Clay Mathematics Proceedings
Volume 7, 2007

Long arithmetic progressions of primes

Ben Green

Abstract. This is an article for a general mathematical audience on the author's work, joint with Terence Tao, establishing that there are arbitrarily long arithmetic progressions of primes.

1. Introduction and history

This is a description of recent work of the author and Terence Tao [**GTc**] on primes in arithmetic progression. It is based on seminars given for a general mathematical audience in a variety of institutions in the UK, France, the Czech Republic, Canada and the US.

Perhaps curiously, the order of presentation is much closer to the order in which we discovered the various ingredients of the argument than it is to the layout in [**GTc**]. We hope that both expert and lay readers might benefit from contrasting this account with [**GTc**] as well as the expository accounts by Kra [**Kra06**] and Tao [**Tao06a, Tao06b**].

As we remarked, this article is based on lectures given to a general audience. It was often necessary, when giving these lectures, to say things which were not strictly speaking true for the sake of clarity of exposition. We have retained this style here. However, it being undesirable to commit false statements to print, we have added numerous footnotes alerting readers to points where we have oversimplified, and directing them to places in the literature where fully rigorous arguments can be found.

Our result is:

Theorem 1.1 (G.–Tao). *The primes contain arbitrarily long arithmetic progressions.*

Let us start by explaining that the truth of this statement is not in the least surprising. For a start, it is rather easy to write down a progression of five primes (for example $5, 11, 17, 23, 29$), and in 2004 Frind, Jobling and Underwood produced

2000 *Mathematics Subject Classification.* Primary 11N13, Secondary 11B25.

© 2007 Ben Green

the example

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

of 23 primes in arithmetic progression. A very crude heuristic model for the primes may be developed based on the prime number theorem, which states that $\pi(N)$, the number of primes less than or equal to N, is asymptotic to $N/\log N$. We may alternatively express this as

$$\mathbb{P}\big(x \text{ is prime} \mid 1 \leqslant x \leqslant N\big) \sim 1/\log N.$$

Consider now the collection of all arithmetic progressions

$$x, x + d, \dots, x + (k-1)d$$

with $x, d \in \{1, \dots, N\}$. Select x and d at random from amongst the N^2 possible choices, and write E_j for the event that $x + jd$ is prime, for $j = 0, 1, \dots, k-1$. The prime number theorem tells us that

$$\mathbb{P}(E_j) \approx 1/\log N.$$

If the events E_j were independent we should therefore have

$$\mathbb{P}(x, x + d, \dots, x + (k-1)d \text{ are all prime}) = \mathbb{P}\big(\bigwedge_{j=0}^{k-1} E_j\big) \approx 1/(\log N)^k.$$

We might then conclude that

$$\#\{x, d \in \{1, \dots, N\} : x, x + d, \dots, x + (k-1)d \text{ are all prime } \} \approx \frac{N^2}{(\log N)^k}.$$

For fixed k, and in fact for k nearly as large as $2 \log N/\log\log N$, this is an increasing function of N. This suggests that there are infinitely many k-term arithmetic progressions of primes for any fixed k, and thus arbitrarily long such progressions.

Of course, the assumption that the events E_j are independent was totally unjustified. If E_0, E_1 and E_2 all hold then one may infer that x is odd and d is even, which increases the chance that E_3 also holds by a factor of two. There are, however, more sophisticated heuristic arguments available, which take account of the fact that the primes $> q$ fall only in those residue classes $a \pmod q$ with a coprime to q. There are very general conjectures of Hardy-Littlewood which derive from such heuristics, and a special case of these conjectures applies to our problem. It turns out that the extremely naïve heuristic we gave above only misses the mark by a constant factor:

Conjecture 1.2 (Hardy-Littlewood conjecture on k-term APs). *For each k we have*

$$\#\{x, d \in \{1, \dots, N\} : x, x+d, \dots, x+(k-1)d \text{ are all prime } \} = \frac{\gamma_k N^2}{(\log N)^k}(1+o(1)),$$

where

$$\gamma_k = \prod_p \alpha_p^{(k)}$$

is a certain product of "local densities" which is rapidly convergent and positive.

We have

$$\alpha_p^{(k)} = \begin{cases} \frac{1}{p}\left(\frac{p}{p-1}\right)^{k-1} & \text{if } p \leqslant k \\ \left(1 - \frac{k-1}{p}\right)\left(\frac{p}{p-1}\right)^{k-1} & \text{if } p \geqslant k. \end{cases}$$

In particular we compute[1]

$$\gamma_3 = 2\prod_{p \geqslant 3}\left(1 - \frac{1}{(p-1)^2}\right) \approx 1.32032$$

and

$$\gamma_4 = \frac{9}{2}\prod_{p \geqslant 5}\left(1 - \frac{3p-1}{(p-1)^3}\right) \approx 2.85825.$$

What we actually prove is a somewhat more precise version of Theorem 1.1, which gives a lower bound falling short of the Hardy-Littlewood conjecture by just a constant factor.

THEOREM 1.3 (G.–Tao). *For each $k \geqslant 3$ there is a constant $\gamma_k' > 0$ such that*

$$\#\{x, d \in \{1, \dots, N\} : x, x+d, \dots, x+(k-1)d \text{ are all prime }\} \geqslant \frac{\gamma_k' N^2}{(\log N)^k}$$

for all $N > N_0(k)$.

The value of γ_k' we obtain is very small indeed, especially for large k.

Let us conclude this introduction with a little history of the problem. Prior to our work, the conjecture of Hardy-Littlewood was known only in the case $k = 3$, a result due to Van der Corput [**vdC39**] (see also [**Cho44**]) in 1939. For $k \geqslant 4$, even the existence of infinitely many k-term progressions of primes was not previously known. A result of Heath-Brown from 1981 [**HB81**] comes close to handling the case $k = 4$; he shows that there are infinitely many 4-tuples $q_1 < q_2 < q_3 < q_4$ in arithmetic progression, where three of the q_i are prime and the fourth is either prime or a product of two primes. This has been described as "infinitely many $3\frac{1}{2}$-term arithmetic progressions of primes".

2. The relative Szemerédi strategy

A number of people have noted that [**GTc**] manages to avoid using any deep facts about the primes. Indeed the only serious number-theoretical input is a zero-free region for ζ of "classical type", and this was known to Hadamard and de la Vallée Poussin over 100 years ago. Even this is slightly more than absolutely necessary; one can get by with the information that ζ has an isolated pole at 1 [**Taoa**].

Our main advance, then, lies not in our understanding of the primes but rather in what we can say about *arithmetic progressions*. Let us begin this section by telling a little of the story of the study of arithmetic progressions from the combinatorial point of view of Erdős and Turán [**ET36**].

[1]For a tabulation of values of γ_k, $3 \leqslant k \leqslant 20$, see [**GH79**]. As $k \to \infty$, $\log \gamma_k \sim k \log\log k$.

DEFINITION 2.1. Fix an integer $k \geqslant 3$. We define $r_k(N)$ to be the largest cardinality of a subset $A \subseteq \{1, \dots, N\}$ which does not contain k distinct elements in arithmetic progression.

Erdős and Turán asked simply: what is $r_k(N)$? To this day our knowledge on this question is very unsatisfactory, and in particular we do not know the answer to

QUESTION 2.2. Is it true that $r_k(N) < \pi(N)$ for $N > N_0(k)$?

If this is so then the primes contain k-term arithmetic progressions on density grounds alone, irrespective of any additional structure that they might have. I do not know of anyone who seriously doubts the truth of this conjecture, and indeed all known lower bounds for $r_k(N)$ are much smaller than $\pi(N)$. The most famous such bound is Behrend's assertion [**Beh46**] that

$$r_3(N) \gg Ne^{-c\sqrt{\log N}};$$

slightly superior lower bounds are known for $r_k(N)$, $k \geqslant 4$ (cf. [**ŁL, Ran61**]).

The question of Erdős and Turán became, and remains, rather notorious for its difficulty. It soon became clear that even seemingly modest bounds should be regarded as great achievements in combinatorics. The first really substantial advance was made by Klaus Roth, who proved

THEOREM 2.3 (Roth, [**Rot53**]). *We have* $r_3(N) \ll N(\log\log N)^{-1}$.

The key feature of this bound is that $\log\log N$ tends to infinity with N, albeit slowly[2]. This means that if one fixes some small positive real number, such as 0.0001, and then takes a set $A \subseteq \{1, \dots, N\}$ containing at least $0.0001N$ integers, then provided N is sufficiently large this set A will contain three distinct elements in arithmetic progression.

The generalisation of this statement to general k remained unproven until Szemerédi clarified the issue in 1969 for $k = 4$ and then in 1975 for general k. His result is one of the most celebrated in combinatorics.

THEOREM 2.4 (Szemerédi [**Sze69, Sze75**]). *We have* $r_k(N) = o(N)$ *for any fixed* $k \geqslant 3$.

Szemerédi's theorem is one of many in this branch of combinatorics for which the bounds, if they are ever worked out, are almost unimaginably weak. Although it is in principle possible to obtain an explicit function $\omega_k(N)$, tending to zero as $N \to \infty$, for which

$$r_k(N) \leqslant \omega_k(N)N,$$

to my knowledge no-one has done so. Such a function would certainly be worse than $1/\log_* N$ (the number of times one must apply the log function to N in order to get a number less than 2), and may even be slowly-growing compared to the inverse of the Ackermann function.

The next major advance in the subject was another proof of Szemerédi's theorem by Furstenberg [**Fur77**]. Furstenberg used methods of ergodic theory, and

[2]cf. the well-known quotation "$\log\log\log N$ has been proved to tend to infinity with N, but has never been observed to do so".

his argument is relatively short and conceptual. The methods of Furstenberg have proved very amenable to generalisation. For example in [**BL96**] Bergelson and Leibman proved a version of Szemerédi's theorem in which arithmetic progressions are replaced by more general configurations $(x + p_1(d), \dots, x + p_k(d))$, where the p_i are polynomials with $p_i(\mathbb{Z}) \subseteq \mathbb{Z}$ and $p_i(0) = 0$. A variety of multidimensional versions of the theorem are also known. A significant drawback[3] of Furstenberg's approach is that it uses the axiom of choice, and so does not give *any* explicit function $\omega_k(N)$.

Rather recently, Gowers [**Gow98, Gow01**] made a major breakthrough in giving the first "sensible" bounds for $r_k(N)$.

THEOREM 2.5 (Gowers). *Let $k \geqslant 3$ be an integer. Then there is a constant $c_k > 0$ such that*

$$r_k(N) \ll N(\log\log N)^{-c_k}.$$

This is still a long way short of the conjecture that $r_k(N) < \pi(N)$ for N sufficiently large. However, in addition to coming much closer to this bound than any previous arguments, Gowers succeeded in introducing methods of harmonic analysis to the problem for the first time since Roth. Since harmonic analysis (in the form of the circle method of Hardy and Littlewood) has been the most effective tool in tackling additive problems involving the primes, it seems fair to say that it was the work of Gowers which first gave us hope of tackling long progressions of primes. The ideas of Gowers will feature fairly substantially in this exposition, but in our paper [**GTc**] much of what is done is more in the ergodic-theoretic spirit of Furstenberg and of more recent authors in that area such as Host–Kra [**HK05**] and Ziegler [**Zie**].

To conclude this discussion of Szemerédi's theorem we mention a variant of it which is far more useful in practice. This applies to *functions*[4] $f : \mathbb{Z}/N\mathbb{Z} \to [0, 1]$ rather than just to (characteristic functions of) sets. It also guarantees *many* arithmetic progressions of length k. This version does, however, follow from the earlier formulation by some fairly straightforward averaging arguments due to Varnavides [**Var59**].

PROPOSITION 2.6 (Szemerédi's theorem, II). *Let $k \geqslant 3$ be an integer, and let $\delta \in (0, 1]$ be a real number. Then there is a constant $c(k, \delta) > 0$ such that for any function $f : \mathbb{Z}/N\mathbb{Z} \to [0, 1]$ with $\mathbb{E}f = \delta$ we have the bound*[5]

$$\mathbb{E}_{x,d\in\mathbb{Z}/N\mathbb{Z}} f(x)f(x + d) \dots f(x + (k - 1)d) \geqslant c(k, \delta).$$

We do not, in [**GTc**], prove any new bounds for $r_k(N)$. Our strategy is to prove a *relative Szemerédi theorem*. To describe this we consider, for brevity of exposition, only the case $k = 4$. Consider the following table.

[3] A discrete analogue of Furstenberg's argument has now been found by Tao [**Taob**]. It does give an explicit function $\omega_k(N)$, but once again it tends to zero incredibly slowly.

[4] When discussing additive problems it is often convenient to work in the context of a finite abelian group G. For problems involving $\{1, \dots, N\}$ there are various technical tricks which allow one to work in $\mathbb{Z}/N'\mathbb{Z}$, for some $N' \approx N$. In this expository article we will not bother to distinguish between $\{1, \dots, N\}$ and $\mathbb{Z}/N\mathbb{Z}$. For examples of the technical trickery required here, see [**GTc**, Definition 9.3], or the proof of Theorem 2.6 in [**Gow01**].

[5] We use this very convenient conditional expectation notation repeatedly. $\mathbb{E}_{x\in A} f(x)$ is defined to equal $|A|^{-1} \sum_{x\in A} f(x)$.

Szemerédi	Relative Szemerédi
$\{1,\dots,N\}$	?
$A \subseteq \{1,\dots,N\}$ $\lvert A\rvert \geqslant 0.0001N$	$\mathcal{P}_N$ $=$ primes $\leqslant N$
Szemerédi's theorem: A contains many 4-term APs.	Green–Tao theorem: $\mathcal{P}_N$ contains many 4-term APs.

On the left-hand side of this table is Szemerédi's theorem for progressions of length 4, stated as the result that a set $A \subseteq \{1,\dots,N\}$ of density 0.0001 contains many 4-term APs if N is large enough. On the right is the result we wish to prove. Only one thing is missing: we must find an object to play the rôle of $\{1,\dots,N\}$. We might try to place the primes inside some larger set $\mathcal{P}'_N$ in such a way that $|\mathcal{P}_N| \geqslant 0.0001|\mathcal{P}'_N|$, and hope to prove an analogue of Szemerédi's theorem for $\mathcal{P}'_N$.

A natural candidate for $\mathcal{P}'_N$ might be the set of *almost primes*; perhaps, for example, we could take $\mathcal{P}'_N$ to be the set of integers in $\{1,\dots,N\}$ with at most 100 prime factors. This would be consistent with the intuition, coming from sieve theory, that almost primes are much easier to deal with than primes. It is relatively easy to show, for example, that there are long arithmetic progressions of almost primes [**Gro80**].

This idea does not quite work, but a variant of it does. Instead of a *set* $\mathcal{P}'_N$ we instead consider what we call a *measure*[6] $\nu : \{1,\dots,N\} \to [0,\infty)$. Define the *von Mangoldt function* Λ by

$$\Lambda(n) := \begin{cases} \log p & \text{if } n = p^k \text{ is prime} \\ 0 & \text{otherwise.} \end{cases}$$

The function Λ is a weighted version of the primes; note that the prime number theorem is equivalent to the fact that $\mathbb{E}_{1\leqslant n\leqslant N}\Lambda(n) = 1 + o(1)$. Our measure ν will satisfy the following two properties.

(i) (ν majorises the primes) We have $\Lambda(n) \leqslant 10000\nu(n)$ for all $1 \leqslant n \leqslant N$.
(ii) (primes sit inside ν with positive density) We have $\mathbb{E}_{1\leqslant n\leqslant N}\nu(n) = 1 + o(1)$.

These two properties are very easy to satisfy, for example by taking $\nu = \Lambda$, or by taking ν to be a suitably normalised version of the almost primes. Remember, however, that we intend to prove a Szemerédi theorem relative to ν. In order to do that it is reasonable to suppose that ν will need to meet more stringent conditions. The conditions we use in [**GTc**] are called the *linear forms condition* and the *correlation condition*. We will not state them here in full generality, referring the reader to [**GTc**, §3] for full details. We remark, however, that verifying these conditions is of the same order of difficulty as obtaining asymptotics for, say,

$$\sum_{n\leqslant N} \nu(n)\nu(n+2).$$

[6]Actually, ν is just a function but we use the term "measure" to distinguish it from other functions appearing in our work.

For this reason there is no chance that we could simply take $\nu = \Lambda$, since if we could do so we would have solved the twin prime conjecture.

We call a measure ν which satisfies the linear forms and correlation conditions *pseudorandom*.

To succeed with the relative Szemerédi strategy, then, our aim is to find a pseudorandom measure ν for which conditions (i) and (ii) and the are satisfied. Such a function[7] comes to us, like the almost primes, from the idea of using a sieve to bound the primes. The particular sieve we had recourse to was the Λ^2-sieve of Selberg. Selberg's great idea was as follows.

Fix a parameter R, and let $\lambda = (\lambda_d)_{d=1}^R$ be any sequence of real numbers with $\lambda_1 = 1$. Then the function

$$\sigma_\lambda(n) := (\sum_{\substack{d|n \\ d \leqslant R}} \lambda_d)^2$$

majorises the primes greater than R. Indeed if $n > R$ is prime then the truncated divisor sum over $d|n$, $d \leqslant R$ contains just one term corresponding to $d = 1$.

Although this works for any sequence λ, some choices are much better than others. If one wishes to minimise

$$\sum_{n \leqslant N} \sigma_\lambda(n)$$

then, provided that R is a bit smaller than $\sqrt{N}$, one is faced with a minimisation problem involving a certain quadratic form in the λ_ds. The optimal weights λ_d^{SEL}, Selberg's weights, have a slightly complicated form, but roughly we have

$$\lambda_d^{\mathrm{SEL}} \approx \lambda_d^{\mathrm{GY}} := \mu(d)\frac{\log(R/d)}{\log R},$$

where $\mu(d)$ is the Möbius function. These weights were considered by Goldston and Yıldırım [**GY**] in some of their work on small gaps between primes (and earlier, in other contexts, by others including Heath-Brown). It seems rather natural, then, to define a function ν by

$$\nu(n) := \begin{cases} \log N & n \leqslant R \\ \dfrac{1}{\log R}\Big(\sum\limits_{\substack{d|n \\ d \leqslant R}} \lambda_d^{\mathrm{GY}}\Big)^2 & n > R. \end{cases}$$

The weight $1/\log R$ is chosen for normalisation purposes; if $R < N^{1/2-\epsilon}$ for some $\epsilon > 0$ then we have $\mathbb{E}_{1 \leqslant n \leqslant N}\nu(n) = 1 + o(1)$.

One may more-or-less read out of the work of Goldston and Yıldırım a proof of properties (i) and (ii) above, as well as pseudorandomness, for this function ν.

[7]Actually, this is a lie. There is no pseudorandom measure which majorises the primes themselves. One must first use a device known as the W-trick to remove biases in the primes coming from their irregular distribution in residue classes to small moduli. This is discussed in §3.

One requires that $R < N^c$ where c is sufficiently small. These verifications use the classical zero-free region for the ζ-function and classical techniques of contour integration.

Goldston and Yıldırım's work was part of their long-term programme to prove that

$$\liminf_{n\to\infty} \frac{p_{n+1} - p_n}{\log n} = 0, \tag{1}$$

where p_n is the nth prime. We have recently learnt that this programme has been successful. Indeed together with J. Pintz they have used weights coming from a higher-dimensional sieve in order to establish (1). It is certain that without the earlier preprints of Goldston and Yıldırım our work would have developed much more slowly, at the very least.

Let us conclude this section by remarking that ν will not play a great rôle in the subsequent exposition. It plays a substantial rôle in [**GTc**], but in a relatively non-technical exposition like this it is often best to merely remark that the measure ν and the fact that it is pseudorandom is used all the time in proofs of the various statements that we will describe.

3. Progressions of length three and linear bias

Let G be a finite abelian group with cardinality N. If $f_1, \dots, f_k : G \to \mathbb{C}$ are any functions we write

$$T_k(f_1, \dots, f_k) := \mathbb{E}_{x,d\in G} f_1(x) f_2(x+d) \dots f_k(x+(k-1)d)$$

for the normalised count of k-term APs involving the f_i. When all the f_i are equal to some function f, we write

$$T_k(f) := T_k(f, \dots, f).$$

When f is equal to 1_A, the characteristic function of a set $A \subseteq G$, we write

$$T_k(A) := T_k(1_A) = T_k(1_A, \dots, 1_A).$$

This is simply the number of k-term arithmetic progressions in the set A, divided by N^2.

Let us begin with a discussion of 3-term arithmetic progressions and the trilinear form T_3. If $A \subseteq G$ is a set, then clearly $T_3(A)$ may vary between 0 (when $A = \emptyset$) and 1 (when $A = G$). If, however, one places some restriction on the cardinality of A then the following question seems natural:

Question 3.1. Let $\alpha \in (0,1)$, and suppose that $A \subseteq G$ is a set with cardinality αN. What is $T_3(A)$?

To think about this question, we consider some examples.

Example 1 (Random set). Select a set $A \subseteq G$ by picking each element $x \in G$ to lie in A independently at random with probability α. Then with high probability $|A| \approx \alpha N$. Also, if $d \neq 0$, the arithmetic progression $(x, x+d, x+2d)$ lies in G with probability α^3. Thus we expect that $T_3(A) \approx \alpha^3$, and indeed it can be shown using simple large deviation estimates that this is so with high probability.

Write $E_3(\alpha) := \alpha^3$ for the *expected* normalised count of three-term progressions in the random set of Example 1. One might refine Question 3.1 by asking:

QUESTION 3.2. Let $\alpha \in (0,1)$, and suppose that $A \subseteq G$ is a set with cardinality αN. Is $T_3(A) \approx E_3(\alpha)$?

It turns out that the answer to this question is "no", as the next example illustrates.

Example 2 (Highly structured set, I). Let $G = \mathbb{Z}/N\mathbb{Z}$, and consider the set $A = \{1, \dots, \lfloor \alpha N \rfloor\}$, an interval. It is not hard to check that if $\alpha < 1/2$ then $T_3(A) \approx \frac{1}{4}\alpha^2$, which is much bigger than $E_3(\alpha)$ for small α.

These first two examples do not rule out a positive answer to the following question.

QUESTION 3.3. Let $\alpha \in (0,1)$, and suppose that $A \subseteq G$ is a set with cardinality αN. Is $T_3(A) \geqslant E_3(\alpha)$?

If this question did have an affirmative answer, the quest for progressions of length three in sets would be a fairly simple one (the primes would trivially contain many three-term progressions on density grounds alone, for example). Unfortunately, there are counterexamples.

Example 3 (Highly structured set, II). Let $G = \mathbb{Z}/N\mathbb{Z}$. Then there are sets $A \subseteq G$ with $|A| = \lfloor \alpha N \rfloor$, yet with $T_3(A) \ll \alpha^{10000}$. We omit the details of the construction, remarking only that such sets can be constructed[8] as unions of intervals of length $\gg_\alpha N$ in $\mathbb{Z}/N\mathbb{Z}$.

Our discussion so far seems to be rather negative, in that our only conclusion is that none of Questions 3.1, 3.2 and 3.3 have particularly satisfactory answers. Note, however, that the three examples we have mentioned are all consistent with the following dichotomy.

DICHOTOMY 3.4 (Randomness vs Structure for 3-term APs). Suppose that $A \subseteq G$ has size αN. Then **either**

- $T_3(A) \approx E_3(\alpha)$ **or**
- A has *structure*.

It turns out that one may clarify, in quite a precise sense, what is meant by *structure* in this context. The following proposition may be proved by fairly straightforward harmonic analysis. We use the Fourier transform on G, which is defined as follows. If $f : G \to \mathbb{C}$ is a function and $\gamma \in \widehat{G}$ a character (i.e., a homomorphism from G to $\mathbb{C}^\times$), then

$$f^\wedge(\gamma) := \mathbb{E}_{x \in G} f(x)\gamma(x).$$

PROPOSITION 3.5 (Too many/few 3APs implies linear bias). *Let $\alpha, \eta \in (0,1)$. Then there is $c(\alpha, \eta) > 0$ with the following property. Suppose that $A \subseteq G$ is a set with $|A| = \alpha N$, and that*

$$|T_3(A) - E_3(\alpha)| \geqslant \eta.$$

[8]Basically one considers a set $S \subseteq \mathbb{Z}^2$ formed as the product of a Behrend set in $\{1, \dots, M\}$ and the interval $\{1, \dots, L\}$, for suitable M and L, and then one projects this set linearly to $\mathbb{Z}/N\mathbb{Z}$.

Then there is some character $\gamma \in \widehat{G}$ with the property that

$$|(1_A - \alpha)^\wedge(\gamma)| \geqslant c(\alpha, \eta).$$

Note that when $G = \mathbb{Z}/N\mathbb{Z}$ every character γ has the form $\gamma(x) = e(rx/N)$. It is the occurrence of the linear function $x \mapsto rx/N$ here which gives us the name *linear bias*.

It is an instructive exercise to compare this proposition with Examples 1 and 2 above. In Example 2, consider the character $\gamma(x) = e(x/N)$. If α is reasonably small then all the vectors $e(x/N)$, $x \in A$, have large positive real part and so when the sum

$$(1_A - \alpha)^\wedge(\gamma) = \mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} \widehat{1}_A(x) e(x/N)$$

is formed there is very little cancellation, with the result that the sum is large.

In Example 1, by contrast, there is (with high probability) considerable cancellation in the sum for $(1_A - \alpha)^\wedge(\gamma)$ for every character γ.

4. Linear bias and the primes

What use is Dichotomy 3.4 for thinking about the primes? One might hope to use Proposition 3.5 in order to count 3-term APs in some set $A \subseteq G$ by showing that A does not have linear bias. One would then know that $T_3(A) \approx E_3(\alpha)$, where $|A| = \alpha N$.

Let us imagine how this might work in the context of the primes. We have the following proposition[9], which is an analogue of Proposition 3.5. In this proposition[10], $\nu : \mathbb{Z}/N\mathbb{Z} \to [0, \infty)$ is the Goldston-Yıldırım measure constructed in §2.

Proposition 4.1. *Let $\alpha, \eta \in (0, 2]$. Then there is $c(\alpha, \eta) > 0$ with the following propety. Let $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ be a function with $\mathbb{E}f = \alpha$ and such that $0 \leqslant f(x) \leqslant 10000\nu(x)$ for all $x \in \mathbb{Z}/N\mathbb{Z}$, and suppose that*

$$|T_3(f) - E_3(\alpha)| \geqslant \eta.$$

Then

$$|\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(f(x) - \alpha)e(rx/N)| \geqslant c(\alpha, \eta) \tag{2}$$

for some $r \in \mathbb{Z}/N\mathbb{Z}$.

This proposition may be applied with $f = \Lambda$ and $\alpha = 1 + o(1)$. If we could rule out (2), then we would know that $T_3(\Lambda) \approx E_3(1) = 1$, and would thus have an asymptotic for 3-term progressions of primes.

[9]There are two ways of proving this proposition. One uses classical harmonic analysis. For pointers to such a proof, which would involve establishing an L^p-restriction theorem for ν for some $p \in (2, 3)$, we refer the reader to [**GT06**]. This proof uses more facts about ν than mere pseudorandomness. Alternatively, the result may be deduced from Proposition 3.5 by a transference principle using the machinery of [**GTc**, §6–8]. For details of this approach, which is far more amenable to generalisation, see [**GTb**]. Note that Proposition 4.1 does not feature in [**GTc**] and is stated here for pedagogical reasons only.

[10]Recall that we are being very hazy in distinguishing between $\{1, \dots, N\}$ and $\mathbb{Z}/N\mathbb{Z}$.

Sadly, (2) does hold. Indeed if N is even and $r = N/2$ then, observing that most primes are odd, it is easy to confirm that

$$\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}(\Lambda(x) - 1)e(rx/N) = -1 + o(1).$$

That is, the primes *do* have linear bias.

Fortunately, it is possible to modify the primes so that they have no linear bias using a device that we refer to as the W-trick. We have remarked that most primes are odd, and that as a result $\Lambda - 1$ has considerable linear bias. However, if one takes the odd primes

$$3, 5, 7, 11, 13, 17, 19, \dots$$

and then rescales by the map $x \mapsto (x - 1)/2$, one obtains the set

$$1, 2, 3, 5, 6, 8, 9, \dots$$

which does not have substantial $(\mathrm{mod}\, 2)$ bias (this is a consequence of the fact that there are roughly the same number of primes congruent to 1 and $3(\mathrm{mod}\, 4)$). Furthermore, if one can find an arithmetic progression of length k in this set of rescaled primes, one can certainly find such a progression in the primes themselves. Unfortunately this set of rescaled primes still has linear bias, because it contains only one element $\equiv 1(\mathrm{mod}\, 3)$. However, a similar rescaling trick may be applied to remove this bias too, and so on.

Here, then, is the W-trick. Take a slowly growing function $w(N) \to \infty$, and set $W := \prod_{p<w(N)} p$. Define the rescaled von Mangoldt function $\widetilde{\Lambda}$ by

$$\widetilde{\Lambda}(n) := \frac{\phi(W)}{W}\Lambda(Wn + 1).$$

The normalisation has been chosen so that $\mathbb{E}\widetilde{\Lambda} = 1 + o(1)$. $\widetilde{\Lambda}$ does not have substantial bias in any residue class to modulus $q < w(N)$, and so there is at least hope of applying a suitable analogue of Proposition 4.1 to it.

Now it is a straightforward matter to define a new pseudorandom measure $\widetilde{\nu}$ which majorises $\widetilde{\Lambda}$. Specifically, we have

(i) ($\widetilde{\nu}$ majorises the modified primes) We have $\widetilde{\lambda}(n) \leqslant 10000\widetilde{\nu}(n)$ for all $1 \leqslant n \leqslant N$.
(ii) (modified primes sit inside $\widetilde{\nu}$ with positive density) We have $\mathbb{E}_{1\leqslant n\leqslant N}\widetilde{\nu}(n) = 1 + o(1)$.

The following modified version of Proposition 4.1 may be proved:

PROPOSITION 4.2. *Let $\alpha, \eta \in (0, 2]$. Then there is $c(\alpha, \eta) > 0$ with the following property. Let $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ be a function with $\mathbb{E}f = \alpha$ and such that $0 \leqslant f(x) \leqslant 10000\widetilde{\nu}(x)$ for all $x \in \mathbb{Z}/N\mathbb{Z}$, and suppose that*

$$|T_3(f) - E_3(\alpha)| \geqslant \eta.$$

Then

$$|\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}(f(x) - \alpha)e(rx/N)| \geqslant c(\alpha, \eta) \tag{3}$$

for some $r \in \mathbb{Z}/N\mathbb{Z}$. □

This may be applied with $f = \widetilde{\Lambda}$ and $\alpha = 1 + o(1)$. Now, however, condition (3) does not so obviously hold. In fact, one has the estimate

$$\sup_{r \in \mathbb{Z}/N\mathbb{Z}} |\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(\widetilde{\Lambda}(x) - 1)e(rx/N)| = o(1). \tag{4}$$

To prove this requires more than simply the good distribution of $\widetilde{\Lambda}$ in residue classes to small moduli. It is, however, a fairly standard consequence of the Hardy-Littlewood circle method as applied to primes by Vinogradov. In fact, the whole theme of linear bias in the context of additive questions involving primes may be traced back to Hardy and Littlewood.

Proposition 4.2 and (4) imply that $T_3(\widetilde{\Lambda}) \approx E_3(1) = 1$. Thus there are infinitely many three-term progressions in the modified (W-tricked) primes, and hence also in the primes themselves[11].

5. Progressions of length four and quadratic bias

We return now to the discussion of §3. There we were interested in counting 3-term arithmetic progressions in a set $A \subseteq G$ with cardinality αN. In this section our interest will be in 4-term progressions.

Suppose then that $A \subseteq G$ is a set, and recall that

$$T_4(A) := \mathbb{E}_{x,d \in G} 1_A(x)1_A(x+d)1_A(x+2d)1_A(x+3d)$$

is the normalised count of four-term arithmetic progressions in A. One may, of course, ask the analogue of Question 3.1:

Question 5.1. Let $\alpha \in (0,1)$, and suppose that $A \subseteq G$ is a set with cardinality αN. What is $T_4(A)$?

Examples 1,2 and 3 make perfect sense here, and we see once again that there is no immediately satisfactory answer to Question 5.1. With high probability the random set of Example 1 has about $E_4(\alpha) := \alpha^4$ four-term APs, but there are structured sets with substantially more or less than this number of APs. As in §3, these examples are consistent with a dichotomy of the following type:

Dichotomy 5.2 (Randomness vs Structure for 4-term APs). Suppose that $A \subseteq G$ has size αN. Then **either**

- $T_4(A) \approx E_4(\alpha)$ **or**
- A has *structure*.

Taking into account the three examples we have so far, it is quite possible that this dichotomy takes *exactly* the form of that for 3-term APs. That is to say "A has structure" could just mean that A has linear bias:

Question 5.3. Let $\alpha, \eta \in (0,1)$. Suppose that $A \subseteq G$ is a set with $|A| = \alpha N$, and that

$$|T_4(A) - E_4(\alpha)| \geqslant \eta.$$

[11] In fact, this analysis does not have to be pushed much further to get a proof of Conjecture 1.2 for $k = 3$, that is to say an asymptotic for 3-term progressions of primes. One simply counts progressions $x, x+d, x+2d$ by splitting into residue classes $x \equiv b (\mathrm{mod}\, W)$, $d \equiv b' (\mathrm{mod}\, W)$ and using a simple variant of Proposition 4.2.

Must there exist some $c = c(\alpha, \eta) > 0$ and some character $\gamma \in \widehat{G}$ with the property that

$$|(1_A - \alpha)^\wedge(\gamma)| \geqslant c(\alpha, \eta)?$$

That the answer to this question is no, together with the nature of the counterexample, is one of the key themes of our whole work. This phenomenon was discovered, in the context of ergodic theory, by Furstenberg and Weiss [**FW96**] and then again, in the discrete setting, by Gowers [**Gow01**].

Example 4 (Quadratically structured set). Define $A \subseteq \mathbb{Z}/N\mathbb{Z}$ to be the set of all x such that $x^2 \in [-\alpha N/2, \alpha N/2]$. It is not hard to check using estimates for Gauss sums that $|A| \approx \alpha N$, and also that

$$\sup_{r \in \mathbb{Z}/N\mathbb{Z}} |\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}}(1_A(x) - \alpha)e(rx/N)| = o(1),$$

that is to say A does not have linear bias. (In fact, the largest Fourier coefficient of $1_A - \alpha$ is just $N^{-1/2+\epsilon}$.) Note, however, the relation

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

valid for arbitrary $x, d \in \mathbb{Z}/N\mathbb{Z}$. This means that if $x, x+d, x+2d \in A$ then automatically we have

$$(x+3d)^2 \in [-7\alpha N/2, 7\alpha N/2].$$

It seems, then, that if we know that $x, x+d$ and $x+2d$ lie in A there is a very high chance that $x+3d$ also lies in A. This observation may be made rigorous, and it does indeed transpire that $T_4(A) \geqslant c\alpha^3$.

How can one rescue the randomness-structure dichotomy in the light of this example? Rather remarkably, "quadratic" examples like Example 4 are the *only* obstructions to having $T_4(A) \approx E_4(\alpha)$. There is an analogue of Proposition 3.5 in which characters γ are replaced by "quadratic" objects[12].

Proposition 5.4 (Too many/few 4APs implies quadratic bias). *Let $\alpha, \eta \in (0,1)$. Then there is $c(\alpha, \eta) > 0$ with the following property. Suppose that $A \subseteq G$ is a set with $|A| = \alpha N$, and that*

$$|T_4(A) - E_4(\alpha)| \geqslant \eta.$$

Then there is some quadratic object $q \in \mathcal{Q}(\kappa)$, where $\kappa \geqslant \kappa_0(\alpha, \eta)$, with the property that

$$|\mathbb{E}_{x \in G}(1_A(x) - \alpha)q(x)| \geqslant c(\alpha, \eta).$$

We have not, of course, said what we mean by the set of *quadratic objects* $\mathcal{Q}(\kappa)$. To give the exact definition, even for $G = \mathbb{Z}/N\mathbb{Z}$, would take us some time, and we refer to [**GTa**] for a full discussion. In the light of Example 4, the reader will not be surprised to hear that quadratic exponentials such as $q(x) = e(x^2/N)$ are members of $\mathcal{Q}$. However, $\mathcal{Q}(\kappa)$ also contains rather more obscure objects[13] such as

$$q(x) = e(x\sqrt{2}\{x\sqrt{3}\})$$

[12]The proof of this proposition is long and difficult and may be found in [**GTa**]. It is heavily based on the arguments of Gowers [**Gow98, Gow01**]. This proposition has no place in [**GTc**], and it is once again included for pedagogical reasons only. It played an important rôle in the development of our ideas.

[13]We are thinking of these as defined on $\{1, \ldots, N\}$ rather than $\mathbb{Z}/N\mathbb{Z}$.

and

$$q(x) = e(x\sqrt{2}\{x\sqrt{3}\} + x\sqrt{5}\{x\sqrt{7}\} + x\sqrt{11}),$$

where $\{x\}$ denotes fractional part. The parameter κ governs the complexity of the expressions which are allowed: smaller values of κ correspond to more complicated expressions. The need to involve these "generalised" quadratics in addition to "genuine" quadratics such as $e(x^2/N)$ was first appreciated by Furstenberg and Weiss in the ergodic theory context, and the matter also arose in the work of Gowers.

6. Quadratic bias and the primes

It is possible to prove[14] a version of Proposition 5.4 which might be applied to primes. The analogue of Proposition 4.1 is true but not useful, for the same reason as before: the primes exhibit significant bias in residue classes to small moduli. As before, this bias may be removed using the W-trick.

PROPOSITION 6.1. *Let $\alpha, \eta \in (0, 2]$. Then there are $c(\alpha, \eta)$ and $\kappa_0(\alpha, \eta) > 0$ with the following propety. Let $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ be a function with $\mathbb{E}f = \alpha$ and such that $0 \leqslant f(x) \leqslant 10000\widetilde{\nu}(x)$ for all $x \in \mathbb{Z}/N\mathbb{Z}$, and suppose that*

$$|T_4(f) - E_4(\alpha)| \geqslant \eta.$$

Then we have

$$|\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}(f(x) - \alpha)q(x)| \geqslant c(\alpha, \eta) \tag{5}$$

for some quadratic object $q \in \mathcal{Q}(\kappa)$ with $\kappa \geqslant \kappa_0(\alpha, \eta)$.

One is interested, of course, in applying this with $f = \widetilde{\Lambda}$. If we could verify that (5) does not hold, that is to say the primes do not have quadratic bias, then it would follow that $T_4(\widetilde{\Lambda}) \approx E_4(1) = 1$. This means that the modified (W-tricked) primes have many 4-term progressions, and hence so do the primes themselves[15].

One wishes to show, then, that for fixed κ one has

$$\sup_{q\in\mathcal{Q}(\kappa)} |\mathbb{E}_{x\in\mathbb{Z}/N\mathbb{Z}}(\widetilde{\Lambda}(x) - 1)q(x)| = o(1). \tag{6}$$

Such a result is certainly not a consequence of the classical Hardy-Littlewood circle method[16]. Generalised quadratic phases such as $q(x) = e(x\sqrt{2}\{x\sqrt{3}\})$ are particularly troublesome. Although we do now have a proof of (6), it is very long and complicated. See [**GTd**] for details.

In the next section we explain how our original paper [**GTc**] managed to avoid the need to prove (6).

---

[14] As with Proposition 4.1, this proposition does not appear in [**GTc**], though it motivated our work and a variant of it is used in our later work [**GTb**]. Once again there are two proofs. One is based on a combination of harmonic analysis and the work of Gowers, is difficult, and requires more facts about ν than mere pseudorandomness. This was our original argument. It is also possible to proceed by a transference principle, deducing the result from Proposition 5.4 using the machinery of [**GTc**, §6–8]. See [**GTb**] for more details.

[15] in fact, just as for progressions of length 3, this allows one to obtain a proof of Conjecture 1.2 for $k = 4$, that is to say an asymptotic for prime progressions of length 4. See [**GTb**].

[16] Though reasonably straightforward extensions of the circle method do permit one to handle genuine quadratic phases such as $q(x) = e(x^2\sqrt{2})$.

7. Quotienting out the bias - the energy increment argument

Our paper [**GTc**] failed to rule out the possibility that $\widetilde{\Lambda} - 1$ correlates with some quadratic function $q \in \mathcal{Q}(\kappa)$. For that reason we did not obtain a proof of Conjecture 1.2, getting instead the weaker statement of Theorem 1.3. In this section[17] we outline the *energy increment* argument of [**GTc**], which allowed us to deal with the possibility that $\widetilde{\Lambda} - 1$ does correlate with a quadratic.

We begin by writing

$$\widetilde{\Lambda} := 1 + f_0. \tag{7}$$

Proposition 6.1 tells us that $T_4(\widetilde{\Lambda}) \approx 1$, unless f_0 correlates with some quadratic $q_0 \in \mathcal{Q}$. Suppose, then, that

$$|\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f_0(x) q_0(x)| \geqslant \eta.$$

Then we revise the decomposition (7) to

$$\widetilde{\Lambda} := F_1 + f_1, \tag{8}$$

where F_1 is a function defined using q_0. In fact, F_1 is basically the average of $\widetilde{\Lambda}$ over approximate level sets of q_0. That is, one picks an appropriate scale[18] $\epsilon = 1/J$, and then defines

$$F_1 := \mathbb{E}(\widetilde{\Lambda}|\mathcal{B}_0),$$

where $\mathcal{B}_0$ is the σ-algebra generated by the sets $x : q_0(x) \in [j/J, (j+1)/J)$.

A variant of Proposition 6.1 implies a new dichotomy: either $T_4(\widetilde{\Lambda}) \approx T_4(F_1)$, or else f_1 correlates with some quadratic $q_1 \in \mathcal{Q}$. Suppose then that

$$|\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f_1(x) q_1(x)| \geqslant \eta.$$

We then further revise the decomposition (8) to

$$\widetilde{\Lambda} := F_2 + f_2,$$

where now

$$F_2 := \mathbb{E}(\widetilde{\Lambda}|\mathcal{B}_0 \wedge \mathcal{B}_1),$$

the σ-algebra being defined by the joint level sets of q_0 and q_1.

We repeat this process. It turns out that the algorithm stops in a finite number s of steps, bounded in terms of η. The reason for this is that each new assumption

$$|\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f_j(x) q_j(x)| \geqslant \eta$$

---

[17] The exposition in this section is rather looser than in other sections. To make the argument rigorous, one must introduce various technical devices, such as the exceptional sets which feature in [**GTc**, §7,8]. We are also being rather vague about the meaning of terms such as "correlate", and the parameter κ involved in the definition of quadratic object. Note also that the argument of [**GTc**] uses *soft* quadratic objects rather than the genuine ones which we are discussing here for expositional purposes. See §8 for a brief discussion of these.

[18] As we remarked, the actual situation is more complicated. There is an averaging over possible decompositions of $[0, 1]$ into intervals of length ϵ, to ensure that the level sets look pleasant. There is also a need to consider exceptional sets, which unfortunately makes the argument look rather messy.

implies an increased lower bound for the *energy* of $\widetilde{\Lambda}$ relative to the σ-algebra $\mathcal{B}_0 \wedge \cdots \wedge \mathcal{B}_{j-1}$, that is to say the quantity

$$E_j := \|\mathbb{E}(\widetilde{\Lambda}|\mathcal{B}_0 \wedge \cdots \wedge \mathcal{B}_{j-1})\|_2.$$

The fact that $\widetilde{\Lambda}$ is dominated by $\widetilde{\nu}$ does, however, provide a universal bound for the energy, by dint of the evident inequality

$$E_j \leqslant 10000\|\mathbb{E}(\widetilde{\nu}|\mathcal{B}_0 \wedge \cdots \wedge \mathcal{B}_{j-1})\|_2.$$

The pseudorandomness of $\widetilde{\nu}$ allows one[19] to bound the right-hand side here by $O(1)$.

At termination, then, we have a decomposition

$$\widetilde{\Lambda} = F_s + f_s,$$

where

$$\sup_{q \in \mathcal{Q}} |\mathbb{E}_{x \in \mathbb{Z}/N\mathbb{Z}} f_s(x) q(x)| < \eta, \tag{9}$$

and F_s is defined by

$$F_s := \mathbb{E}(\widetilde{\Lambda}|\mathcal{B}_0 \wedge \mathcal{B}_1 \wedge \cdots \wedge \mathcal{B}_{s-1}). \tag{10}$$

A variant of Proposition 6.1 implies, together with (9), that

$$T_4(\widetilde{\Lambda}) \approx T_4(F_s). \tag{11}$$

What can be said about $T_4(F_s)$? Let us note two things about the function F_s. First of all the definition (10) implies that

$$\mathbb{E}F_s = \mathbb{E}\widetilde{\Lambda} = 1 + o(1). \tag{12}$$

Secondly, F_s is not too large pointwise; this is again an artifact of $\widetilde{\Lambda}$ being dominated by $\widetilde{\nu}$. We have, of course,

$$\|F_s\|_\infty = \|\mathbb{E}(\widetilde{\Lambda}|\mathcal{B}_0 \wedge \mathcal{B}_1 \wedge \cdots \wedge \mathcal{B}_{s-1})\|_\infty \leqslant 10000\|\mathbb{E}(\widetilde{\nu}|\mathcal{B}_0 \wedge \mathcal{B}_1 \wedge \cdots \wedge \mathcal{B}_{s-1})\|_\infty.$$

The pseudorandomness of $\widetilde{\nu}$ can again be used[20] to show that the right-hand side here is $10000 + o(1)$; that is,

$$\|F_s\|_\infty \leqslant 10000 + o(1). \tag{13}$$

The two properties (12) and (13) together mean that F_s behaves rather like the characteristic function of a subset of $\mathbb{Z}/N\mathbb{Z}$ with density at least $1/10000$. This suggests the use of Szemerédi's theorem to bound $T_4(F_s)$ below. The formulation of that theorem given in Proposition 2.6 applies to exactly this situation, and it tells us that

$$T_4(F_s) > c$$

for some absolute constant $c > 0$. Together with (11) this implies a similar lower bound for $T_4(\widetilde{\Lambda})$, which means that there are infinitely many 4-term arithmetic progressions of primes.

---

[19]This deduction uses the machinery of the Gowers U^3-norm, which we do not discuss in this survey. See [**GTc**, §6] for a full discussion. Of specific relevance is the fact that $\|\widetilde{\nu}\|_{U^3} = o(1)$, which is a consequence of the pseudorandomness of $\widetilde{\nu}$.

[20]Again, the machinery of the Gowers U^3-norm is used.

Let us conclude this section with an overview of what it is we have proved. The only facts about $\widetilde{\Lambda}$ that we used were that it is dominated pointwise by $10000\widetilde{\nu}$, and that $\mathbb{E}\widetilde{\Lambda}$ is not too small. The argument sketched above applies equally well in the general context of functions with these properties, and in the context of an arbitrary pseudorandom measure (not just the Goldston-Yıldırım measure).

PROPOSITION 7.1 (Relative Szemerédi Theorem). *Let $\delta \in (0,1]$ be a real number and let ν be a psuedorandom measure. Then there is a constant $c'(4,\delta) > 0$ with the following property. Suppose that $f : \mathbb{Z}/N\mathbb{Z} \to \mathbb{R}$ is a function such that $0 \leqslant f(x) \leqslant \nu(x)$ pointwise, and for which $\mathbb{E}f \geqslant \delta$. Then we have the estimate*

$$T_4(f) \geqslant c'(4,\delta).$$

In [**GTc**] we prove the same[21] theorem for progressions of any length $k \geqslant 3$.

Proposition 7.1 captures the spirit of our argument quite well. We first deal with arithmetic progressions in a rather general context. Only upon completion of that study do we concern ourselves with the primes, and this is simply a matter of constructing an appropriate pseudorandom measure. Note also that Szemerédi's theorem is used as a "black box". We do not need to understand the proof of it, or to have good bounds for it.

Observe that one consequence of Proposition 7.1 is a *Szemerédi theorem relative to the primes*: any subset of the primes with positive relative density contains progressions of arbitrary length. Applying this to the set of primes congruent to $1 (\mathrm{mod}\, 4)$, we see that there are arbitrarily long progressions of numbers which are sums of two squares.

8. Soft obstructions

Readers familiar with [**GTc**] may have been confused by our exposition thus far, since "quadratic objects" play essentially no rôle in that paper. The purpose of this brief section is to explain why this is so, and to provide a bridge between this survey and our paper. Further details and discussion may be found in [**GTc**, §6].

Let us start by recalling §3, where a set of "obstructions" to a set $A \subseteq G$ having roughly $E_3(\alpha)$ three-term APs was obtained. This was just the collection of characters $\gamma \in \widehat{G}$, and we used the term *linear bias* to describe correlation with one of these characters.

Let $f : G \to \mathbb{C}$ be a function with $\|f\|_\infty \leqslant 1$. Now we observe the formula

$$\mathbb{E}_{a,b\in G}\overline{f(x+a)f(x+b)}f(x+a+b) = \sum_{\gamma\in\widehat{G}} |\widehat{f}(\gamma)|^2\overline{\widehat{f}(\gamma)}\overline{\gamma(x)},$$

which may be verified by straightforward harmonic analysis on G. Coupled with the fact that

$$\sum_{\gamma\in\widehat{G}} |\widehat{f}(\gamma)|^2 \leqslant 1,$$

[21]Note, however, that the definition of pseudorandom measure is strongly dependent on k.

a consequence of Parseval's identity, this means that the[22] "dual function"

$$\mathcal{D}_2 f := \mathbb{E}_{a,b\in G}\overline{f(x+a)f(x+b)}f(x+a+b)$$

can be approximated by the weighted sum of a few characters. Every character is actually equal to a dual function; indeed we clearly have $\mathcal{D}_2(\gamma) = \overline{\gamma}$.

We think of the dual functions $\mathcal{D}_2(f)$ as *soft linear obstructions*. They may be used in the iterative argument of §7 in place of the genuinely linear functions, after one has established certain algebraic closure properties of these functions (see [**GTc**, Proposition 6.2])

The great advantage of these soft obstructions is that it is reasonably obvious how they should be generalised to give objects appropriate for the study of longer arithmetic progressions. We write $\mathcal{D}_3(f)$ for

$$\mathbb{E}_{a,b,c}\overline{f(x+a)f(x+b)f(x+c)}f(x+a+b)f(x+a+c)f(x+b+c)\overline{f(x+a+b+c)}.$$

This is a kind of sum of f over parallelepipeds (minus one vertex), whereas $\mathcal{D}_2(f)$ was a sum over parallelograms (minus one vertex). This we think of as a *soft quadratic obstruction*. Gone are the complications of having to deal with explicit generalised quadratic functions which, rest assured, only become worse when one deals with progressions of length 5 and longer.

The idea of using these soft obstructions came from the ergodic-theory work of Host and Kra [**HK05**], where very similar objects are involved.

We conclude by emphasising that soft obstructions lead to relatively soft results, such as Theorem 1.3. To get a proof of Conjecture 1.2 it will be necessary to return to generalised quadratic functions and their higher-order analogues.

9. Acknowledgements

I would like to thank James Cranch for reading the manuscript and advice on using Mathematica, and Terry Tao for several helpful comments.

References

[Beh46] F. A. Behrend – "On sets of integers which contain no three terms in arithmetical progression", *Proc. Nat. Acad. Sci. U. S. A.* **32** (1946), p. 331–332.

[BL96] V. Bergelson & A. Leibman – "Polynomial extensions of van der Waerden's and Szemerédi's theorems", *J. Amer. Math. Soc.* **9** (1996), no. 3, p. 725–753.

[Cho44] S. Chowla – "There exists an infinity of 3—combinations of primes in A. P", *Proc. Lahore Philos. Soc.* **6** (1944), no. 2, p. 15–16.

[ET36] P. Erdős, & P. Turán – "On some sequences of integers", *J. London Math. Soc.* **11** (1936), p. 261–264.

[Fur77] H. Furstenberg – "Ergodic behavior of diagonal measures and a theorem of Szemerédi on arithmetic progressions", *J. Analyse Math.* **31** (1977), p. 204–256.

[FW96] H. Furstenberg & B. Weiss – "A mean ergodic theorem for $(1/N)\sum_{n=1}^{N} f(T^n x)g(T^{n^2}x)$", in *Convergence in ergodic theory and probability (Columbus, OH, 1993)*, Ohio State Univ. Math. Res. Inst. Publ., vol. 5, de Gruyter, Berlin, 1996, p. 193–227.

[22]The subscript 2 refers to the Gowers U^2-norm, which is relevant to the study of progressions of length 3.

[GH79] E. Grosswald & P. Hagis, Jr. – "Arithmetic progressions consisting only of primes", *Math. Comp.* **33** (1979), no. 148, p. 1343–1352.

[Gow98] W. T. Gowers – "A new proof of Szemerédi's theorem for arithmetic progressions of length four", *Geom. Funct. Anal.* **8** (1998), no. 3, p. 529–551.

[Gow01] ———, "A new proof of Szemerédi's theorem", *GAFA* **11** (2001), p. 465–588.

[Gro80] E. Grosswald – "Arithmetic progressions of arbitrary length and consisting only of primes and almost primes", *J. Reine Angew. Math.* **317** (1980), p. 200–208.

[GTa] B. J. Green & T. C. Tao – "An inverse theorem for the Gowers U^3-norm, with applications", to appear in Proc. Edinbrugh Math. Soc.

[GTb] ———, "Linear equations in primes", `math.NT/0606088`.

[GTc] ———, "The primes contain arbitrarily long arithmetic progressions", to appear in Ann. of Math.

[GTd] ———, "Quadratic uniformity of the Möbius function", `math.NT/0606087`.

[GT06] ———, "Restriction theory of Selberg's sieve, with applications", *J. Théorie des Nombres de Bordeaux* **18** (2006), p. 147–182.

[GY] D. Goldston & C. Y. Yildirim – "Small gaps between primes, I", `math.NT/0504336`.

[HB81] D. R. Heath-Brown – "Three primes and an almost-prime in arithmetic progression", *J. London Math. Soc. (2)* **23** (1981), no. 3, p. 396–414.

[HK05] B. Host & B. Kra – "Nonconventional ergodic averages and nilmanifolds", *Ann. of Math. (2)* **161** (2005), no. 1, p. 397–488.

[Kra06] B. Kra – "The Green-Tao theorem on arithmetic progressions in the primes: an ergodic point of view", *Bull. Amer. Math. Soc. (N.S.)* **43** (2006), no. 1, p. 3–23 (electronic).

[ŁL] I. Łaba & M. Lacey – "On sets of integers not containing long arithmetic progressions", `math.CO/0108155`.

[Ran61] R. A. Rankin – "Sets of integers containing not more than a given number of terms in arithmetical progression", *Proc. Roy. Soc. Edinburgh Sect. A* **65** (1960/1961), p. 332–344 (1960/61).

[Rot53] K. F. Roth – "On certain sets of integers", *J. London Math. Soc.* **28** (1953), p. 104–109.

[Sze69] E. Szemerédi – "On sets of integers containing no four elements in arithmetic progression", *Acta Math. Acad. Sci. Hungar.* **20** (1969), p. 89–104.

[Sze75] ———, "On sets of integers containing no k elements in arithmetic progression", *Acta Arith.* **27** (1975), p. 199–245, Collection of articles in memory of Juriĭ Vladimirovič Linnik.

[Taoa] T. C. Tao – "A note on Goldston-Yıldırım correlation estimates", unpublished.

[Taob] ———, "A quantitative ergodic theory proof of Szemerédi's theorem", preprint.

[Tao06a] ———, "Arithmetic progressions and the primes", *Collectanea Mathematica* **extra volume** (2006), p. 37–88, Proceedings, 7th International Conference on Harmonic Analysis and Partial Differential Equations.

[Tao06b] ———, "Obstructions to uniformity, and arithmetic patterns in the primes", *Quart. J. Pure App. Math.* **2** (2006), p. 199–217, special issue in honour of John H. Coates.

[Var59] P. Varnavides – "On certain sets of positive density", *J. London Math. Soc.* **34** (1959), p. 358–360.

[vdC39] J. G. van der Corput – "Über Summen von Primzahlen und Primzahlquadraten", *Math. Ann.* **116** (1939), p. 1–50.

[Zie] T. Ziegler – "Universal characteristic factors and Furstenberg averages", to appear in J. Amer. Math. Soc.

School of Mathematics, University Walk, Bristol BS8 1TW, England
E-mail address: `b.j.green@bristol.ac.uk`