$

Since each of these expectations is a finite sum, the proof of the lemma is immediate by expanding both sides and using a change of variable.

The *d^{th}-Gowers (uniformity) norm* $\|f\|_{U^d}$ of a function $f\colon \mathbb{Z}_N \to \mathbb{C}$ is defined inductively. Set

$$\|f\|_{U^1} := |\mathbb{E}(f(x) \,|\, x \in \mathbb{Z}_N)|$$

Thus for $d = 1$, $\|f\|_{U^d}$ is only a seminorm.[4] For $d \geq 2$, we mimic successive uses of the van der Corput Lemma and define

$$\|f\|_{U^d} := \mathbb{E}\left(\|\overline{f} f_h\|_{U^{d-1}}^{2^{d-1}} \,|\, h \in \mathbb{Z}_N\right)^{1/2^d} \ , \tag{4.2}$$

where $f_h(x) = f(x+h)$. By definition, $\|f\|_{U^d}$ is non-negative for $d = 1$ and therefore also for all higher d. Furthermore, Equation (4.2) shows that the d^{th}-Gowers norm is shift invariant, meaning that $\|f(x)\|_{U^d} = \|f(x+h)\|_{U^d}$ for any $h \in \mathbb{Z}_N$. To justify the notation and terminology, we need to check that the Gowers norms are actually norms.

It follows immediately from the definitions and a change of variable that

$$\|f\|_{U^1} = \left(\mathbb{E}(f(x)\overline{f(x+h)} \,|\, x, h \in \mathbb{Z}_N)\right)^{1/2} \ . \tag{4.3}$$

The second Gowers norm can also be expressed in familiar terms. Using the Fourier expansion of f and computing, we have that

$$\|f\|_{U^2} = \left(\sum_{\xi \in \mathbb{Z}_N} \left|\hat{f}(\xi)\right|^4\right)^{1/4} \ ,$$

where $\hat{f}$ denotes the Fourier transform of f. It follows that for $d = 2$, $\|f\|_{U^d}$ is nondegenerate and so it is a norm.

For higher d, the situation is more complicated. To see that $\|f\|_{U^d}$ is a norm for $d \geq 2$, we give an equivalent characterization of the d^{th}-Gowers norm as a certain average over a d-dimensional cube. This also allows us to express the definition of the norm in a closed form. We first need to introduce some more notation.

[4] A *seminorm* on a vector space V is a non-negative real valued function such that $\|f+g\| \leq \|f\| + \|g\|$ and $\|cf\| = |c| \cdot \|f\|$ for all $f, g \in V$ and all scalars c. Thus unlike a norm, one may have $\|f\| = 0$ for some $f \neq 0$.

We consider $\{0,1\}^d$ as the set of vertices of the d-dimensional Euclidean cube, meaning it consists of points $\omega = (\omega_1, \ldots, \omega_d)$ with $\omega_j \in \{0,1\}$ for $j = 1, \ldots, d$. For $\omega \in \{0,1\}^d$, define $|\omega| = \omega_1 + \ldots + \omega_d$ and if $\omega \in \{0,1\}^d$ and $\mathbf{h} = (h_1, \ldots, h_d) \in \mathbb{Z}_N^d$, we define $\omega \cdot \mathbf{h} := \omega_1 h_1 + \ldots + \omega_d h_d$. Then if $f \colon \mathbb{Z}_N \to \mathbb{C}$ is a complex valued function, it follows by inductively applying the definition in (4.2) that

$$(4.4) \quad \|f\|_{U^d} := \mathbb{E}\left(\prod_{\omega \in \{0,1\}^d} C^{|\omega|} f(x + \omega \cdot \mathbf{h}) \,|\, x \in \mathbb{Z}_N, \mathbf{h} \in \mathbb{Z}_N^d\right)^{1/2^d},$$

where C is the conjugation operator $Cf(x) := \overline{f(x)}$. Thus the Gowers norms can be viewed as an average over the cube $\{0,1\}^d$.

By repeated applications of the Cauchy-Schwarz Inequality and the definitions of the norms, one obtains the *Gowers Cauchy-Schwarz Inequality* for 2^d functions $f_\omega \colon \mathbb{Z}_N \to \mathbb{C}$:

(4.5)

$$\left|\mathbb{E}\left(\prod_{\omega \in \{0,1\}^d} C^{|\omega|} f_\omega(x + \omega \cdot \mathbf{h}) \,|\, x \in \mathbb{Z}_N, \mathbf{h} \in \mathbb{Z}_N^d\right)\right| \leq \prod_{\omega \in \{0,1\}^d} \|f_\omega\|_{U^d} \; .$$

From this, one can show that $\|f\|_{U^d}$ is subadditive and so is a seminorm. Furthermore, using the Gowers Cauchy-Schwarz Inequality, one has the chain of inequalities

$$(4.6) \qquad \|f\|_{U^1} \leq \|f\|_{U^2} \leq \ldots \leq \|f\|_{L^\infty} \; .$$

Since $\|f\|_{U^d}$ is nondegenerate for $d = 2$, Inequality (4.6) implies that $\|f\|_{U^d}$ is nondegenerate for all higher d, giving that the d^{th}- Gowers norm is actually a norm for $d \geq 2$.

We can also rewrite the Gowers norms in notation that is closer in spirit to the ergodic theoretic setup. Consider $\mathbb{Z}_N$ endowed with the transformation $T(x) = x + 1 \mod N$ and the uniform measure m assigning weight $1/N$ to each element of $\mathbb{Z}_N$. Then the definition of Equation (4.4) becomes:

(4.7)

$$\|f\|_{U^d} = \left(\int \prod_{\omega \in \{0,1\}^d} C^{|\omega|} f(T^{\omega \cdot \mathbf{h}} x) \, dm(x) dm(h_1) \ldots dm(h_d)\right)^{1/2^d} .$$

These norms are used by Gowers (as well as by Host and Kra [21] and more recently by Green and Tao [15] and by Tao [28]) to control the average along arithmetic progressions, meaning the quantity in Equation (4.1). This type of control can be viewed as a generalized version of the von Neumann Ergodic Theorem, which states that the average of a bounded function on a finite measure space converges in mean to its integral. We formalize this control, as stated by Tao [28]:

Theorem 4.3 (**Generalized von Neumann Theorem** [28])**.** *Let $k \geq 2$ be an integer, N be a prime number, and $f_0, \dots, f_{k-1} \colon \mathbb{Z}_N \to \mathbb{C}$ be functions with $\|f_0\|_\infty, \dots, \|f_{k-1}\|_infty \leq 1$. Then*

$$\bigl|\mathbb{E}\bigl(f_0(x)f_1(x+n)\dots f_{k-1}(x+(k-1)n) \mid x, n \in \mathbb{Z}_N\bigr)\bigr| \leq \min_{0\leq j\leq k-1} \|f_j\|_{U^{k-1}} \ .$$

The proof of this theorem is based on an induction argument, using the Cauchy-Schwarz Inequality and the van der Corput Lemma for $\mathbb{Z}_N$ (Lemma 4.2).

To prove Szemerédi's Theorem, Gowers [12] studies the indicator function $\mathbf{1}_A$ of a set $A \subset \mathbb{Z}_N$. As in Furstenberg's proof, there are two distinct phenomena to consider. If $\|\mathbf{1}_A - |A|/N\|_{U^{k-1}}$ is small, then a constant function is substituted for $\mathbf{1}_A$ and the average along arithmetic progressions is large. If $\|\mathbf{1}_A - |A|/N\|_{U^{k-1}}$ is large, then the restriction of $\mathbf{1}_A$ to some sufficiently large subset of $\mathbb{Z}_N$ has many useful arithmetic properties and the average in Equation (4.1) is once again large. As in Furstenberg's proof, a structure theorem is needed to analyze the second case: here the structure is a nested sequence of arithmetic progressions. The difficulty in this proof lies in showing that a usable version of the dichotomy between large and small always occurs.

5. Gowers (semi-)norms in ergodic theory

Furstenberg's proof of Theorem 3.1 left open the question of the existence of the limit in the left hand side of Equation (3.1). In [21], we show that this lim inf is actually a limit:

Theorem 5.1 (**Multiple Convergence** [21])**.** *Assume that $(X, \mathcal{X}, \mu, T)$ is a measure preserving probability system, $k \geq 1$ is an integer, and $f_1, f_2, \dots, f_k$ are bounded functions on X. Then the averages*

$$\frac{1}{N}\sum_{n=0}^{N-1} f_1(T^n x)f_2(T^{2n}x)\dots f_k(T^{kn}x) \tag{5.1}$$

converge in $L^2(\mu)$ as $N \to \infty$.

The existence of the limit for $k = 1$ is von Neumann's ergodic theorem, existence for $k = 2$ was proven by Furstenberg [8], and for $k = 3$ was proven with a technical assumption by Conze and Lesigne [5] and

Furstenberg and Weiss [11], and in general by Host and Kra [20]. More recently, Ziegler [32] has an alternate approach for all k.

The first step in proving Theorem 5.1 is showing that instead of taking the average in the system $(X, \mathcal{X}, \mu, T)$, it suffices to consider the average over some (ostensibly simpler) system $(Y, \mathcal{Y}, \nu, S)$. This amounts to proving a generalized von Neumann Theorem, as in Gowers' proof. This idea is implicit in Furstenberg's [8] proof of Szemerédi's Theorem and made explicit in the proof of Theorem 5.1.

In [21], we introduced seminorms that generalize the Gowers norms; although the language is quite different, the form of the definition can be taken to closely resemble that of the Gowers norms. We consider a general probability measure preserving space $(X, \mathcal{X}, \mu)$ with an invertible measurable, measure preserving transformation $T\colon X \to X$ on it. For a function $f \in L^\infty(\mu)$, we define (compare with Equation (4.2))

$$\|f\|_{U^1} := \left| \int f(x)\, d\mu(x) \right|$$

and inductively we define the d^{th}-seminorm by

$$\|f\|_{U^d}^{2^d} := \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \|f\overline{T^n f}\|_{U^{d-1}}^{2^{d-1}} \,. \tag{5.2}$$

To recover the Gowers norms, we take the space $\mathbb{Z}_N$ with the transformation $x \mapsto x+1 \mod N$ and the uniform measure assigning each element of $\mathbb{Z}_N$ weight $1/N$.

Once again, there is an alternate presentation, analogous to that of Equation (4.7), as the integral with respect to a certain measure and this second presentation makes many properties of $\|f\|_{U^d}$ more transparent. We need some notation to define this measure. (A reader not interested in the technical definition of this measure can omit this alternate presentation.)

Assume that $(X, \mathcal{X}, \mu)$ is a probability space. If $f \in L^1(\mu)$ and $\mathcal{Y} \subset \mathcal{X}$ is a sub-σ-algebra, then the *conditional expectation* of f on $\mathcal{Y}$ is the $\mathcal{Y}$-measurable function $\mathbb{E}(f \mid \mathcal{Y})$ such that

$$\int_A f\, d\mu = \int_A \mathbb{E}(f \mid \mathcal{Y})\, d\mu$$

for all $A \in \mathcal{Y}$.

Fix an ergodic[5] measure preserving probability system $(X, \mathcal{X}, \mu, T)$. Define $X^{[d]} = X^{2^d}$ and write points of $X^{[d]}$ as $\mathbf{x} = (x_\omega \colon \omega \in \{0,1\}^d)$.

[5] The system $(X, \mathcal{X}, \mu, T)$ is *ergodic* if the only sets $A \in \mathcal{X}$ with $\mu(T^{-1}A) = \mu(A)$ have measure 0 or 1. Every system has a decomposition into ergodic components and so we can assume that the system being studied in Theorem 5.1 is ergodic.

Let $T^{[d]} = T \times T \times \ldots \times T$ taken 2^d times. There is a natural identification between $X^{[d+1]}$ and $X^{[d]} \times X^{[d]}$, with a point $\mathbf{x} \in X^{[d+1]}$ being identified with $(\mathbf{x}', \mathbf{x}'') \in X^{[d]} \times X^{[d]}$, where $x'_\omega = x_{\omega,0}$ and $x''_\omega = x_{\omega,1}$ for each $\omega \in \{0,1\}^d$.

For each integer $d \geq 0$, we inductively define a $T^{[d]}$-invariant measure $\mu^{[d]}$ on $X^{[d]}$. Define $\mu^{[0]} := \mu$. Assume that $\mu^{[d]}$ is defined for some $d \geq 0$. Let $\mathcal{I}^{[d]}$ denote the $T^{[d]}$-invariant σ-algebra of $(X^{[d]}, \mu^{[d]}, T^{[d]})$. Using the natural identification of $X^{[d+1]}$ with $X^{[d]} \times X^{[d]}$, define the measure preserving (probability) system $(X^{[d+1]}, \mu^{[d+1]}, T^{[d+1]})$ to be the *relatively independent joining* of $(X^{[d]}, \mu^{[d]}, T^{[d]})$ with itself over $\mathcal{I}^{[d]}$: this means that the measure $\mu^{[d+1]}$ is the measure such that for all bounded functions F' and F'' on $X^{[d]}$, we have

$$\int_{X^{[d+1]}} F'(\mathbf{x}')F''(\mathbf{x}'')\, d\mu^{[d+1]}(\mathbf{x}) = \int_{X^{[d]}} \mathbb{E}(F' \,|\, \mathcal{I}^{[d]})\mathbb{E}(F'' \,|\, \mathcal{I}^{[d]})\, d\mu^{[d]} \ .$$

The measure $\mu^{[d+1]}$ is invariant under $T^{[d+1]}$ and the two natural projections on $X^{[d]}$ are each $\mu^{[d]}$. Using induction, this gives that each of the 2^d natural projections of $\mu^{[d]}$ on X is equal to μ. Thus for a bounded function f on X, the integral

$$\int_{X^{[d]}} \prod_{\omega \in \{0,1\}^d} C^{|\omega|} f(x_\omega)\, d\mu^{[d]}(\mathbf{x})$$

is real and nonnegative, where as before $Cf(x) := \overline{f(x)}$. An alternate definition of the seminorms is:

$$\|f\|_{U^d} = \left(\int_{X^{[d]}} \prod_{\omega \in \{0,1\}^d} C^{|\omega|} f(x_\omega)\, d\mu^{[d]}(\mathbf{x}) \right)^{1/2^d} . \tag{5.3}$$

To show that these are seminorms, one proceeds in the same manner as in the combinatorial setup, deriving a version of the Cauchy-Schwarz Inequality (analogous to Equation (4.5)) and using it to show subadditivity. Positivity follows immediately from definition (5.2). From the definition of these measures and the Ergodic Theorem, we obtain that this second definition is equivalent to the first definition given in Equation (5.2).

The definition of Equation (5.3) can once again be viewed as an average over the cube $\{0,1\}^d$. A convergence theorem for general averages along cubes is also proved in [21].

The first step in proving Theorem 5.1 is showing that the averages along arithmetic progressions are once again controlled by the d-seminorms, meaning an analog of Theorem 4.3:

Theorem 5.2 (**Generalized von Neumann, revisited** [21])**.** *Assume that $(X, \mathcal{X}, \mu, T)$ is an ergodic measure preserving probability system. Let $k \geq 2$ be an integer and assume that $f_1, \dots, f_k$ are bounded functions on X with $\|f_1\|_\infty, \dots \|f_k\|_\infty \leq 1$. Then*

$$\limsup_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} f_1(T^n x) f_2(T^{2n} x) \dots f_k(T^{kn} x) \right\|_2 \leq \min_{1\leq j\leq k} (j\|f_j\|_{U^k}) \ .$$

The added factor of j which appears on the right hand side of this bound and not in Theorem 4.3 is due to the change in underlying space. In Theorem 4.3, we assumed that N is prime; in this case, for any integer j that is not a multiple of N, the map $n \mapsto jn$ is onto in $\mathbb{Z}_N$, and this is not the case in $\mathbb{Z}$. As for the earlier Generalized von Neumann Theorem, Theorem 5.2 is proved using induction, the Cauchy-Schwarz Inequality and a van der Corput lemma. This time we need a Hilbert space variation of this lemma:

Lemma 5.3 (**van der Corput Lemma, revisited** [3])**.** *Assume that $\mathcal{H}$ is a Hilbert space with inner product $\langle \ , \ \rangle$ and norm $\|\cdot\|$, and that ξ_n, $n \geq 0$, is a sequence in $\mathcal{H}$ with $\|\xi_n\| \leq 1$ for all n. Then*

$$\limsup_{N\to\infty} \left\| \frac{1}{N} \sum_{n=0}^{N-1} \xi_n \right\|^2 \leq \limsup_{H\to\infty} \frac{1}{H} \sum_{h=0}^{H-1} \limsup_{N\to\infty} \left| \frac{1}{N} \sum_{n=0}^{N-1} \langle \xi_{n+h}, \xi_n \rangle \right| \ .$$

By Theorem 5.2, one can consider an average along arithmetic progressions on an appropriate factor, rather than the whole space. We make this notion more precise.

For a measure preserving system $(X, \mathcal{X}, \mu, T)$, the word *factor* is used with two different but equivalent meanings. First, it is a T-invariant σ-algebra of $\mathcal{X}$. (Strictly speaking, this is a sub-σ-algebra, but throughout we omit the use of the word "sub".) Secondly, if $(Y, \mathcal{Y}, \nu, S)$ is a measure preserving system, a map $\pi \colon X \to Y$ is a *factor map* if π maps μ to ν and $S \circ \pi = \pi \circ T$. Then Y is said to be a factor of X and the two definitions coincide up to the identification of $\mathcal{Y}$ with $\pi^{-1}(\mathcal{Y})$. For $f \in L^1(\mu)$, we view $\mathbb{E}(f \mid \mathcal{Y})$ as a function on $\mathcal{X}$ and let $\mathbb{E}(f \mid Y)$ denote the function on Y defined by $\mathbb{E}(f \mid Y) \circ \pi = \mathbb{E}(f \mid \mathcal{Y})$. It is characterized by

$$\int_Y \mathbb{E}(f \mid Y)(y) \cdot g(y) \, d\nu(y) = \int_X f(x) \cdot g(\pi(x)) \, d\mu(x)$$

for all $g \in L^\infty(\mu)$.

The seminorms are used to define factors of the system $(X, \mathcal{X}, \mu, T)$. One presentation of these factors is by defining their orthogonal complements: for $d \geq 1$, define $\mathcal{Z}_{d-1}$ to be the σ-algebra of $\mathcal{X}$ such that for

$f \in L^\infty(\mu)$:

$$\|f\|_{U^d} = 0 \text{ if and only if } \mathbb{E}(f \mid \mathcal{Z}_{d-1}) = 0 \ .$$

Thus a bounded function f is measurable with respect to $\mathcal{Z}_{d-1}$ if and only if $\int fgd\mu = 0$ for all functions $g \in L^\infty(\mu)$ with $\|g\|_{U^{d-1}} = 0$. This motivates an equivalent definition of the factors $\mathcal{Z}_d$ with respect to a dual norm. Namely, defining the dual norm $\|f\|_{(U^d)^*}$ by

$$\|f\|_{(U^d)^*} := \sup_{g \in L^\infty(\mu)} \left\{ \int_X fg\, d\mu \colon \|g\|_{U^d} \leq 1 \right\} \ , \tag{5.4}$$

we have that the space of functions with finite $(U^d)^*$ norms is a dense subset (in L^2) of the bounded functions that are measurable with respect to $\mathcal{Z}_{d-1}$.

Letting Z_j denote the factor associated to the σ-algebra $\mathcal{Z}_j$, we have that Z_0 is the trivial factor and Z_1 is the *Kronecker* factor, meaning the σ-algebra which is spanned by the eigenfunctions of T. Furthermore, the sequence of factors is increasing (compare with Equation (4.6)):

$$Z_0 \leftarrow Z_1 \leftarrow Z_2 \leftarrow \ldots \leftarrow X$$

and if T is weakly mixing, then Z_d is the trivial factor for all d.

Theorem 5.2 states that the factor $\mathcal{Z}_{d-1}$ is a *characteristic factor* for the average of Equation (5.1), meaning that the limit behavior of the averages in $L^2(\mu)$ remains unchanged when each function is replaced by its conditional expectation on this factor. Thus it suffices to prove convergence when one of the factors Z_d is substituted for the original system. For a progression of length k, this amounts to decomposing a bounded function $f = g+h$ with $g = f - \mathbb{E}(f \mid \mathcal{Z}_{k-1})$. The function g is the uniform component and has zero $k-1$ seminorm and so contributes zero to the average along arithmetic progressions. The second term h is the anti-uniform component and belongs to the algebra of functions measurable with respect to the factor $\mathcal{Z}_{k-1}$ and must be analyzed via a structure theorem for the characteristic factors. This decomposition of an arbitrary bounded function into uniform and anti-uniform components is unique. In the combinatorial setting, a similar decomposition (see Section 6) can only be carried out approximately. Ergodic theory is more precise than combinatorics in describing the second component of this decomposition.

When the description of a characteristic factor is "simple", one has a better chance of proving convergence in this factor. For the given decomposition, the description of the characteristic factor is as an inverse limit of *nilsystems*, meaning that it can be approximated arbitrarily well by a rotation on a homogeneous space of a nilpotent Lie group.[6]

6. Quantitative ergodic theory

Tao [28] gave a new proof of Szemerédi's Theorem, along the lines of Furstenberg's original proof, but proving it in the finite system $\mathbb{Z}_N$ rather than for an arbitrary measure space. This allows him to extract explicit bounds for $N_0(\delta, k)$ in the finite version (Theorem 2.1), although the bounds are nowhere near as good as those obtained by Gowers [12].

Once again, a generalized von Neumann Theorem (analogous to Theorems 4.3 and 5.2) is used to start the proof. Then, as in the ergodic setup, an arbitrary bounded function f on $\mathbb{Z}_N$ is decomposed into pieces, each of which can be analyzed. This time the decomposition is into a term with small Gowers norm and a structured component, with the wrinkle that one also has to deal with a small error term. The first term corresponds to a uniform component $f - \mathbb{E}(f \,|\, \mathcal{Z})$ for a well chosen σ-algebra $\mathcal{Z}$ (similar to the use of a characteristic factor in the ergodic setup) which has small Gowers norm and makes a small contribution to the average in Equation (4.1). Since the space being used is $\mathbb{Z}_N$, the σ-algebra $\mathcal{Z}$ is nothing more than a finite partition of $\mathbb{Z}_N$: elements of a σ-algebra are unions of elements (also called *atoms*) of some partition of $\mathbb{Z}_N$ and a function is measurable with respect to this σ-algebra if it is constant on each element of the partition. The second term is the conditional expectation of f relative to $\mathcal{Z}$, meaning that it is the function measurable with respect to $\mathcal{Z}$ defined by

$$\mathbb{E}(f|\mathcal{Z})(x) = \frac{\mathbb{E}(\mathbf{1}_A(x) f(x) \,|\, x \in \mathbb{Z}_N)}{\mathbb{E}(\mathbf{1}_A(x) \,|\, x \in \mathbb{Z}_N)} \ ,$$

where A is the atom of $\mathcal{Z}$ containing x. This component is analyzed using a form of recurrence similar to that needed for a Furstenberg tower.

The second component of the decomposition, called the *anti-uniform* functions by Tao, is essentially dual to the uniform component where

[6] If G is a k-step nilpotent Lie group and Γ is a discrete cocompact subgroup, then $a \in G$ naturally acts on G/Γ by left translation by $T_a(x\Gamma) = (ax)\Gamma$. The Haar measure μ is the unique Borel probability measure on G/Γ that is invariant under this action of G by left translations. For a fixed element $a \in G$, the system $(G/\Gamma, \mathcal{G}/\Gamma, T_a, \mu)$ is a k-step nilsystem. The structure theorem in [21] states that the factor $\mathcal{Z}_{k-1}$, which is a characteristic factor for the average along an arithmetic progressions of length k, is an inverse limit of such $(k-1)$-step nilsystems.

the anti-uniform (dual) norm $\|g\|_{(U^d)^*}$ is defined by (compare with Equation (5.4))

$$\|g\|_{(U^d)^*} := \sup_{f\colon \mathbb{Z}_N \to \mathbb{C}} \{|\langle f, g\rangle| : \|f\|_{U^d} \leq 1\} \ .$$

The contribution of this term to the average is bounded from below by van der Waerden's Theorem,[7] with the idea being that these functions lie in a sufficiently compact space so that a finite coloring argument can be used. Applying this idea to a function with positive expectation, the average along arithmetic progressions is positive.

This proof follows Furstenberg's proof closely. One advantage is the elimination of the compactness argument, leading to explicit bounds on the size of the set needed to guarantee the existence of a progression of length k. The structure theorem corresponds to the tower of compact extensions used by Furstenberg and does not need an understanding of the precise structure of the chosen σ-algebra, such as the nilsystems in the structure theorem of Host and Kra. However, a more precise understanding of this structure should clarify the apparent link between the anti-uniform functions of level k appearing in Tao's proof and the k-step nilsystems used to prove Theorem 5.1.

7. Arithmetic progressions in the primes

Green and Tao continue in this vein to prove the existence of arithmetic progressions in the primes. The starting point is clear: study the averages of Equation (4.1) for the indicator function of the primes. However, since the primes have density 0, any function that is 0 other than on the primes can not be bounded without its average on $\mathbb{Z}_N$ becoming arbitrarily small as N tends to infinity. Since such a function cannot be bounded independently of N, Szemerédi's Theorem can not be applied directly.

Instead, Green and Tao begin with the closely related *von Mangoldt function* $\Lambda(n)$, where

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

and make use of the fact that this function is more natural analytically than Λ. Although the von Mangoldt function is supported on the primes and their powers, the powers are sparsely enough distributed so

[7] Van der Waerden's Theorem [31] states that if the integers are partitioned into finitely many pieces, then one of these pieces contains arbitrarily long arithmetic progressions. This theorem motivated Erdös and Turán [6] to conjecture Szemerédi's Theorem.

that they only contribute a small error term in the calculations. This function has had many uses in number theory; for example, the unique factorization theorem for integers is equivalent to the statement

$$\log n = \sum_{d|n} \Lambda(d) \text{ for all positive integers } n \ ,$$

and the Prime Number Theorem is equivalent to the statement

$$\frac{1}{N} \sum_{1 \le n \le N} \Lambda(n) = 1 + o(1) \ .$$

(Throughout, by $o(1)$, we mean a quantity that tends to 0 as $N \to \infty$, and when this quantity depends on other constants, we include them as subscripts on o.)

The function Λ mostly avoids giving weight to arithmetic progressions congruent to $a \mod q$ when a and q are not relatively prime. Such arithmetic progressions are more dense when q has many small prime factors, making Λ too irregularly distributed for their purposes. Therefore Green and Tao are forced to modify Λ, quotienting out the small primes and make it more evenly distributed over all congruence classes. They then majorize the modified function by something *pseudorandom* and much of the work is carried out for pseudorandom functions. The precise definition and modification is given below, but the idea is that the values of a pseudorandom function should be distributed so that using any statistic to measure the values, one gets approximately the same measurement as that arising from a random set of the same density.

The goal then becomes to extend Szemerédi's Theorem, showing that not only does a dense subset of the integers contain arbitrarily long arithmetic progressions, but a dense subset of a pseudorandom collection of integers also contains arbitrarily long arithmetic progressions. Green and Tao do this by "transferring" Szemerédi's Theorem to a more general setting: the hypothesis in Theorem 4.1 that $f\colon \mathbb{Z}_N \to \mathbb{R}$ satisfies $0 \le f(x) \le 1$ is replaced by f being bounded by a more general function $\nu\colon \mathbb{Z}_N \to \mathbb{R}^+$ with certain useful properties. (More precisely, for each $N \in \mathbb{N}$ we have a function $\nu = \nu_N\colon \mathbb{Z}_N \to \mathbb{R}^+$.) The function $\nu\colon \mathbb{Z}_N \to \mathbb{R}^+$ is assumed to be a *measure*,[8] meaning that $\mathbb{E}(\nu(x) \,|\, x \in \mathbb{Z}_N) = 1 + o(1)$, and ν is also assumed to be pseudorandom. They show:

[8] As noted by Green and Tao, the name measure is a misnomer, and ν should more accurately be called a density relative to the uniform measure on $\mathbb{Z}_N$.

Theorem 7.1 (**Transference Theorem** [15])**.** *Let $0 < \delta \leq 1$ be a real number and let $k \geq 2$ be an integer. If N is sufficiently large, $\nu\colon \mathbb{Z}_N \to \mathbb{R}^+$ is a k-pseudorandom measure, and $f\colon \mathbb{Z}_N \to \mathbb{R}$ is function with $0 \leq f(x) \leq \nu(x)$ for all $x \in \mathbb{Z}_N$ and $\mathbb{E}(f(x) \,|\, x \in \mathbb{Z}_N) \geq \delta$, then*

$$\mathbb{E}\big(f(x)f(x+r)\dots f(x+(k-1)r) \,|\, x, r \in \mathbb{Z}_N\big) \geq c(k,\delta) - o_{k,\delta}(1) \ , \tag{7.1}$$

where the constant $c(k,\delta)$ is the same as that in Theorem 4.1.

Other than the bounds on f, the only additional modification caused by bounding f by a pseudorandom measure instead of the constant function 1 is the introduction of the error term $o_{k,\delta}(1)$, which tends to 0 as $N \to \infty$. The dependence of this error is only on k and δ.

Before giving an indication of the proof of Theorem 7.1, we make the notion of a pseudorandom measure more precise. (A reader not interested in the technical details can skip the next few paragraphs.) The measure $\nu\colon \mathbb{Z}_N \to \mathbb{R}^+$ is said to be *k-pseudorandom* if ν satisfies a *k-linear forms condition* and a *k-correlation condition.*

To define the linear forms condition, fix k, the length of the arithmetic progression and assume that N is prime and larger than k. Assume that we have m linear forms ψ_i, $1 \leq i \leq m$, with $m \leq k \cdot 2^{k-1}$ and t variables with $t \leq 3k - 4$. (The exact values of these constants are not important for the proof; the importance lies in showing that a particular choice of a pseudorandom function satisfies these conditions. For this, it only matters that the values depend on nothing but k.) Let $L = (L_{ij})$ be an $m \times t$ matrix, whose entries are rational numbers with numerator and denominator bounded in absolute value by k. By choice of N, we can view the entries of L as elements of $\mathbb{Z}_N$ (recall that N is prime). Assume further that each of the t columns of L are not identically zero and that the columns are pairwise independent. Let $\psi_i(\mathbf{x}) = b_i + \sum_{j=1}^{t} L_{ij}x_j$ denote the m linear forms, where $\mathbf{x} \in \mathbb{Z}_N^t$ and $b_i \in \mathbb{Z}_N$ for $1 \leq i \leq m$. The measure $\nu\colon \mathbb{Z}_N \to \mathbb{R}^+$ is said to satisfy the (m,t,L)-linear forms condition if

$$\mathbb{E}\big(\nu\left(\psi_1(\mathbf{x})\right)\dots\nu(\psi_m(\mathbf{x})) \,|\, \mathbf{x} \in \mathbb{Z}_N^t\big) = 1 + o_{m,t,L}(1) \ ,$$

where the dependence on N is assumed to be uniform in the choice of the linear forms ψ_i and so in particular uniform in the choice of the b_i. The case $m = 1$ with $\psi(x) = x$ corresponds to the measure ν with $\mathbb{E}(\nu) = 1 + o(1)$, and this is the bound used in the Reformulated Szemerédi Theorem (Theorem 4.1). For higher m, the values of the measure ν evaluated on linear forms up to a certain complexity are, on average, independent. If there were no restriction on the complexity, the measure would be close to the ergodic theoretic notion of weak mixing, meaning its values along any distinct linear forms would be, on average, independent.

We now define the correlation condition. The measure $\nu\colon \mathbb{Z}_N \to \mathbb{R}^+$ is said to satisfy a 2^{k-1}-correlation condition if for each m with $1 \leq m \leq 2^{k-1}$, there exists a weight function $\tau = \tau_m\colon \mathbb{Z}_N \to \mathbb{R}^+$ with

$$\mathbb{E}(\tau^q \mid z \in \mathbb{Z}_N) \leq C(m, q)$$

for a constant $C(m, q)$, for all $1 \leq q < \infty$, and that

$$\mathbb{E}\big(\nu(x+h_1)\nu(x+h_2)\ldots\nu(x+h_m) \mid x \in \mathbb{Z}_N\big) \leq \sum_{1\leq i<j\leq m} \tau(h_i - h_j)$$

for all $h_1, h_2, \ldots, h_m \in \mathbb{Z}_N$.

The correlation condition arises in Goldston and Yildirim's [17] work and is used for specific estimates applied to the prime numbers. Although the linear forms condition does not arise in their work, fortunately their estimates also apply to ν satisfying this condition.

In some sense, the Transference Theorem can be thought of as a generalization of Furstenberg's Multiple Recurrence Theorem. In the ergodic set up, a natural choice of measure is the uniform one, assigning each integer in $\{1, \ldots, N\}$ the equal weight $1/N$. This measure is invariant with respect to the shift map $x \mapsto x+1 \mod N$. In Green and Tao's generalization, the measure behaves in a pseudorandom manner with respect to the shift. For a certain choice of R (discussed below), to each number in $\{1, \ldots, N\}$ having no prime factors less than R, the new measure assigns the weight $\log R/N$, and in order to make the measure more regular, it assigns an appropriately chosen small value to each of the other numbers in $\{1, \ldots, N\}$.

Lending credence to the idea that Szemerédi's Theorem should hold for a function bounded by a pseudorandom measure is the fact that a pseudorandom measure is close to the constant function 1 in Gowers norm:

Lemma 7.2 ([15])**.** *Fix an integer $k \geq 1$, let $N > k$ be a prime number, and assume that $\nu\colon \mathbb{Z}_N \to \mathbb{R}^+$ is a k-pseudorandom measure. Then*

$$\|\nu - 1\|_{U^d} = o(1)$$

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

The broad outline of the proof of Theorem 7.1 is similar to that of Tao's proof of Szemerédi's Theorem sketched in the last section, but both the technical details and the combination of ideas from seemingly unrelated areas of mathematics make it a significantly more ambitious undertaking. The innovation is the reduction of Theorem 7.1 to Szemerédi's Theorem. The key argument, again, is a structure theorem, but this time not only is there an error term in the decomposition, but the decomposition is only valid on most of the space. Green and Tao show:

Theorem 7.3 (**Decomposition Theorem** [15])**.** *Let $k \geq 2$ be an integer, let $0 < \epsilon \ll 1$ be a small parameter, and let $N = N(\epsilon)$ be sufficiently large. Assume that $\nu\colon \mathbb{Z}_N \to \mathbb{R}^+$ is a k-pseudorandom measure and that $f \in L^1(\mathbb{Z}_N)$ is a function satisfying $0 \leq f(x) \leq \nu(x)$ for all $x \in \mathbb{Z}_N$. Then there exists a σ-algebra $\mathcal{Z}$ and an exceptional set $\Omega \in \mathcal{Z}$ with $\mathbb{E}(\nu(x)\mathbf{1}_\Omega(x) \,|\, x \in \mathbb{Z}_N) = o_\epsilon(1)$ such that*

$$\|\mathbf{1}_{\Omega^C}\mathbb{E}(\nu - 1 \,|\, \mathcal{Z})\|_{L^\infty} = o_\epsilon(1)$$

and

$$\|\mathbf{1}_{\Omega^C}(f - \mathbb{E}(f \,|\, \mathcal{Z}))\|_{U^{k-1}} \leq \epsilon^{1/2^k} \ ,$$

where Ω^C denotes the complement of Ω.

This means that outside a small subset Ω of $\mathbb{Z}_N$, a function f that is bounded by a pseudorandom measure can be decomposed into a sum of a uniform function g and an anti-uniform function h, plus a small error term. The function g has small Gowers norm and corresponds to $f - \mathbb{E}(f \,|\, \mathcal{Z})$ in the ergodic theoretic setup, while the non-negative function h is bounded and corresponds to $\mathbb{E}(f \,|\, \mathcal{Z})$. Other than the error terms, this parallels the ergodic theoretic decomposition associated to a characteristic factor described in Section 5 and the decomposition used by Tao described in Section 6.

The proof of the Decomposition Theorem follows a complicated iterative procedure, designed to produce the σ-algebra $\mathcal{Z}$. Starting with the trivial σ-algebra $\mathcal{B} = \{\emptyset, \mathbb{Z}_N\}$, if the function $f - \mathbb{E}(f|\mathcal{B})$ has small U^{k-1} norm, the algorithm terminates. If not, an appropriate addition is made to the σ-algebra $\mathcal{B}$, taking care to increase $\mathbb{E}(f|\mathcal{B})$, yet keep it uniformly bounded. This process is repeated until $f - \mathbb{E}(f|\mathcal{B})$ is sufficiently small and the algorithm terminates.

The next ingredient in the proof of Theorem 7.1 is analogous to the generalized von Neumann Theorem; it gives a way to control the contribution of the Gowers uniform portion in the decomposition, meaning a way to bound the contribution coming from a function with small Gowers norm. Once again, the bound on the functions changes: instead of being bounded by the constant 1, the functions are now bounded pointwise by 1 plus a pseudorandom measure.

Theorem 7.4 (**Pseudorandom Generalized von Neumann Theorem** [15])**.** *Let $k \geq 2$ be an integer, let N be a prime number, and* *assume that* $\nu\colon \mathbb{Z}_N \to \mathbb{R}^+$ *is a* k*-pseudorandom measure. Assume that* $f_0, \dots, f_{k-1} \in L^1(\mathbb{Z}_N)$ *are functions such that*

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

Then

$$\begin{aligned}\big|\mathbb{E}\big(f_0(x)f_1(x+n)\dots f_{k-1}(x+(k-1)n) \,|\, x,n \in \mathbb{Z}_N\big)\big| \\ = O\left(\min_{0\leq j\leq k-1} \|f_j\|_{U^{k-1}}\right) + o(1) \ .\end{aligned}$$

We are now ready to outline the proof of Theorem 7.1, still glossing over many technical details. We fix a function f that is bounded by a pseudorandom measure and that has positive expectation on $\mathbb{Z}_N$. Using the Decomposition Theorem, the expectation on the left hand side of Equation (7.1), which is the average along arithmetic progressions, is larger than the same expectation with $\mathbf{1}_{\Omega^C} f$ substituted for f, where Ω is some small set. Ignoring the error term, we now use the decomposition of this new function into $g + h$, where g is the Gowers uniform portion and h is some bounded function. Much like the idea of a characteristic factor in ergodic theory, we now want to discard the Gowers uniform portion g and replace our function by h. Making the substitution $g + h$ for $\mathbf{1}_{\Omega^C} f$, the expectation on the left hand side of Equation (7.1) can be expanded as a sum of 2^k expectations of the form

$$\mathbb{E}(f_0(x)f_1(x+n)\dots f_{k-1}(x+(k-1)n \,|\, x,n \in \mathbb{Z}_N) \ ,$$

where each f_i is equal either to g or to h. All terms but one contain an occurrence of g in it and each term containing a g is small by the Pseudorandom Generalized von Neumann Theorem. We are left only with a single term making a large contribution to the expectation, which is the only term with nothing but occurrences of the function h. The good news is that now this function h is bounded and so the usual Szemerédi Theorem applies. Furthermore, f and h have approximately the same expectation, and in particular the expectation on $\mathbb{Z}_N$ of h is also positive. Thus by Szemerédi's Theorem, the expectation in Equation (7.1) with f replaced by h is positive. Therefore, the same result holds for f.

Lastly we give an indication of the choice of the function f and measure ν needed to use the Transference Theorem for the primes. The function f is a variation on the von Mangoldt function, cut off at a certain point, in order to make a function that is (vaguely speaking) supported on primes of magnitude $\log N$. Unfortunately, it does not suffice to simply use a multiple of this function for ν, since as we noted earlier, the primes, and therefore any multiple of the von Mangoldt function, are not uniformly distributed across all residue classes, whereas a pseudorandom function is. Instead, the measure ν is taken to have its support (again, vaguely speaking) on numbers n such that all the prime factors of $n-1$ are greater than some integer R. One can view this measure as approximately $\log R$ times the characteristic function of such numbers.

In the third century B.C., the scholar Eratosthenes came up with a simple algorithm for listing all the prime numbers up to a given N, referred to as the sieve of Eratosthenes. Given a list of the numbers between 1 and N, starting with 2, erase all multiples of 2 up to N, other thatn 2 itself. Call the remaining set P_2. Returning to the beginning, take the first number greater than 2 and erase all of its multiples up to N, again other than the number itself. In general, the *level R almost primes* $P_R(N)$ are defined to be the set of all numbers between 1 and N that contain no nontrivial factors less than or equal to R. Thus if $R=\sqrt{N}$, we have that $P_{\sqrt{N}}(N)$ consists exactly of the prime numbers up to N. Mertens [24] proved that the size $|P_R(N)|$ is approximately $cN/\log R$ for some positive constant c. Combining this with the estimate from the Prime Number Theorem that the number of primes up to N is approximately $N/\log N$, we have that the density of primes in the almost primes $P_R(N)$ is about a multiple of $\log R/\log N$. Therefore if R is a small power of N, then the primes have positive density in the level R almost primes. This motivates the function and measure Green and Tao use. For completeness, we give the technical definitions.

Let W be the product of the primes up to some $\omega=\omega(N)$, where $\omega(N)$ tends to infinity sufficiently slowly so that $\tilde{\Lambda}$ has mean one ($\omega(N)=\log\log N$ suffices). Taking place of the indicator function of the primes is the *modified von Mangoldt function* $\tilde{\Lambda}$, defined by

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

Thus the function $\tilde{\Lambda}$ is supported on the set

$$Q_W=\{n\in\mathbb{Z}\colon Wn+1 \text{ is prime}\}\ .$$

It suffices to find an arithmetic progression in Q_W, since if

$$\{x,x+n,\dots,x+(k-1)n\}$$

is an arithmetic progression in Q_W, then

$$\{Wx+1,W(x+n)+1,\dots,W(x+(k-1)n)+1\}$$

is an arithmetic progression in the primes with common difference Wn. This modification is needed because the primes bounded by x are not uniformly spread out in arithmetic progressions. For example, there is only one prime congruent to $0 \mod 2$, while there are approximately $x/\log x$ congruent to $1 \mod 2$. Furthermore, if a and q are relatively prime integers, the number of primes in the arithmetic progression $a \mod q$ up to x is approximately $\frac{x}{\log x} \cdot \frac{1}{\phi(q)}$. If one considers integers n with $n \equiv a \mod q$ and for which $Wn+1$ is prime, then there are none only when q and $Wa+1$ are not relatively prime and this can only happen when q and W are relatively prime. This means that q has no small prime factors and the set Q_W is more uniformly distributed among the arithmetic progressions.

Green and Tao do not show directly that $\tilde{\Lambda}$ is k-pseudorandom, but instead majorize it by a measure ν whose values are more uniformly distributed and then are able to show that ν is k-pseudorandom. Before defining the measure ν, we need one more variation on the von Mangoldt function. The *truncated von Mangoldt function* is defined to be

$$\Lambda_R(n) = \sum_{d|n, d \leq R} \mu(d) \log(R/d) \ ,$$

where μ is the Möbius function.[9] The restriction $d \leq R$ is needed to guarantee that the log remains positive. This is a cut off version of the von Mangoldt function, since if $R > n$ then $\Lambda_R(n) = \Lambda(n)$. This useful approximation to $\Lambda(n)$ has been widely used in analytic number theory, notably by Selberg [26] to give strong upper bounds for the number of primes predicted by an application of the Hardy-Littlewood k-tuple Conjecture. More recently, Goldston and Yildirim [17] use it in their work on gaps in the primes.

Fix $R = N^{k^{-1}2^{-k-4}}$. (One can think of this choice of R as N^{ϵ} for some $\epsilon < 1$.) The measure $\nu\colon \mathbb{Z}_N \to \mathbb{R}^+$ is defined for $0 \leq n < N$ to be

$$\nu(n) = \begin{cases} \frac{\phi(W)}{W} \frac{(\Lambda_R(Wn+1))^2}{\log R} & \text{for } N/(2^k(k+4)!) \leq n \leq 2N/(2^k(k+4)!) \\ 1 & \text{otherwise} \ , \end{cases}$$

[9] The *Möbius function* $\mu(n)$ is defined by $\mu(1) = 1$, $\mu(n) = 0$ if n is not a square free integer and $\mu(n) = (-1)^r$ if n is a square free integer and has r distinct prime factors.

where $\{0, 1, \ldots, N-1\}$ is naturally identified with $\mathbb{Z}_N$ and ϕ denotes the Euler totient function.[10] We make a few comments on this definition. The bounds on n are designed to avoid counting progressions that are not progressions $\mathbb{Z}$ because they wrap around 0 in $\mathbb{Z}_N$, and the factors $\phi(w)/w$ guarantee that one has the correct average as $N \to \infty$. Since we are interested in majorizing $\tilde{\Lambda}$, the function Λ_R is evaluated at $Wn + 1$.

One can quickly verify that this choice of ν majorizes the modified von Mangoldt function $\tilde{\Lambda}$. The last major step is verifying that ν is k-pseudorandom. This relies on techniques from analytic number theory, using and extending recent results of Goldston and Yildirim [17] on finding small gaps between primes. An alternate approach to this portion of the proof is given by Tao in [29], using only elementary properties of the primes and basic properties of the Riemann ζ function.

8. Further directions

At this time, Green and Tao's Theorem seems out of the reach of ergodic theory. All combinatorial number theorems that have been proved using ergodic theory rely on some variant of Furstenberg's Correspondence Principle, which only applies to sets of integers with positive upper density. However, the many similarities between Green and Tao's approach and proofs in ergodic theory suggest that a connection exists. The ultimate goal would be to use translations of the proof techniques of Green and Tao to obtain new convergence results in ergodic theory; in particular one may be able to use ergodic theory to show the existence of some patterns in certain subsets of density zero.

Tao's proof [28] of Szemerédi's Theorem removes the compactness argument needed in Furstenberg's proof and replaces it by a lengthy induction. This induction only needs finitely many steps, but the number of steps is not explicitly known. A better understanding of the structure theorem used would probably improve the bounds that Tao extracts with this method. It seems that finding the exact link between the anti-uniform functions of level k and the k-step nilsystems found in the work of Host and Kra [21] would clarify the connections between the two fields and probably lead to new and interesting developments.

A natural question arises from these considerations. Bergelson and Leibman [4] used ergodic theory to establish a polynomial Szemerédi

[10] The *Euler totient function* $\phi(n)$ is defined to be the number of positive integers less than or equal to n that are relatively prime to n, with 1 being counted as relatively prime to all numbers.

type theorem [11] and perhaps it is possible to carry out a similar program to that of Green and Tao for this situation. Namely, prove a transference polynomial Szemerédi Theorem and show that not only do subsets of the integers with positive upper density contain polynomial patterns, but also dense subsets of pseudorandom sets contain polynomial patterns. This would prove, for example, that there exist infinitely many triples (p, k, n) of integers with $p, n \geq 1$ and $k > 1$ such that $p, p+n, p+n^2, \dots, p+n^k$ consists only of prime numbers. Finding polynomial patterns in the primes seems to have the added difficulty of lifting the result from $\mathbb{Z}_N$ to $\mathbb{Z}$.

One can also hope to use some variation of Green and Tao's method to study the existence of other patterns in the primes, such as pairs of primes $p, p+2$ (the twin prime conjecture) or pairs of primes $p, 2p+1$. Closer in spirit to arithmetic progressions, one might look for infinitely many pairs (p, d) with p a prime and d a positive integer such that $p, p+2d, p+4d, 2p+d$ are all prime.

References

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

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

[3] V. Bergelson. Weakly mixing PET. *Erg. Th. & Dyn. Sys.*, **7** (1987), 337–349.

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

[5] J. -P. Conze and E. Lesigne. Sur un théorème ergodique pour des mesures diagonales. *C. R. Acad. Sci. Paris, Série I*, **306** (1988), 491–493.

[6] P. Erdös and P. Turán. On some sequences of integers. *J. Lond. Math. Soc.*, **11** (1936), 261-264.

[11] More precisely, Bergelson and Leibman's Theorem [4] generalizes Furstenberg's Multiple Ergodic Theorem (Theorem 3.1). They show that if $(X, \mathcal{X}, \mu, T)$ is an invertible measure preserving probability system, $k \geq 1$ is an integer, $p_1, \dots, p_k$ are polynomials taking integer values on the integers with $p_1(0) = \dots = p_k(0) = 0$, and $A \in \mathcal{X}$ with $\mu(A) > 0$, then

$$\liminf_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mu\big(T^{p_1(n)}A \cap \ldots \cap T^{p_k(n)}A\big) > 0 \ .$$

As for arithmetic progressions, the limits of the related averages

$$\frac{1}{N} \sum_{n=0}^{N-1} f_1(T^{p_1(n)}x) \ldots f_k(T^{p_k(n)}x)$$

for bounded functions $f_1, \dots, f_k$ are known to exist in $L^2(\mu)$ (see [22] and [23]).

[7] M. Frind, P. Jobling and P. Underwood. 23 primes in arithmetic progression. Available at http://primes.plentyoffish.com/

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

[9] H. Furstenberg and Y. Katznelson. An ergodic Szemerédi theorem for commuting transformations. *J. d'Analyse Math.*, **34** (1979), 275–291.

[10] H. Furstenberg and Y. Katznelson. A density version of the Hales-Jewett Theorem. *J. d'Analyse Math.*, **57** (1991), 64–119.

[11] H. Furstenberg and B. Weiss. A mean ergodic theorem for $\frac{1}{N}\sum_{n=1}^{n} f(T^n x)g(T^{n^2}x)$. In *Convergence in Ergodic Theory and Probability*, Walter de Gruyter & Co, New York, 1996, 193–227.

[12] T. Gowers. A new proof of Szemerédi's Theorem. *GAFA*, **11** (2001), 465–588.

[13] A. Granville. Personal communication.

[14] B. Green. Roth's Theorem in the primes. To appear, *Ann. Math.*

[15] B. Green and T. Tao. The primes contain arbitrarily long arithmetic progressions. To appear, *Ann. Math.*

[16] B. Green and T. Tao. A bound for progressions of length k in the primes. Preprint.

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

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

[19] D. R. Heath-Brown. Three primes and an almost prime in arithmetic progression. *J. Lond. Math. Soc. (2)*, **23** (1981), 396–414.

[20] B. Host and B. Kra. Convergence of Conze-Lesigne Averages. *Erg. Th. & Dyn. Sys.*, **21** (2001), 493–509.

[21] B. Host and B. Kra. Nonconventional ergodic averages and nilmanifolds. *Ann. Math.*​**161** (2005), 397–488.

[22] B. Host and B. Kra. Convergence of polynomial ergodic averages. *Isr. J. Math.*​**149** (2005), 1–19.

[23] A. Leibman. Convergence of multiple ergodic averages along polynomials of several variables. *Isr. J. Math.*​**146**, 303–316.

[24] F. Mertens. Ein Beitrag zur analytischen Zahlentheorie. *Journal für Math.*, **78** (1874), 46–62.

[25] A. Moran, P. Pritchard and A. Thyssen. Twenty-two primes in arithmetic progression. *Math. Comp.*, **64** (1995), 1337–1339.

[26] A. Selberg. The general sieve method and its place in prime number theory. *Proc. ICM*, vol 1, Cambridge (1950), 286–292.

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

[28] T. Tao. A quantitative ergodic theory proof of Szemerédi's theorem. Preprint.

[29] T. Tao. A remark on Goldston-Yildirim correlation estimates. Preprint.

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

[31] B. L. van der Waerden. Beweis einer Baudetschen Vermutung. *Nieuw Arch. Wisk.*, **15** (1927), 212–216.

[32] T. Ziegler. Universal characteristic factors and Furstenberg averages. Preprint.

Department of Mathematics, Northwestern University, 2033 Sheridan Road, Evanston, IL 60208-2730

E-mail address: kra@math.northwestern.edu