

Generalising the Hardy–Littlewood method for primes

Ben Green*

Abstract. The Hardy–Littlewood method is a well-known technique in analytic number theory. Among its spectacular applications are Vinogradov’s 1937 result that every sufficiently large odd number is a sum of three primes, and a related result of Chowla and Van der Corput giving an asymptotic for the number of 3-term progressions of primes, all less than N . This article surveys recent developments of the author and T. Tao, in which the Hardy–Littlewood method has been generalised to obtain, for example, an asymptotic for the number of 4-term arithmetic progressions of primes less than N .

Mathematics Subject Classification (2000). 11B25.

Keywords. Hardy–Littlewood method, prime numbers, arithmetic progressions, nilsequences.

1. Introduction

Godfrey Harold Hardy and John Edensor Littlewood wrote, in the 1920s, a famous series of papers *Some problems of “partitio numerorum”*. In these papers, whose content is elegantly surveyed by Vaughan [31], they developed techniques having their genesis in work of Hardy and Ramanujan on the partition function [18] to well-known questions in additive number theory such as Waring’s problem and the Goldbach problem.

Papers III and V in the series, [16], [17], were devoted to the sequence of primes. In particular it was established on the assumption of the Generalised Riemann Hypothesis that every sufficiently large odd number is the sum of three primes. In 1937 Vinogradov [33] made a further substantial advance by removing the need for any unproved hypothesis.

The Hardy–Littlewood–Vinogradov method may be applied to give an asymptotic count for the number of solutions in primes p_i to any fixed linear equation

$$a_1 p_1 + \cdots + a_t p_t = b$$

in, say, the box $p_1, \dots, p_t \leq N$, provided that at least 3 of the a_i are non-zero. This includes the three-primes result, and also the result that there are infinitely many

*This research was partially conducted during the period the author served as a Clay Research Fellow. He would like to express his sincere gratitude to the Clay Institute, and also to the Massachusetts Institute of Technology, where he was a Visiting Professor for the academic year 2005-06.

triples of primes $p_1 < p_2 < p_3$ in arithmetic progression, due to Chowla [4] and van der Corput [29].

More generally the Hardy–Littlewood method may also be used to investigate systems such as $\mathbf{A}\mathbf{p} = \mathbf{b}$, where \mathbf{A} is an $s \times t$ matrix with integer entries and, potentially, $s > 1$. A natural example of such a system is given by the $(k-2) \times k$ matrix

$$\mathbf{A} := \begin{pmatrix} 1 & -2 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 & 0 \\ & & & & \dots & & & \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 & 1 \end{pmatrix}, \quad (1)$$

in which case a solution to $\mathbf{A}\mathbf{p} = 0$ is just a k -term arithmetic progression of primes.

Here, unfortunately, the Hardy–Littlewood method falters in that it generally requires $t \geq 2s + 1$. In particular it cannot be used to handle progressions of length four or longer. There are certain special systems with fewer variables which *can* be handled. In this context we take the opportunity to mention a beautiful result of Balog [2], where it is shown that for any m there are distinct primes $p_1 < \dots < p_m$ such that each number $\frac{1}{2}(p_i + p_j)$ is also prime, or in other words that the system

$$\begin{aligned} p_1 + p_2 &= 2p_{12} \\ &\vdots \\ p_{m-1} + p_m &= 2p_{m-1,m} \end{aligned} \quad (2)$$

has a solution in primes $p_1, \dots, p_m, p_{12}, \dots, p_{m-1,m}$. There is also a result of Heath-Brown [19], in which it is established that there are infinitely many four-term progressions in which three members are prime and the fourth is either a prime or a product of two primes.

The survey of Kumchev and Tolev [25] gives a detailed account of applications of the Hardy–Littlewood method to additive prime number theory.

The aim of this survey is to give an overview of recent joint work with Terence Tao [13], [14], [15]. Our aim, which has been partially successful, is to extend the Hardy–Littlewood method so that it is capable of handling a more-or-less arbitrary system $\mathbf{A}\mathbf{p} = \mathbf{b}$, subject to the proviso that we do not expect to be able to handle any system which secretly encodes a “binary” problem such as Goldbach or Twin Primes.

This is a large and somewhat technical body of work. Perhaps my main aim here is to give a guide to our work so far, pointing out ways in which the various papers fit together, and future directions we plan to take. A subsidiary aim is to focus as far as possible on key concepts, rather than on details. Of course, one would normally aim to do this in a survey article. However in our case we expect that many of these details will be substantially cleaned up in future incarnations of the theory, whilst the key concepts ought to remain more-or-less as they are.

I will say rather little about our paper [12] establishing that there are arbitrarily long arithmetic progressions of primes. Whilst there is considerable overlap between

that paper and the ideas we discuss here, those methods were somewhat “soft” whereas the flavour of our more recent work is distinctly “hard”. We refer the reader to the survey of Tao in Volume II of these Proceedings, and also to the surveys [11], [23], [27], [28].

To conclude this introduction let me remark that the reader should not be under the impression that the Hardy–Littlewood method only applies to linear equations in primes, or even that this is the most popular application of the method. There has, for example, been a huge amount done on the circle of questions surrounding Waring’s problem. For a survey see [32]. More generally there are many spectacular results where variants of the method are used to locate integer points on quite general varieties, provided of course that there are sufficiently many variables. The reader may consult Wooley’s survey [34] for more information on this.

2. The Hardy–Littlewood heuristic

We have stated our interest in systems of linear equations in primes. While we are still somewhat lacking in theoretical results, there are heuristics which predict what answers we should expect in more-or-less any situation.

It is natural, when working with primes, to introduce the von Mangoldt function $\Lambda : \mathbb{N} \rightarrow \mathbb{R}_{\geq 0}$, defined by

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

The prime powers with $k \geq 2$ make a negligible contribution to any additive expression involving Λ . Thus, for example, the prime number theorem is equivalent to the statement that

$$\mathbb{E}_{n \leq N} \Lambda(n) = 1 + o(1).$$

Here we have used the very convenient notation of expectation from probability theory, setting $\mathbb{E}_{x \in X} := |X|^{-1} \sum_{x \in X}$ for any set X . The notation $o(1)$ refers to a quantity which tends to zero as $N \rightarrow \infty$.

We now discuss a version of the Hardy–Littlewood heuristic for systems of linear equations in primes. Here, and for the rest of the article, we restrict attention to homogeneous systems for simplicity of exposition.

Conjecture 2.1 (Hardy–Littlewood). Let A be a fixed $s \times t$ matrix with integer entries and such that there is at least one non-zero solution to $A\mathbf{x} = 0$ with $x_1, \dots, x_t \geq 0$. Then

$$\mathbb{E}_{\substack{x_1, \dots, x_t \leq N \\ A\mathbf{x} = 0}} \Lambda(x_1) \dots \Lambda(x_t) = \mathfrak{S}(A)(1 + o(1))$$

as $N \rightarrow \infty$, where the *singular series* $\mathfrak{S}(A)$ is equal to a product of local factors $\prod_p \alpha_p$, where

$$\alpha_p := \left(\frac{p}{p-1} \right)^t \lim_{M \rightarrow \infty} \frac{\mathbb{P}(\mathbf{Ax} = 0, (x_1, p) = \cdots = (x_t, p) = 1 | \mathbf{x} \in [-M, M]^t)}{\mathbb{P}(\mathbf{Ax} = 0 | \mathbf{x} \in [-M, M]^t)}.$$

The singular series reflects “local obstructions” to having solutions to $\mathbf{Ax} = 0$ in primes; in the simple example $A = \begin{pmatrix} 1 & 9 & -27 \end{pmatrix}$, where the associated equation $p_1 + 9p_2 - 27p_3 = 0$ has no solutions, one has $\alpha_3 = 0$. A more elegant formulation of the conjecture would include a “local obstruction at ∞ ” α_∞ , in exchange for removing the hypothesis on A .

Chowla and van der Corput’s results concerning three-term progressions of primes confirm the prediction Conjecture 2.1 for the matrix $A = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$. From this it is easy to derive an *asymptotic* for the number of triples (p_1, p_2, p_3) , $p_1 < p_2 < p_3 \leq N$, of primes in arithmetic progression.

Theorem 2.2 (Chowla, van der Corput, [4], [29]). *The number of triples of primes (p_1, p_2, p_3) , $p_1 < p_2 < p_3 \leq N$, in arithmetic progression is*

$$\mathfrak{S}_3 N^2 \log^{-3} N (1 + o(1)),$$

where

$$\mathfrak{S}_3 := \frac{1}{2} \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2} \right) \approx 0.3301.$$

The singular series \mathfrak{S}_3 is equal to $\frac{1}{4} \mathfrak{S}(A)$, where $A = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$, and is also half the twin prime constant.

Certain systems $\mathbf{Ap} = \mathbf{b}$ should be thought of as very difficult indeed, since their understanding implies an understanding of a binary problem such as the Goldbach or twin prime problem. If A has the property that every non-zero vector in its row span (over \mathbb{Q}) has at least three non-zero entries then there is no such reason to believe that it should be fantastically hard to solve.

Definition 2.3 (Non-degenerate systems). Suppose that s, t are positive integers with $t \geq s + 2$. We say that an $s \times t$ matrix A with integer entries is *non-degenerate* if it has rank s , and if every non-zero vector in its row span (over \mathbb{Q}) has at least three non-zero entries.

The reader may care to check that the system (1) defining a progression of length k is non-degenerate.

Our eventual goal is to prove Conjecture 2.1 for all non-degenerate systems. This goal may be subdivided into subgoals according to the value of s .

Conjecture 2.4 (Asymptotics for s simultaneous equations). Fix a value of $s \geq 1$ and suppose that $t \geq s + 2$ and that A is a non-degenerate $s \times t$ matrix. Then Conjecture 2.1 holds for the system $\mathbf{Ap} = 0$.

One can also formulate an appropriate conjecture for non-homogeneous systems $\mathbf{A}\mathbf{p} = \mathbf{b}$, and one would not expect to encounter significant extra difficulties in proving it. One might also try to count prime solutions to $\mathbf{A}\mathbf{p} = 0$ in which the primes p_i are subject to different constraints $p_i \leq N_i$, or perhaps are constrained to lie in a fixed arithmetic progression $p_i \equiv a_i \pmod{q_i}$. One would expect all of these extensions to be relatively straightforward.

The classical Hardy–Littlewood method can handle the case $s = 1$ of Conjecture 2.4. Our new developments have led to a solution of the case $s = 2$. In particular we can obtain an asymptotic for the number of 4-term arithmetic progressions of primes, all less than N :

Theorem 2.5 (Green–Tao [15]). *The number of quadruples of primes (p_1, p_2, p_3, p_4) , $p_1 < p_2 < p_3 < p_4 \leq N$, in arithmetic progression is*

$$\mathfrak{S}_4 N^2 \log^{-4} N (1 + o(1)),$$

where

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

3. The Hardy–Littlewood method for primes

The aim of this section is to describe the Hardy–Littlewood method as it would normally be applied to linear equations in primes. We will sketch the proof of Theorem 2.2, the asymptotic for the number of 3-term progressions of primes. This is equivalent to the $s = 1$ case of Conjecture 2.4 for the specific matrix $\mathbf{A} = \begin{pmatrix} 1 & -2 & 1 \end{pmatrix}$. Very similar means may be used to handle the general case $s = 1$ of that conjecture.

The Hardy–Littlewood method is, first and foremost, a method of harmonic analysis. The primes are studied by introducing the *exponential sum* (a kind of Fourier transform)

$$S(\theta) := \mathbb{E}_{n \leq N} \Lambda(n) e(\theta n)$$

for $\theta \in \mathbb{R}/\mathbb{Z}$, where $e(\alpha) := e^{2\pi i \alpha}$. It is the appearance of the circle \mathbb{R}/\mathbb{Z} here which gives the Hardy–Littlewood method its alternative name. Now it is easy to check that

$$\mathbb{E}_{x_1, x_2, x_3 \leq N} \Lambda(x_1) \Lambda(x_2) \Lambda(x_3) 1_{x_1 - 2x_2 + x_3 = 0} = \int_0^1 S(\theta)^2 S(-2\theta) d\theta.$$

whence

$$\mathbb{E}_{\substack{x_1, x_2, x_3 \leq N \\ x_1 - 2x_2 + x_3 = 0}} \Lambda(x_1) \Lambda(x_2) \Lambda(x_3) = (2N + O(1)) \int_0^1 S(\theta)^2 S(-2\theta) d\theta. \quad (3)$$

The method consists of gathering information about $S(\theta)$, and then using this formula to infer an asymptotic for the left-hand side.

The process of gathering information about $S(\theta)$ leads us to another key feature of the Hardy–Littlewood method: the realisation that one must split the set of θ into two classes, the *major arcs* \mathfrak{M} in which $\theta \approx a/q$ for some small q and the *minor arcs* $\mathfrak{m} := [0, 1) \setminus \mathfrak{M}$. To see why, let us attempt some simple evaluations. First of all we note that

$$S(0) := \mathbb{E}_{n \leq N} \Lambda(n) = 1 + o(1),$$

this being equivalent to the prime number theorem. To evaluate $S(1/2)$, observe that almost all of the support of Λ is on odd numbers n , for which $e(n/2) = -1$. Thus

$$S(1/2) := \mathbb{E}_{n \leq N} \Lambda(n)e(n/2) = -1 + o(1).$$

The evaluation of $S(1/3)$ is a little more subtle. Most of the support of Λ is on n not divisible by 3, and for those n the character $e(n/3)$ takes two values according as $n \equiv 1 \pmod{3}$ or $n \equiv 2 \pmod{3}$. We have

$$\begin{aligned} S(1/3) &= e(1/3)\mathbb{E}_{n \leq N} \mathbf{1}_{n \equiv 1 \pmod{3}} \Lambda(n) + e(2/3)\mathbb{E}_{n \leq N} \mathbf{1}_{n \equiv 2 \pmod{3}} \Lambda(n) + o(1) \\ &= -1/2 + o(1), \end{aligned}$$

this being a consequence of the fact that the primes are asymptotically equally divided between the congruence classes $1 \pmod{3}$ and $2 \pmod{3}$.

In similar fashion one can get an estimate for $S(a/q)$ for small q , and indeed for $S(a/q + \eta)$ for sufficiently small η , if one uses the prime number theorem in arithmetic progressions. The set of such θ is called the *major arcs* and is denoted \mathfrak{M} . (The notion of “small q ” might be $q \leq \log^A N$, for some fixed A . The notion of “small η ” might be $|\eta| \leq \log^A N/qN$. The flexibility allowed here depends on what type of prime number theorem along arithmetic progressions one is assuming. Unconditionally, the best such theorem is due to Siegel and Walfisz and it is this theorem which leads to these bounds on q and $|\eta|$.)

Suppose by contrast that $\theta \notin \mathfrak{M}$, that is to say θ is not close to a/q with q small. We say that $\theta \in \mathfrak{m}$, the *minor arcs*. It is hard to imagine that in the sum

$$S(\sqrt{2} - 1) = \mathbb{E}_{n \leq N} \Lambda(n)e(n\sqrt{2}) \tag{4}$$

the phases $e(n\sqrt{2})$ could conspire with $\Lambda(n)$ to prevent cancellation. It turns out that indeed there *is* substantial cancellation in this sum. This was first proved by Vinogradov, and nowadays it is most readily established using an identity of Vaughan [30], which allows one to decompose (4) into three further sums which are amenable to estimation. We will discuss a variant of this method in §5. For the particular value $\theta = \sqrt{2} - 1$, and for other highly irrational values, one can obtain an estimate of the shape $|S(\theta)| \ll N^{-c}$ for some $c > 0$, which is quite remarkable since applying the best-known error term in the prime number theorem only allows one to estimate $S(0)$ with the much larger error $O(\exp(-C_\varepsilon \log^{3/5-\varepsilon} N))$. By defining parameters suitably (that is by taking a suitable value of the constant A in the precise definition

of \mathfrak{M}), one can arrange that $S(\theta)$ is always very small indeed on the minor arcs \mathfrak{m} , say

$$\sup_{\theta \in \mathfrak{m}} |S(\theta)| \ll \log^{-10} N. \tag{5}$$

Recall now the formula (3). Splitting the integral into that over \mathfrak{M} and that over \mathfrak{m} , we see from Parseval’s identity that

$$\left| \int_{\mathfrak{m}} S(\theta)^2 S(-2\theta) d\theta \right| \leq \sup_{\theta \in \mathfrak{m}} |S(\theta)| \int_0^1 |S(\theta)|^2 d\theta \ll \frac{\log^{-9} N}{N}. \tag{6}$$

Thus in the effort to establish Theorem 2.2 the contribution from the minor arcs \mathfrak{m} may essentially be ignored. The proof of that theorem is now reduced to showing that

$$\int_{\mathfrak{M}} S(\theta)^2 S(-2\theta) d\theta = (1 + o(1)) \frac{1}{N} \prod_{p \geq 3} \left(1 - \frac{1}{(p-1)^2} \right).$$

Since one has asymptotic formulae for $S(\theta)$ (and $S(-2\theta)$) on \mathfrak{M} , this is essentially just a computation, albeit not a particularly straightforward one.

It is instructive to look for the point in the above argument where we used the fact that A was non-degenerate, that is to say that our problem had at least three variables. Why can we not use the same ideas to solve the twin prime or Goldbach problems? The answer lies in the bound (6). In the twin prime problem we would be looking to bound

$$\left| \int_{\theta \in \mathfrak{m}} |S(\theta)|^2 e(2\theta) \right|,$$

and the only obvious means of doing this is via an inequality of the form

$$\left| \int_{\theta \in \mathfrak{m}} |S(\theta)|^2 e(2\theta) \right| \leq \sup_{\theta \in \mathfrak{m}} |S(\theta)|^c \int_0^1 |S(\theta)|^{2-c} d\theta.$$

Now, however, Parseval’s identity does not permit one to place a bound on

$$\int_0^1 |S(\theta)|^{2-c} d\theta.$$

Indeed this whole endeavour is rather futile since heuristics predict that the minor arcs actually make a significant contribution to the asymptotic for twin primes.

An attempt to count 4-term progressions in primes via the circle method is beset by difficulties of a similar kind.

4. Exponential sums with Möbius

The presentation in the next two sections (and in our papers) is influenced by that in the beautiful book of Iwaniec and Kowalski [21].

In the previous section we described what is more-or-less the standard approach to solving linear equations in primes using the Hardy–Littlewood method. In [21, Ch. 19] one may find a very elegant variant in which the Möbius function μ is made to play a prominent rôle. As we saw above the behaviour of the exponential sum $S(\theta)$ was a little complicated to describe, depending as it does on how close to a rational θ is. By contrast the exponential sum

$$M(\theta) := \mathbb{E}_{n \leq N} \mu(n) e(\theta n)$$

has a very simple behaviour, as the following result of Davenport shows.

Proposition 4.1 (Davenport’s bound). *We have the estimate*

$$|M(\theta)| \ll_A \log^{-A} N$$

uniformly in $\theta \in [0, 1)$ for any $A > 0$.

In fact on the GRH Baker and Harman [1] obtain the superior bound $|M(\theta)| \ll N^{-3/4+\varepsilon}$. By analogy with results of Salem and Zygmund [26] concerning random trigonometric series one might guess that the truth is that $\sup_{\theta \in [0, 1)} |M(\theta)| \sim c\sqrt{\log N/N}$. This is far from known even on GRH; so far as I am aware no *lower* bound of the form $\sup_{\theta \in [0, 1)} |M(\theta)|\sqrt{N} \rightarrow \infty$ is known.

Although Davenport’s result is easy to describe its proof has the same ingredients as used in the analysis of $S(\theta)$. One must again divide \mathbb{R}/\mathbb{Z} into major and minor arcs. On the major arcs one must once more use information equivalent to a prime number theorem along arithmetic progressions, that is to say information on the zeros of L -functions $L(s, \chi)$ close to the line $\Re s = 1$. On the minor arcs one uses an appropriate version of Vaughan’s identity. One of the attractions of working with Möbius is that this identity takes a particularly simple form (see [21, Ch. 13] or [14]).

We offer a rough sketch of how Proposition 4.1 may be used as the main ingredient in a proof of Theorem 2.2, referring the reader to [21, Ch. 19] for the details. The key point is that one has the identity

$$\Lambda(n) = \sum_{d|n} \mu(d) \log(n/d).$$

One splits the sum over d into the ranges $d \leq N^{1/10}$ and $d > N^{1/10}$ (say), obtaining a decomposition $\Lambda = \Lambda^{\sharp} + \Lambda^{\flat}$. One has

$$S^{\flat}(\theta) := \mathbb{E}_{n \leq N} \Lambda^{\flat}(n) e(n\theta) = \sum_{d \leq N^{1/10}} \log d \sum_{N^{1/10} \leq k \leq N/d} \mu(k) e(\theta kd),$$

from which it follows easily using Davenport’s bound that

$$S^{\flat}(\theta) \ll_A \log^{-A} N \tag{7}$$

uniformly in $\theta \in [0, 1)$.

One may then write the expression

$$\mathbb{E}_{\substack{x_1, x_2, x_3 \leq N \\ x_1 - 2x_2 + x_3 = 0}} \Lambda(x_1)\Lambda(x_2)\Lambda(x_3)$$

as a sum of eight terms using the splitting $\Lambda = \Lambda^{\sharp} + \Lambda^{\flat}$. The basic idea is now that the main term $\prod_p \alpha_p$ in Theorem 2.2 comes from the term with three copies of Λ^{\sharp} , whilst the other 7 terms (each of which contains at least one Λ^{\flat}) provide a negligible contribution in view of (7) and simple variants of the formula (3).

We have extolled the virtues of the Möbius function by pointing to the aesthetic qualities of Davenport’s bound. A more persuasive argument for focussing on it is the following basic metaprinciple of analytic number theory:

Principle (Möbius randomness law). *The Möbius function is highly orthogonal to any “reasonable” bounded function $f: \mathbb{N} \rightarrow \mathbb{C}$. That is to say*

$$\mathbb{E}_{n \leq N} \mu(n) f(n) = o(1),$$

and usually one would in fact expect

$$\mathbb{E}_{n \leq N} \mu(n) f(n) \ll N^{-1/2+\varepsilon}. \tag{8}$$

In the category “reasonable” in this context one would certainly include polynomials and other somewhat continuous objects, but one should exclude functions f which are closely related to the primes ($f = \mu$ and $f = \Lambda$, for example, are clearly not orthogonal to Möbius).

At a finer level than is relevant to our work, the Möbius randomness law is more reliable than other heuristics that one might formulate, for example concerning Λ . In [22] it is shown that

$$\mathbb{E}_{n \leq N} \Lambda(n) \lambda(n) e(-2\sqrt{n}) \sim cN^{-1/4},$$

where $\lambda(n) := n^{-11/2} \tau(n)$ is a normalised version of Ramanujan’s τ -function. One could hardly be called naïve for expecting square root cancellation here.

5. Proving the Möbius randomness law

In the last section we mentioned a principle, the *Möbius randomness law*, which is very useful as a guiding principle in analytic number theory. Unfortunately it is not possible to prove the strong version (8) of the principle in any case – even when $f(n) \equiv 1$ it is equivalent to the Riemann hypothesis.

It is, however, possible to prove weaker estimates of the form

$$\mathbb{E}_{n \leq N} \mu(n) f(n) \ll_A \log^{-A} N, \tag{9}$$

for arbitrary $A > 0$, for a wide variety of functions f . Davenport’s bound is precisely this result when $f(n) = e(\theta n)$ (and, furthermore, this result is uniform in θ). Similar statements are also known for polynomial phases and for Dirichlet characters (uniformly over all characters of a fixed conductor).

Now when it comes to proving an estimate of the form (9), one should think of there being two different classes of behaviour for f . In the first class are those f which are in a vague sense multiplicative, or linear combinations of a few multiplicative functions. Then the behaviour of $\mathbb{E}_{n \leq N} \mu(n) f(n)$ can be intimately connected with the zeros of L -functions. One has, for example, the formula

$$\sum_{n=1}^{\infty} \mu(n) \chi(n) n^{-s} = \frac{1}{L(s, \chi)}$$

for any fixed Dirichlet character χ . By the standard contour integration technique (Perron’s formula) of analytic number theory one sees that $\mathbb{E}_{n \leq N} \mu(n) \chi(n)$ is small provided that $L(s, \chi)$ does not have zeros close to $\Re s = 1$. (In fact, as reported on [21, p. 124], there are complications caused by possible multiple zeros of L , and it is better to work first with the sum $\mathbb{E}_{n \leq N} \Lambda(n) \chi(n)$ of χ over primes.)

The need to consider zeros of L -functions can also be felt when considering *additive* characters $e(an/q)$, for relatively small q . Indeed any Dirichlet character to the modulus q may be expressed as a linear combination of such characters. Conversely any additive character $e(an/q)$ may be written as a linear combination of Dirichlet characters to moduli dividing q by using Gauss sums. By applying Siegel’s theorem, which gives the best unconditional information concerning the location of zeros of $L(s, \chi)$ near to $\Re s = 1$, one obtains for any A the estimate

$$\mathbb{E}_{n \leq N} \mu(n) e(an/q) \ll_A \log^{-A} N,$$

uniformly for $q \leq \log^A N$. By partial summation the same estimate holds when a/q is replaced by $\theta = a/q + \eta$ for suitably small η , that is to say for all θ which lie in the set \mathfrak{M} of major arcs.

We turn now to a completely different technique for bounding $\mathbb{E}_{n \leq N} \mu(n) f(n)$. Remarkably this is at its most effective when the previous technique fails, that is to say when f is somehow *far* from multiplicative.

Proposition 5.1 (Type I and II sums control sums with Möbius). *Let $f : \mathbb{N} \rightarrow \mathbb{C}$ be a function with $\|f\|_{\infty} \leq 1$, and suppose that the following two estimates hold.*

1. (Type I sums are small) *For all $D \leq N^{2/3}$, and for all sequences $(a_d)_{d=D}^{2D}$ with $\|a\|_{l^2[D, 2D]} = 1$, we have*

$$\left| \sum_{d=D}^{2D} \sum_{1 \leq w < N/d} a_d f(wd) \right| \ll_A N (\log N)^{-A-3}. \tag{10}$$

2. (Type II sums are small) For all $D, W, N^{1/3} \leq D \leq N^{2/3}, N^{1/3} \leq W \leq N/D$ and all choices of complex sequences $(a_d)_{d=D}^{2D}, (b_w)_{w=W}^{2W}$ with $\|a\|_{l^2[D,2D]} = \|b\|_{l^2[W,2W]} = 1$, we have

$$\left| \sum_{d=D}^{2D} \sum_{W \leq w \leq 2W} a_d b_w f(wd) \right| \ll_A N(\log N)^{-A-5}. \tag{11}$$

Then

$$\mathbb{E}_{n \leq N} \mu(n) f(n) \ll_A \log^{-A} N. \tag{12}$$

The reader may find a proof of this statement in [14, Ch. 6]. It is proved by decomposing the Möbius function into two parts using an identity of Vaughan [30]. When one multiplies by $f(n)$ and sums, one of these parts leads to Type I sums and the other to Type II sums. Note that there is considerable flexibility in arranging the ranges of D in which Type I and II estimates are required, but it is not important to have such flexibility in our arguments.

The statement of Proposition 5.1 may look complicated. What has been achieved, however, is the elimination of μ . Strictly speaking, one actually only *needs* Type I and II estimates for some rather specific choices of coefficients a_d, b_w whose definition involves μ . The important realisation is that it is best to forget about the precise forms of these coefficients, the general expressions (10) and (11) laying bare the important underlying information required of f .

Note that if f is close to multiplicative then there is no hope of obtaining enough cancellation in Type II sums to make use of Proposition 5.1. If f is actually completely multiplicative, for example, one may take $a_d = \overline{f(d)}$ and $b_w = \overline{f(w)}$ and there is manifestly no cancellation at all in (11). If this is not the case, however, then very often it is possible to verify the bounds (10) and (11). An example of this is a linear phase $e(\theta n)$ where θ lies in the minor arcs \mathfrak{m} , that is to say θ is not close to a/q with q small. By verifying these two estimates for such θ , one has from (12) that Davenport’s bound holds when $\theta \in \mathfrak{m}$. This completes the proof of Davenport’s bound, since the major arcs \mathfrak{M} have already been handled using L -function technology.

To see how this is usually achieved in practice we refer the reader to [5, Ch. 24]. There the reader will see that a key device is the Cauchy–Schwarz inequality, which allows one to eliminate the arbitrary coefficients a_d, b_w .

In [14] there is also a discussion of this result. Although logically equivalent, this discussion takes a point of view which turns out to be invaluable when dealing with more complicated situations. Taking $f(n) = e(\theta n)$ in Proposition 5.1, we suppose that either (10) or (11) does not hold, that is to say that either a Type I or a Type II sum is large. We then *deduce* that θ must be close to a rational with small denominator, that is to say θ must be major arc. This *inverse* approach to bounding sums with Möbius means that there is no need to make an *a priori* definition of what a “major” or “minor” object is. In situations to be discussed later this helps enormously.

6. The insufficiency of harmonic analysis

What did we mean when we stated that the Hardy–Littlewood method was a method of harmonic analysis? In §3 we saw that there is a formula, (3), which expresses the number of 3-term progressions in a set (such as the primes) in terms of the exponential sum over that set. The following proposition is an easy consequence of a slightly generalised version of that formula:

Proposition 6.1. *Suppose that $f_1, f_2, f_3: [N] \rightarrow [-1, 1]$ are three functions and that*

$$\left| \mathbb{E}_{\substack{x_1, x_2, x_3 \\ x_1 - 2x_2 + x_3 = 0}} f_1(x_1) f_2(x_2) f_3(x_3) \right| \geq \delta.$$

Then for any $i = 1, 2, 3$ we have

$$\sup_{\theta \in [0, 1)} \left| \mathbb{E}_{n \leq N} f_i(n) e(n\theta) \right| \geq (1 + o(1))\delta/2. \quad (13)$$

We think of this as a statement the effect that the linear exponentials $e(n\theta)$ form a *characteristic system* for the linear equation $x_1 - 2x_2 + x_3 = 0$. It follows immediately from Proposition 6.1 and Davenport’s bound that Möbius exhibits cancellation along 3-term APs, in the sense that

$$\mathbb{E}_{\substack{x_1, x_2, x_3 \\ x_1 - 2x_2 + x_3 = 0}} \mu(x_1) \mu(x_2) \mu(x_3) \ll_A \log^{-A} N.$$

Proposition 6.1 is also useful for counting progressions in sets $A \subseteq [N]$, in which context one would take various of the f_i to equal the *balanced function* $f_A := 1_A - \alpha$ of A , where $\alpha := |A|/N$. It is easy to deduce from Proposition 6.1 the following variant, which covers this situation.

Proposition 6.2. *Suppose that $A \subseteq [N]$ is a set with $|A| = \alpha N$ and that*

$$\left| \mathbb{E}_{\substack{x_1, x_2, x_3 \\ x_1 - 2x_2 + x_3 = 0}} 1_A(x_1) 1_A(x_2) 1_A(x_3) - \alpha^3 \right| \geq \delta.$$

Write

$$f_A := 1_A - \alpha$$

for the balanced function of A . Then we have

$$\sup_{\theta \in [0, 1)} \left| \mathbb{E}_{n \leq N} f_A(n) e(n\theta) \right| \geq (1 + o(1))\delta/14. \quad (14)$$

If a function f correlates with a linear exponential as in (13) or (14) then we sometimes say that f has *linear bias*.

In this section we give examples which show that the linear exponentials do not form a characteristic system for the pair of equations $x_1 - 2x_2 + x_3 = x_2 - 2x_3 + x_4 = 0$ defining a four-term progression. These examples show, in a strong sense, that the Hardy–Littlewood method in its traditional form *cannot* be used to study 4-term

progressions. An interesting feature of these two examples is that they were both essentially discovered by Furstenberg and Weiss [6] in the context of ergodic theory. Much of our work is paralleled in, and in fact motivated by, the work of the ergodic theory community. See the lecture by Tao in Volume II of these proceedings, or the elegant surveys of Kra [23], [24] for more discussion and references. The examples were rediscovered, in the finite setting, by Gowers [8], [10] in his work on Szemerédi’s theorem.

Example 6.1 (Quadratic and generalised quadratic behaviour). Let $\alpha > 0$ be a small, fixed, real number, and define the following sets. Let A_1 be defined by

$$A_1 := \{x \in [N] : \{x^2\sqrt{2}\} \in [-\alpha/2, \alpha/2]\}$$

(here, $\{t\}$ denotes the fractional part of t , and lies in $(-1/2, 1/2]$). Define also

$$A_2 := \{x \in [N] : \{x\sqrt{2}\{x\sqrt{3}\}\} \in [-\alpha/2, \alpha/2]\}.$$

Now it can be shown (not altogether straightforwardly) that $|A_1|, |A_2| \approx \alpha N$, and furthermore that

$$\sup_{\theta \in [0,1)} |\mathbb{E}_{n \leq N} f_A(n)e(n\theta)| \ll N^{-c}$$

for $i = 1, 2$. Thus neither of the sets A_1, A_2 has linear bias in a rather strong sense. If the analogue of Proposition 6.2 were true for four term progressions, then, one would expect both A_1 and A_2 to have approximately $\alpha^4 N^2/6$ four-term progressions.

The set A_1 , however, has considerably more 4-term APs than this in view of the identity

$$x^2 - 3(x + d)^2 + 3(x + 2d)^3 - (x + 3d)^2 = 0. \tag{15}$$

This means that if $x, x + d, x + 2d \in A_1$ then

$$\{(x + 3d)^2\sqrt{2}\} \in [-7\alpha/2, 7\alpha/2],$$

which would suggest that $x + 3d \in A_1$ with probability $\gg 1$. In fact one can show using harmonic analysis that (15) is the only relevant constraint in the sense that

$$\begin{aligned} &\mathbb{P}(x + 3d \in A_1 | x, x + d, x + 2d \in A_1) \\ &\approx \mathbb{P}(y_1 - 3y_2 + 3y_3 \in [-1, 1] | y_1, y_2, y_3 \in [-1, 1]) = 8/27. \end{aligned}$$

The number of 3-term progressions in A_1 is $\approx \alpha^3 N/4$, and so it follows that the number of 4-term progressions in A_1 is $\approx 2\alpha^3/27$.

The analysis of A_2 is rather more complicated. However one may check that if $|\{x\sqrt{3}\}|, |\{d\sqrt{3}\}| \leq 1/10$ and if $|\{y\sqrt{2}\{y\sqrt{3}\}\}| \leq \alpha/10$ for $y = x, x + d, x + 2d$, then $x + 3d \in A_2$. One can show that there are $\gg \alpha^3 N^2$ choices of x, d satisfying these constraints, and hence once again A_2 contains $\gg \alpha^3 N^2$ 4-term progressions.

7. Generalised quadratic obstructions

We saw in the last section that the set of linear exponentials $e(\theta n)$ is not a characteristic system for 4-term progressions. There we saw examples involving quadratics $n^2\theta$ and generalised quadratics $n\theta_1\{n\theta_2\}$, and these must clearly be addressed by any generalisation of Propositions 6.1 and 6.2 to 4-term APs. Somewhat remarkably, these quadratic and generalised quadratic examples are in a sense the only ones.

Proposition 7.1. *Suppose that $f_1, f_2, f_3, f_4: [N] \rightarrow [-1, 1]$ are four functions and that*

$$|\mathbb{E}_{\substack{x_1, x_2, x_3, x_4 \\ x_1 - 2x_2 + x_3 = 0 \\ x_2 - 2x_3 + x_4 = 0}} f_1(x_1) f_2(x_2) f_3(x_3) f_4(x_4)| \geq \delta. \quad (16)$$

Then for any $i = 1, 2, 3, 4$ there is a generalised quadratic polynomial

$$\phi(n) = \sum_{r, s \leq C_1(\delta)} \beta_{rs} \{\theta_r n\} \{\theta_s n\} + \gamma_r \{\theta_r n\}, \quad (17)$$

where $\beta_{rs}, \gamma_r, \theta_r \in \mathbb{R}$, such that

$$|\mathbb{E}_{n \leq N} f_i(n) e(\phi(n))| \geq c_2(\delta).$$

We can take $C_1(\delta) \sim \exp(\delta^{-C})$ and $c_2(\delta) \sim \exp(-\delta^{-C})$.

Note that

$$\theta n^2 = 100\theta N^2 \left\{ \frac{n}{10N} \right\}^2$$

and

$$\theta_1 n \{\theta_2 n\} = 10\theta_1 N \left\{ \frac{n}{10N} \right\} \{\theta_2 n\}$$

for $n \leq N$, and so the phases which can be written in the form (17) do include all those which were discovered to be relevant in the preceding section.

The proof of Proposition 7.1 is given in [13]. It builds on earlier work of Gowers [8], [10]. In [13] (see also [14]) several results of a related nature are given, in which other characteristic systems for the equation $x_1 - 2x_2 + x_3 = x_2 - 2x_3 + x_4 = 0$ are given. These systems all have a “quadratic” flavour. We will discuss the family of *2-step nilsequences*, which is perhaps the most conceptually appealing, in §9. In §11 we will mention the family of *local quadratics*, which are useful for computations involving the Möbius function. The only real merit of the generalised quadratic phases $e(\phi(n))$ discussed above is that they are easy to describe from first principles.

8. The Gowers norms and inverse theorems

The proof of Proposition 7.1 is long and complicated: there does not seem to be anything so simple as Formula (3) in the world of 4-term progressions. Very roughly

speaking one assumes that (16) holds, and then one proceeds to place more and more structure on each function f_i until eventually one establishes that f_i correlates with a generalised quadratic phase. There is a finite field setting for this argument, and we would recommend that the interested reader read this first: it may be found in [13, Ch. 5]. The ICM lecture of Gowers [9] is a fine introduction to the ideas in his paper [8], which is the foundation of our work.

There is only one part of the existing theory which we feel sure will play some rôle in future incarnations of these methods. This is the first step in the long series of deductions from (16), in which one shows that each f_i has large *Gowers norm*. For the purposes of this exposition¹ we define the Gowers U^2 -norm $\|f\|_{U^2}$ of a function $f: [N] \rightarrow [-1, 1]$ by

$$\|f\|_{U^2}^4 := \mathbb{E}_{\substack{x_{00}, x_{01}, x_{10}, x_{11} \leq N \\ x_{00} + x_{11} = x_{01} + x_{10}}} f(x_{00})f(x_{01})f(x_{10})f(x_{11}),$$

which is a sort of average of f over two dimensional parallelograms. The U^k norm, $k \geq 3$, is an average of f over k -dimensional parallelepipeds. Written down formally it looks much more complicated than it is:

$$\|f\|_{U^k}^{2^k} := \mathbb{E}_{\substack{x_{0,\dots,0}, \dots, x_{1,\dots,1} \\ x_{\omega(1)} + x_{\omega(2)} = x_{\omega(3)} + x_{\omega(4)}}} f(x_{0,\dots,0}) \dots f(x_{1,\dots,1}),$$

where there are 2^k variables x_ω , $\omega = (\omega_1, \dots, \omega_k) \in \{0, 1\}^k$, and the constraints range over all quadruples $(\omega^{(1)}, \omega^{(2)}, \omega^{(3)}, \omega^{(4)}) \in (\{0, 1\}^k)^4$ with $\omega^{(1)} + \omega^{(2)} = \omega^{(3)} + \omega^{(4)}$.

The Gowers U^k norm governs the behaviour of any non-degenerate system $A\mathbf{x} = 0$ in which A has $(k - 1)$ rows.

Proposition 8.1 (Generalised von Neumann theorem). *Suppose that A is a non-degenerate $s \times t$ matrix with integer entries. Suppose that $f_1, \dots, f_t: [N] \rightarrow [-1, 1]$ are functions and that*

$$|\mathbb{E}_{\substack{x_1, \dots, x_t \\ A\mathbf{x} = 0}} f_1(x_1) \dots f_t(x_t)| \geq \delta.$$

Then for each $i = 1, \dots, t$ we have

$$\|f_i\|_{U^{s+1}} \geq c_A \delta.$$

The proof involves $s + 1$ applications of the Cauchy–Schwarz inequality. In this generality, the result was obtained in [14], though the proof technique is the same as in [10]. There are results in ergodic theory of the same general type, in which “non-conventional ergodic averages” are bounded using seminorms which are analogous to the U^k -norms: see [20].

¹In practice we do all our work on the group $\mathbb{Z}/N'\mathbb{Z}$ for some prime $N' \geq N$ with $N' \approx M(A)N$, where $M(A)$ is some constant depending on the system of equations $A\mathbf{x} = 0$ one is interested in. One advantage of this is that the number of solutions to $A\mathbf{x} = 0$ in $\mathbb{Z}/N'\mathbb{Z}$ is much easier to count than the number of solutions in $[N]$. The Gowers norms defined here differ from the Gowers norms in those settings by constant factors, so for expository purposes they may be thought of as the same. In the group setting the constant c_A in Proposition 8.1 is simply 1.

Taking $s = k - 2$ and \mathbf{A} as in (1), we see that in particular the Gowers U^{k-1} -norm “controls” k -term progressions. The Gowers norms are, of course, themselves defined by a system of linear equations, and so they must be studied as part of a generalised Hardy–Littlewood method with as broad a scope as we would like. The Generalised von Neumann Theorem may be regarded as a statement to the effect that in a sense they represent the *only* systems of equations that need to be studied.

The Gowers norms do not feature in the classical Hardy–Littlewood method. It is, however, possible to prove a somewhat weaker version of Proposition 6.1 by combining the case $k = 3$ of Proposition 8.1 with the following *inverse theorem*:

Proposition 8.2 (Inverse theorem for U^2). *Suppose that N is large and that $f : [N] \rightarrow [-1, 1]$ is a function with $\|f\|_{U^2} \geq \delta$. Then we have*

$$\sup_{\theta \in [0, 1)} |\mathbb{E}_{n \leq N} f(n)e(n\theta)| \geq 2\delta^2.$$

To prove this we note the formula

$$\mathbb{E}_{x_{00}, x_{01}, x_{10}, x_{11}} f(x_{00})f(x_{01})f(x_{10})f(x_{11})1_{x_{00}+x_{11}=x_{01}+x_{10}} = \int_0^1 |\hat{f}(\theta)|^4 d\theta,$$

where $\hat{f}(\theta) := \mathbb{E}_{n \leq N} f(n)e(n\theta)$. This implies that

$$\|f\|_{U^2}^4 = (3N + O(1))\|\hat{f}\|_4^4.$$

In view of the fact that $\|\hat{f}\|_2^2 \leq 1/N$, this and the assumption that $\|f\|_{U^2} \geq \delta$ imply that

$$\|\hat{f}\|_\infty^2 \geq (3 + o(1))\delta^4,$$

which implies the result.

This argument should be compared to the argument in (6), to which it corresponds rather closely.

To deduce Proposition 6.1 by passing through Proposition 8.2 is rather perverse, since the derivation is longer than the one that proceeds via an analogue of (3) and it leads to worse dependencies. With our current technology, however, this is the only method which is amenable to generalisation.

Similarly, one may deduce Proposition 7.1 from Proposition 8.1 and the following result.

Proposition 8.3 (Inverse theorem for the U^3 -norm). *Suppose that $f : [N] \rightarrow \mathbb{R}$ is a function for which $\|f\|_\infty \leq 1$ and $\|f\|_{U^3} \geq \delta$. Then there is a generalised quadratic phase*

$$\phi(n) = \sum_{r,s \leq C_1(\delta)} \beta_{rs}\{\theta_r n\}\{\theta_s n\} + \gamma_r\{\theta_r n\}, \tag{18}$$

where $\beta_{rs}, \gamma_r, \theta_r \in \mathbb{R}$, such that

$$|\mathbb{E}_{n \leq N} f(n)e(\phi(n))| \geq c_2(\delta).$$

We can take $C_1(\delta) \sim \exp(\delta^{-C})$ and $c_2(\delta) \sim \exp(-\delta^{-C})$.

This result (and variations of it involving other “quadratic families”) is in fact the main theorem in [13].

As we mentioned, one may find a series of seminorms which are analogous to the Gowers norms in the ergodic-theoretic work of Host and Kra [20]. There are no such seminorms in the related work of Ziegler [35], however, and this suggests that (as in the classical case) the Gowers norms may not be completely fundamental to a generalised Hardy–Littlewood method.

9. Nilsequences

In the previous section we introduced the Gowers U^k -norms, and stated inverse theorems for the U^2 - and U^3 -norms. These inverse theorems provide lists of rather algebraic functions which are *characteristic* for a given system of equations $A\mathbf{x} = 0$. Roughly speaking, the linear phases $e(\theta n)$ are characteristic for single linear equations in which A is a $1 \times t$ matrix. Generalised quadratic phases $e(\phi(n))$ are characteristic for pairs of linear equations in which A is a non-degenerate $2 \times t$ matrix.

These two results leave open the question of whether there is a similar list of functions which is characteristic for the U^k -norm, $k \geq 4$ and hence, by the Generalised von Neumann Theorem, for non-degenerate systems defined by an $s \times t$ matrix with $s \geq 3$. The form of Propositions 8.2 and 8.3 does not suggest a particularly natural form for such a result, however, and indeed Proposition 8.2 is already rather unnatural-looking.

To make more natural statements, we introduce a class of functions called *nilsequences*.

Definition 9.1. Let G be a connected, simply connected, k -step nilpotent Lie group. That is, the central series $G_0 := G$, $G_{i+1} = [G, G_i]$ terminates with $G_k = \{e\}$. Let $\Gamma \subseteq G$ be a discrete, cocompact subgroup. The quotient G/Γ is then called a k -step nilmanifold. The group G acts on G/Γ via the map $T_g(x\Gamma) = xg\Gamma$. If $F: G/\Gamma \rightarrow \mathbb{C}$ is a bounded, Lipschitz function and $x \in G/\Gamma$ then we refer to the sequence $(F(T_g^n \cdot x))_{n \in \mathbb{N}}$ as a k -step nilsequence.

By analogy with the results of Host and Kra [20] in ergodic theory, we expect the collection of $(k - 1)$ -step nilsequences to be characteristic for the U^k -norm. The following conjecture is one of the guiding principles of the generalised Hardy–Littlewood method.

Conjecture 9.2 (Inverse conjecture for U^k -norms). Suppose that $k \geq 2$ and that $f: [N] \rightarrow [-1, 1]$ has $\|f\|_{U^k} \geq \delta$. Then there is a $(k - 1)$ -step nilmanifold G/Γ with dimension at most $C_{1,k}(\delta)$, together with a function $F: G/\Gamma \rightarrow \mathbb{C}$ with $\|F\|_\infty \leq 1$ and Lipschitz constant at most $C_{2,k}(\delta)$ and elements $g \in G$, $x \in G/\Gamma$ such that

$$|\mathbb{E}_{n \leq N} f(n) F(T_g^n \cdot x)| \geq c_{3,k}(\delta). \tag{19}$$

We can at least be sure that Conjecture 9.2 is no more complicated than necessary, since in [13, Ch. 12] we showed that if a bounded function f correlates with a $(k-1)$ -step nilsequence as in (19) then f *does* have large Gowers U^k -norm. This, incidentally, is another reason to believe that the Gowers norms play a fundamental rôle in the theory. It is not the case that correlation of a function f with a $(k-1)$ -step nilsequence prohibits f from enjoying cancellation along k -term arithmetic progressions, for example. In the case $k=3$ an example of this phenomenon is given by the function f which equals α for $1 \leq n \leq N/3$ and -1 for $N/3 < n \leq N$, where α is the root between 1 and 2 of $\alpha^3 - \alpha^2 + 3\alpha - 4 = 0$. This f correlates with the constant nilsequence 1 yet exhibits cancellation along 3-term progressions, as the reader may care to check.

Conjecture 9.2 seems, at first sight, to be completely unrelated to Propositions 8.2 and 8.3. However after a moment's thought one realises that a linear phase $e(\theta n)$ can be regarded as a 1-step nilsequence in which $G = \mathbb{R}$, $\Gamma = \mathbb{Z}$, $g = \theta$ and $x = 0$. Thus Proposition 8.2 immediately implies the case $k=2$ of Conjecture 9.2.

The case $k=3$ is proved in [13]. One first proves Proposition 8.3, and then one shows how any generalised quadratic phase $e(\phi(n))$ may be approximated by a 2-step nilsequence. Let us discuss a simple example, the *Heisenberg nilmanifold*, to convince the reader that 2-step nilsequences can give rise to “generalised quadratic” behaviour.

Example 9.1 (The Heisenberg nilmanifold). Consider

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

Then G/Γ is a 2-step nilmanifold. By using the identification

$$(x, y, z) \equiv \begin{pmatrix} 1 & x & y \\ 0 & 1 & z \\ 0 & 0 & 1 \end{pmatrix} \Gamma,$$

we can identify G/Γ (as a set) with \mathbb{R}^3 , quotiented out by the equivalence relations

$$(x, y, z) \sim (x + a, y + b + cx, z + c) \quad \text{for all } a, b, c \in \mathbb{Z}.$$

This can in turn be coordinatised by the cylinder $(\mathbb{R}/\mathbb{Z})^2 \times [-1/2, 1/2]$ with the identification $(x, y, -1/2) \sim (x, x + y, 1/2)$. Let $F: G/\Gamma \rightarrow \mathbb{C}$ be a function. We may lift this to a function $\tilde{F}: G \rightarrow \mathbb{C}$, defined by $\tilde{F}(g) := F(g\Gamma)$. In coordinates, this lift takes the form

$$\tilde{F}(x, y, z) = F(x \pmod{1}, y - [z]x \pmod{1}, \{z\})$$

where $[z] = z - \{z\}$ is the nearest integer to x . Let

$$g := \begin{pmatrix} 1 & \alpha & \beta \\ 0 & 1 & \gamma \\ 0 & 0 & 1 \end{pmatrix}$$

be an element of G . Then the shift $T_g : G \rightarrow G$ is given by

$$T_g(x, y, z) = (x + \alpha, y + \beta + \gamma x, z + \gamma).$$

A short induction confirms, for example, that

$$T_g^n(0, 0, 0) = (n\alpha, n\beta + \frac{1}{2}n(n+1)\alpha\gamma, n\gamma).$$

Therefore if $F : G/\Gamma \rightarrow G/\Gamma$ is any Lipschitz function, written as a function $F : (\mathbb{R}/\mathbb{Z})^2 \times [-1/2, 1/2] \rightarrow \mathbb{C}$ with $F(-1/2, y, z) = F(1/2, y + z, z)$, then we have

$$F(T_g^n(0, 0, 0)) = F(n\alpha \pmod{1}, n\beta + \frac{1}{2}n(n+1)\alpha\gamma - [n\gamma]n\alpha \pmod{1}, \{n\gamma\}).$$

The term $[n\gamma]n\alpha$ which appears here certainly exhibits a sort of generalised quadratic behaviour. For a complete description of how an arbitrary generalised quadratic phase $e(\phi(n))$ can be approximated by a two-step nilsequence, we refer the reader to [13, Ch. 12].

Let us conclude this section by stating, for the reader’s convenience, a result/conjecture which summarises much of our discussion so far in one place.

Theorem 9.3 (Green–Tao [13]). *We have the following two statements.*

- (i) (Generalised von Neumann) *Suppose that s and t are positive integers with $s + 2 \leq t$. Suppose that A is a non-degenerate $s \times t$ matrix with integer entries. Suppose that $f_1, \dots, f_t : [N] \rightarrow [-1, 1]$ are functions and that*

$$|\mathbb{E}_{\substack{x_1, \dots, x_t \\ Ax=0}} f_1(x_1) \dots f_t(x_t)| \geq \delta. \tag{20}$$

Then for each $i = 1, \dots, t$ we have

$$\|f_i\|_{U^{s+1}} \geq c_A \delta.$$

- (ii) (Gowers inverse result: proved for $k = 2, 3$, conjectural for $k \geq 4$) *Suppose that $f : [N] \rightarrow [-1, 1]$ has $\|f\|_{U^k} \geq \delta$. Then there is a $(k-1)$ -step nilmanifold G/Γ with dimension at most $C_{1,k}(\delta)$, together with a function $F : G/\Gamma \rightarrow \mathbb{C}$ with $\|F\|_\infty \leq 1$ and Lipschitz constant at most $C_{2,k}(\delta)$ and elements $g \in G$, $x \in G/\Gamma$ such that*

$$|\mathbb{E}_{n \leq N} f(n) F(T_g^n \cdot x)| \geq c_{3,k}(\delta). \tag{21}$$

In particular when $s = 1$ or 2 and (20) holds for some A and some δ then for each $i = 1, \dots, t$ there is a 2-step nilsequence $(F(T_g^n \cdot x))_{n \in \mathbb{N}}$ such that

$$|\mathbb{E}_{n \leq N} f_i(n) F(T_g^n \cdot x)| \geq c_A(\delta). \tag{22}$$

10. Working with the primes

Let us suppose that we wish to count four-term progressions in the primes. One might try to apply Theorem 9.3 with the functions f_i equal to the balanced function of A , the set of primes $p \leq N$, and then hope to rule out a correlation such as (21) for some $\delta = o(\alpha^t)$ (here, of course, $\alpha \approx \log^{-1} N$ by the prime number theorem). This would then lead to an asymptotic using various instances of (20) together with the triangle inequality.

There are two reasons why this is a hopeless strategy. First of all, the primes *do* correlate with nilsequences. In fact since all primes other than 2 are odd it is easy to see that

$$\mathbb{E}_{n \leq N} f_A(n) e(n/2) \approx -\alpha.$$

There is a way to circumvent this problem, which we call the W -trick. The idea is that if $W = 2 \times 3 \times \dots \times w(N)$ is the product of the first several primes, then for any b coprime to W the set

$$A_b := \{n \leq N : Wn + b \text{ is prime}\}$$

does not exhibit significant bias in progressions with common difference $q \leq w(N)$. One can then count 4-term progressions in the primes by counting 4-term progressions in $A_{b_1} \times \dots \times A_{b_4}$ for each quadruple $(b_1, \dots, b_4) \in (\mathbb{Z}/W\mathbb{Z})^{\times 4}$ in arithmetic progression and adding.

We refer to any set A_b as a set of “ W -tricked primes”. In practice one is only free to take $w(N) \sim \log \log N$, since one must be able to understand the distribution of primes in progressions with common difference W (note that even on GRH one could only take $w(N) \sim c \log N$). Even assuming we could obtain optimal results concerning the correlation of the W -tricked primes with 2-step nilsequences, this information will be very weak indeed.

This highlights a more serious problem with the suggested strategy. Suppose that $A \subseteq [N]$ is a set of density α for which there is no obvious reason why A should have an unexpectedly large or small number of 4-term APs, that is to say for which we might hope to prove that

$$\mathbb{E}_{\substack{x_1, x_2, x_3, x_4 \\ x_1 - 2x_2 + x_3 = 0 \\ x_2 - 2x_3 + x_4 = 0}} 1_A(x_1) 1_A(x_2) 1_A(x_3) 1_A(x_4) \approx \alpha^4. \tag{23}$$

For example, A might be the W -tricked primes less than N , in which case $\alpha \sim \frac{W}{\phi(W)} \log^{-1} N$.

We might prove (23) by writing $1_A = \alpha + f_A$, expanding as the sum of sixteen terms, and showing that fifteen of these are $o(\alpha^4)$ by appealing to Theorem 9.3, and ruling out a correlation with a 2-step nilsequence as in (22). Unfortunately we will be operating with $\delta = o(\alpha^4) \ll \log^{-4+\varepsilon} N$, and the dependence of $c_A(\delta)$ on δ is very weak, being of the form $\exp(-\delta^{-C})$. Thus we are asking to rule out the possibility

that

$$|\mathbb{E}_{n \leq N} f_A(n) F(T_g^n \cdot x)| \gg \exp(-\log^C N)$$

for some potentially rather large C . This is a problem, since one would never expect more than square root cancellation in any such expression. In fact for the W -tricked primes one only has a small amount (depending on $w(N)$) of potential cancellation to work with and to all intents and purposes one should not bank on having available any estimate stronger than

$$\mathbb{E}_{n \leq N} f_A(n) F(T_g^n \cdot x) = o(1).$$

What one really needs is a version of Proposition 16 which applies to functions which need not be bounded by 1. Then one could hope to work with the von Mangoldt function Λ instead of the far less natural characteristic function 1_A , or more accurately with W -tricked variants of the von Mangoldt function such as

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

Such a result is the main result of our forthcoming paper [15]. It would take us too far afield to say anything concerning its proof, other than that it uses one of the key tools from our paper [12] on long progressions of primes, the “ergodic transference” technology of [12, Chs. 6,7,8].

Proposition 10.1 (Transference principle, [15]). *Suppose that $\nu: [N] \rightarrow \mathbb{R}^+$ is a pseudorandom measure. Then*

- (i) *The generalised von Neumann theorem, Theorem 9.3 (i), continues to hold for functions $f_1, \dots, f_t: [N] \rightarrow \mathbb{R}^+$ such that $|f_i(x)| \leq 1 + \nu(x)$ pointwise (the value of c_A may need to be reduced slightly).*
- (ii) *If the Gowers inverse conjecture, Theorem 9.3 (ii), holds for a given value of k then it continues to hold for a function f such that $|f(x)| \leq 1 + \nu(x)$ pointwise. In particular such an extension of the Gowers inverse conjecture is true when $k = 2, 3$.*

The reader may consult [12, Ch. 3] for a definition of the term *pseudorandom measure* and a discussion concerning it. For the purposes of this article the reader can merely accept that there is such a notion, and furthermore that one may construct a pseudorandom measure $\nu: [N] \rightarrow \mathbb{R}^+$ such that $\nu + 1$ dominates any fixed W -tricked von Mangoldt function $\Lambda_{W,b}$. The construction of ν comes from sieve theoretic ideas originating with Selberg. The confirmation that ν is pseudorandom is essentially due, in a very different context, to Goldston and Yıldırım [7].

Applying these two results, one may see that Conjecture 2.1 for a given non-degenerate $s \times t$ matrix A is a consequence of the Gowers inverse conjecture in the case $k = s + 1$ together with a bound of the form

$$\mathbb{E}_{n \leq N} (\Lambda_{b,W}(n) - 1) F(T_g^n \cdot x) = o_{G/\Gamma, F}(1) \tag{24}$$

for every s -step nilsequence $(F(T_g^n \cdot x))_{n \in \mathbb{N}}$.

By effecting a decomposition of $\Lambda_{b,W}$ as $\Lambda_{b,W}^\sharp + \Lambda_{b,W}^\flat$ rather like that in §4, the proof of this statement may be further reduced to a similar result for the Möbius function:

Conjecture 10.2 (Möbius and nilsequences). For all $A > 0$ we have the bound

$$\mathbb{E}_{n \leq N} \mu(n) F(T_g^n \cdot x) \ll_{A,G/\Gamma,F} \log^{-A} N$$

for every k -step nilsequence $(F(T_g^n \cdot x))_{n \in \mathbb{N}}$.

Note that we require more cancellation (a power of a logarithm) here than in (24). This is because in passing from μ to $\Lambda_{b,W}^\flat$ one loses a logarithm in performing partial summation as in the derivation of (7). The method we have in mind to prove Conjecture 10.2, however, is likely to give this strong cancellation at no extra cost.

Conjecture 10.2 posits a rather vast generalisation of Davenport’s bound. The conjecture is, of course, highly plausible in view of the Möbius randomness law.

Let us remark that the derivation of (24) from Conjecture 10.2 is not at all immediate, since one must also handle the contribution from $\Lambda_{b,W}^\sharp$. To do this one uses methods of classical analytic number theory rather similar to those of Goldston and Yıldırım [7].

11. Möbius and nilsequences

The main result of [14] is a proof of Conjecture 10.2 in the case $k = 2$. This leads, by the reasoning outlined in the previous section, to a proof of Conjecture 2.4 in the case $s = 2$.

We remarked that the classical Hardy–Littlewood method was a technique of harmonic analysis. We also highlighted the idea of dividing into major and minor arcs. We have said much on the subject of generalising the underlying harmonic analysis, but as yet there has been nothing said about a suitable extension of major and minor arcs. In this section we describe such an extension by making some remarks concerning the proof of the case $k = 2$ of Conjecture 10.2.

In §5 we discussed how bounds on Type I and II sums may be used to show that a given function f does not correlate with Möbius. Recalling our “inverse” strategy for proving Davenport’s bound, one might be tempted to go straight into Proposition 5.1 with $f(n) = F(T_g^n \cdot x)$, a 2-step nilsequence, posit largeness of either a Type I or a Type II sum, and then use this to say that the nilsequence is somehow “major arc”. One might then hope to handle the major nilsequences by some other method, perhaps the theory of L -functions.

Such an attempt is a little too simplistic, for the following reason. Returning to the 1-step case, note that the sum of two 1-step nilsequences is also a 1-step nilsequence (on the product nilmanifold $G_1/\Gamma_1 \times G_2/\Gamma_2$). In particular, the function

$f(n) = e(n/5) + e(n\sqrt{2})$ is a 1-step nilsequence. We know, however, that to handle correlation of Möbius with $e(n/5)$ we need to know something about L -functions, whereas we do not have an L -function method of handling $e(n\sqrt{2})$. This suggests that some sort of preliminary decomposition of the function f is in order, and such a suggestion turns out to be correct.

In the 2-step case, a nilsequence $F(T_g^n \cdot x)$ can be decomposed into *local quadratics*. These are objects of the form

$$f(n) := 1_{B_N}(n)e(\phi(n)), \tag{25}$$

where B_N is a set of the form

$$B_N := \{n : N/2 \leq n < N : F_1(n) \neq 0\}$$

for some 1-step nilsequence F_1 depending on $F, G/\Gamma, g$ and x , and $\phi : B_N \rightarrow \mathbb{R}/\mathbb{Z}$ is *locally quadratic*. This means that one may unambiguously define the second derivative $\phi''(h_1, h_2)$ to equal

$$\phi(x + h_1 + h_2) - \phi(x + h_1) - \phi(x + h_2) + \phi(x)$$

for any x such that $x, x + h_1, x + h_2, x + h_1 + h_2 \in B_N$.

It turns out that for the purposes of analysing Type I and II sums the cutoff 1_{B_N} plays a subservient rôle. The phase ϕ , on the other hand, is crucial. The bulk of [14] is devoted to showing that if either a Type I or a Type II sum involving some f as in (25) is large, then ϕ is *major arc*. This is a direct analogue of the proof of Davenport’s bound as phrased at the end of §5 (the “inverse” approach). Roughly speaking, ϕ is said to be major arc if $q\phi''(h_1, h_2)$ is small for some smallish q and all h_1, h_2 , which in turn essentially means that ϕ is slowly varying on B_N intersected with any fixed progression $a \pmod{q}$. For a detailed discussion see [14]. Suffice it to say that the passage from large Type I/II sum to ϕ being major arc is long and difficult, and requires many applications of the Cauchy–Schwarz inequality to manipulate the phase ϕ into a helpful form, as well as basic tools of equidistribution such as a version of the Erdős–Turán inequality.

Recalling Proposition 5.1, one has reduced the case $s = 2$ of Conjecture 10.2 to the statement that

$$\mathbb{E}_{n \leq N} \mu(n) 1_{B_N}(n) e(\phi(n)) \ll_A \log^{-A} N$$

for any major arc phase ϕ . It turns out that $1_{B_N}(n)e(\phi(n))$ can, in this case, be closely approximated by a sum of linear phases $e(\theta n)$, and so we may conclude using Proposition 4.1.

Note that this analysis has the flavour of an induction on s , the step of the nilsequence we are considering. We expect to see this more clearly when addressing the general case of Conjecture 10.2 in future work.

12. Future directions

The most obvious avenue of research left open is to generalise everything we have done for $s = 2$ to the case $s \geq 3$. In particular we would like inverse theorems for the U^k -norms for $k \geq 4$, and a proof of Conjecture 10.2 for $s \geq 3$. We are currently working towards this goal. We expect that the methods of Gowers [10] can be adapted to achieve the inverse theorem, though this will not be straightforward. It is also very likely that the “inverse” approach to handling Type I and II sums can be adapted to the higher-step case of Conjecture 10.2, though again we do not expect this to be wholly straightforward.

It would be very desirable to have good bounds for error terms such as the $o(1)$ in Theorem 2.5. We are sure that our current estimate for the error in Theorem 2.5 is the worst that has ever featured in analytic number theory – the error term is a *completely ineffective* $o(1)$! Ultimately this is because to show that the error is less than δ one finds oneself needing to rule out a real zero of some $L(s, \chi)$, χ a primitive quadratic character to the modulus q , with $s > 1 - Cq^{-\varepsilon}$, where $\varepsilon = \varepsilon(\delta) \rightarrow 0$ as $\delta \rightarrow 0$. Siegel’s theorem states that for any $\varepsilon > 0$ there is such a C , but it is, of course, not possible to specify C effectively.

It is clear that the spectre of ineffectivity does not rear its head under the assumption of GRH, and we believe that our methods lead to an error term of the form $\log^{-c} N$ in Theorem 2.5.

There are other, presumably more tractable, ways in which one might obtain an explicit error term. Improvements to the combinatorial tools used in [13], particularly advances on the circle of conjectures known as the “polynomial Freiman–Ruzsa conjecture”, could be very helpful here.

We turn now to goals which lie further away. I have hinted at various places in this survey that the way in which we see nilsequences arising is very long-winded and, presumably, not the “right” way. The ergodic theorists [20], [35] do admittedly discover the rôle of these functions somewhat less painfully (albeit after setting up a good deal of notation). Nilsequences seem such natural objects, however, that there ought to be a much better way of appreciating their place in the study of systems of linear equations. Recalling that $\|f\|_{U^2}$ is essentially the L^4 norm of \hat{f} one might even ask, for example,

Question 12.1. Is there a usable “formula” relating $\|f\|_{U^3}$ and certain of the “nil-fourier coefficients” $\mathbb{E}_{n \leq N} f(n) F(T_g^n \cdot x)$?

Such a formula would assuredly have to be very exotic on account of the vast profusion of nilsequences which might enter into consideration. The nilsequences are not naturally parametrised by anything so simple as the circle S^1 , which gave its name to the classical circle method.

Let us conclude with some speculations on non-linear systems of equations, where our knowledge is at present essentially non-existent. We have seen in Conjecture 9.2 that the behaviour of an any system $A\mathbf{x} = \mathbf{b}$, where A is non-degenerate in the sense

of Definition 2.3, should be governed by a very “hard” or “algebraic” collection of *characteristic functions*, in this case the nilsequences.

On the other hand degenerate linear systems, such as $x_1 - x_2 = 1$, do not have this property. To see this, suppose that $N = 2m$ is even and let $A \subseteq [N]$ be a set formed by setting $A \cap \{2i, 2i + 1\} = \{2i\}$ or $\{2i + 1\}$, these choices being independent in i for $i = 0, \dots, m - 1$. Then $|A| = N/2$, and A is indistinguishable from a truly random set by taking inner products with any conceivable “hard” character such as a linear or quadratic phase. However, A is expected to have about $N/8$ solutions to $x_1 - x_2 = 1$, whereas a random set has about twice this many.

One might call an equation or system of equations for which a “hard” characteristic system exists a *mixing* system. We do not have a precise definition of this notion. Some non-linear equations are known to be mixing – for example, the linear phases $e(\theta n)$ form a characteristic system for the equation $x_1 + x_2 = x_3^2$. Many more are not. It would be very interesting to know, for example, whether the equation $x_1 x_2 - x_3 x_4 = 1$ is mixing and, if so, what a characteristic system for it might be. This seems to be a very difficult question as the analysis of this equation even in very specific situations involves deep methods from the theory of automorphic forms.

References

- [1] Baker, R. C. and Harman, G., Exponential sums formed with the Möbius function. *J. London Math. Soc.* (2) **43** (2) (1991), 193–198.
- [2] Balog, A., Linear equations in primes. *Mathematika* **39** (1992), 367–378.
- [3] Bourgain, J., On triples in arithmetic progression. *Geom. Funct. Anal.* **9** (1999), 968–984.
- [4] Chowla, S., There exists an infinity of 3-combinations of primes in A.P. *Proc. Lahore. Philos. Soc.* **6** (2) (1944), 15–16.
- [5] Davenport, H., *Multiplicative number theory*. Third edition, Grad. Texts in Math. 74, Springer-Verlag, New York 2000.
- [6] Furstenberg, H. and Weiss, B., A mean ergodic theorem for $1/N \sum_{n=1}^N f(T^n x)g(T^{n^2} x)$. In *Convergence in ergodic theory and probability* (Columbus, OH, 1993), Ohio State Univ. Math. Res. Inst. Publ. 5, Walter de Gruyter, Berlin 1996, 193–227.
- [7] Goldston, D. A. and Yıldırım, C. Y., Small gaps between primes, I. Preprint.
- [8] Gowers, W. T., A new proof of Szemerédi’s theorem for arithmetic progressions of length four. *Geom. Funct. Anal.* **8** (1998), 529–551.
- [9] Gowers, W. T., Fourier analysis and Szemerédi’s theorem. In *Proceedings of the International Congress of Mathematicians* (Berlin, 1998), Vol. I, Doc. Math., J. DMV, Extra Vol. ICM Berlin, 1998, 617–629.
- [10] Gowers, W. T., A new proof of Szemerédi’s theorem. *Geom. Funct. Anal.* **11** (2001), 465–588.
- [11] Green, B. J., Long arithmetic progressions of primes. Preprint, submitted to *Proceedings of the Gauss-Dirichlet Conference*, Göttingen 2005.

- [12] Green, B. J. and Tao, T. C., The primes contain arbitrarily long arithmetic progressions. *Ann. of Math.*, to appear.
- [13] Green, B. J. and Tao, T. C., An inverse theorem for the Gowers U^3 -norm, with applications. Submitted.
- [14] Green, B. J. and Tao, T. C., Quadratic uniformity of the Möbius function. Preprint.
- [15] Green, B. J. and Tao, T. C., Linear equations in primes. In preparation.
- [16] Hardy, G. H. and Littlewood, J. E., Some problems of “Partitio Numerorum”. III. On the expression of a number as a sum of primes. *Acta. Math.* **44** (1923), 1–70.
- [17] Hardy, G. H. and Littlewood, J. E., Some problems of “Partitio Numerorum”. V. A further contribution to the study of Goldbach’s problem. *Proc. London Math. Soc. (2)* **22** (1923), 46–56.
- [18] Hardy, G. H. and Ramanujan, S., Asymptotic formulæ in combinatory analysis. *Proc. London Math. Soc. (2)* **17** (1918), 75–115.
- [19] Heath-Brown, D. R. Three primes and an almost prime in arithmetic progression. *J. London Math. Soc. (2)* **23** (1981), 396–414.
- [20] Host, B. and Kra, B. Non-conventional ergodic averages and nilmanifolds. *Ann. of Math.* **161** (1) (2005), 397–488.
- [21] Iwaniec, H. and Kowalski, E. *Analytic number theory*. Amer. Math. Soc. Colloq. Publ. 53, Amer. Math. Soc., Providence, RI, 2004.
- [22] Iwaniec, H., Luo, W and Sarnak, P., Low lying zeroes of families of L -functions. *Inst. Hautes Études Sci. Publ. Math.* **91** (2000), 55–131.
- [23] Kra, B., The Green-Tao Theorem on arithmetic progressions in the primes: an ergodic point of view. *Bull. Amer. Math. Soc.* **43** (2006), 3–23.
- [24] Kra, B., From combinatorics to ergodic theory and back again. In *Proceedings of the International Congress of Mathematicians* (Madrid, 2006), Volume III, EMS Publishing House, Zürich 2006, 57–76.
- [25] Kumchev, A. V. and Tolev, D. I., An invitation to additive prime number theory. *Serdica Math. J.* **31** (1–2) (2005), 1–74.
- [26] Salem, R. and Zygmund, A., Some properties of trigonometric series whose terms have random signs. *Acta Math.* **91** (1954), 245–301.
- [27] Tao, T. C., Arithmetic progressions and the primes. *El Escorial Proceedings 2004*, to appear.
- [28] Tao, T. C., Obstructions to uniformity, and arithmetic patterns in the primes. Preprint.
- [29] Van der Corput, J. G., Über Summen von Primzahlen und Primzahlquadraten. *Math. Ann.* **116** (1939), 1–50.
- [30] Vaughan, R. C., Sommes trigonométriques sur les nombres premiers, *C. R. Acad. Sci. Paris Sér. A* **285** (16) (1977), A981–A983.
- [31] Vaughan, R. C., Hardy’s Legacy to Number Theory. *J. Austral. Math. Soc. Ser. A* **65** (1998), 238–266.
- [32] Vaughan, R. C. and Wooley, T. D., Waring’s problem: a survey. In *Number theory for the millennium, III* (Urbana, IL, 2000), A K Peters, Natick, MA, 2002, 301–340.
- [33] Vinogradov, I. M., Representation of an odd number as the sum of three primes. *Dokl. Akad. Nauk SSSR* **15** (1937), 291–294.

- [34] Wooley, T. D., Diophantine problems in many variables: the rôle of additive number theory. In *Topics in Number Theory* (ed. by S. D. Ahlgren et al.), Kluwer Academic Publishers, Dordrecht 1999, 49–83.
- [35] Ziegler, T., Universal characteristic factors and Furstenberg averages. *J. Amer. Math. Soc.*, to appear.

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