

# APPROXIMATE ALGEBRAIC STRUCTURE

BEN GREEN

ABSTRACT. We discuss a selection of recent developments in arithmetic combinatorics having to do with “approximate algebraic structure” together with some of their applications.

## 1 Introduction

Given an inequality, an extremely natural question to ask is

When does equality occur?

If a satisfactory answer to this is available, one might then ask

When does equality *almost* occur?

To be a little more precise, suppose that we have some family of functions  $\mathcal{F}$  and some map (functional)  $v : \mathcal{F} \rightarrow \mathbb{R}$ . The inequality we are considering might then be of the form

$$v(f) \leq M \quad \text{for all } f \in \mathcal{F}.$$

To give an example, the well-known isoperimetric inequality on  $\mathbb{R}^n$  may be stated in this form, with  $\mathcal{F}$  being the set of all functions  $1_A$  where  $A \subset \mathbb{R}^n$  is bounded and open (say),  $v(1_A)$  being the isoperimetric ratio  $\frac{|A|}{|\partial A|^{n/(n-1)}}$ , and  $M$  being the isoperimetric ratio of any Euclidean ball.

The first natural question, the *equality question*, is then

For which  $f \in \mathcal{F}$  do we have  $v(f) = M$ ?

In the case of the isoperimetric inequality, it is well-known (and invariably stated as part of the inequality) that  $f = 1_A$  with  $A$  being a Euclidean ball.

The second natural question is

For which  $f \in \mathcal{F}$  do we have  $v(f) \approx M$ ?

Of course, to make this precise we must specify what is meant by  $\approx$ . We further distinguish between what might be called the *stability question*, which asks

For which  $f \in \mathcal{F}$  do we have  $v(f) \geq (1 + o(1))M$ ?

and what I shall term the *robustness question*, which asks

For which  $f \in \mathcal{F}$  do we have  $v(f) \geq \frac{1}{100}M$  (say)?

Most of this article will be concerned with the robustness question for two particular inequalities, an instance of *Young’s inequality for convolutions* and an inequality concerning the *Gowers norms*. In both situations the equality cases are easily established and are highly algebraic in nature (essentially they characterise finite

---

The author is supported by ERC Starting Grant number 274938 *Approximate algebraic structure and applications*.

groups and polynomial phases respectively). In both cases study of the robustness question has proven to be surprisingly subtle and has led to diverse applications in areas as different as group theory, additive prime number theory and theoretical computer science.

The stability question for these same inequalities is much better understood, though it is still nontrivial and has many applications. For want of space, we will not say a great deal about it. As it turns out the stability question for the isoperimetric inequality and related inequalities such as the Brunn-Minkowski inequality is the subject of much current research, not entirely unrelated to the topics discussed in this article: see for example [26, 27].

## 2 Approximate groups

### 2.1 Young's inequality

Let  $G$  be a group with identity element  $\text{id}_G$ , and let  $\mathcal{F}$  be the collection of all finitely-supported functions  $f : G \rightarrow [0, \infty)$  with  $\sum_{x \in G} f(x) = 1$ ,  $f(x) = f(x^{-1})$  for all  $x$  and  $f(\text{id}_G) > 0$ . One may think of  $f$  as a probability measure on  $G$ , the measure of a set  $A \subset G$  being  $\sum_{x \in A} f(x)$ . A particular (rather simple) case of a well-known inequality of Young [101] for convolutions is the bound

$$v(f) \leq 1 \quad \text{for all } f \in \mathcal{F},$$

where

$$v(f) = \frac{\|f * f\|_2^2}{\|f\|_2^2} = \frac{\sum_{x \in G} (\sum_{y \in G} f(y)f(yx))^2}{\sum_{x \in G} f(x)^2}.$$

Let us give the proof, which follows in a couple of lines using the Cauchy-Schwarz inequality: for each  $x \in G$  we have

$$\sum_{y \in G} f(y)f(yx) \leq (\sum_{y \in G} f(y)^2)^{1/2} (\sum_{y \in G} f(yx)^2)^{1/2} = \sum_{y \in G} f(y)^2, \quad (2.1)$$

and thus

$$(\sum_{y \in G} f(y)f(yx))^2 \leq (\sum_{y \in G} f(y)f(yx)) \sum_{y \in G} f(y)^2. \quad (2.2)$$

Summing over  $x \in G$  and using the fact that  $\sum_{t \in G} f(t) = 1$ , we obtain the result.

Let us address the equality question, that is to say let us characterise those  $f \in \mathcal{F}$  for which  $v(f) = 1$ . For this to happen, we must have equality in (2.2) for every  $x$ . For a given  $x$  this means that either  $\sum_y f(y)f(yx) = 0$ , or else equality occurs in (2.1). The first case implies that for all  $y$  at least one of  $f(y)$  and  $f(yx)$  is zero. The second case may be analysed using the well-known criterion for equality in the Cauchy-Schwarz inequality. This implies that there is some  $\lambda(x)$  such that  $f(y) = \lambda(x)f(yx)$  for all  $y$ ; using the fact that  $\sum_{t \in G} f(t) = 1$ , it follows that  $\lambda(x) = 1$  and therefore  $f(y) = f(yx)$  for all  $y$ .

Thus  $v(f) = 1$  if and only if for all  $x \in G$  we have one of the following two mutually exclusive options:

- (1) For all  $y \in G$ , either  $f(y)$  or  $f(yx)$  is zero;
- (2) For all  $y \in G$ ,  $f(y) = f(yx)$ .

It follows immediately from this that  $f$  cannot take more than one non-zero value, and therefore  $f(x) = \frac{1}{|A|}1_A(x)$  for some (finite) symmetric set  $A \subset G$  containing

the identity. The above two properties then tell us that for all  $x \in G$  we have one of the following two mutually exclusive options:

- (1)  $A$  and  $Ax$  are disjoint;
- (2)  $A = Ax$ .

The set of  $x$  for which (2) is satisfied is a subgroup of  $G$  (the stabiliser of  $A$  when  $G$  acts on finite subsets of itself by right multiplication). Call this group  $H$ . If  $x \in A$  then, since  $\text{id}_G \in A$ , the sets  $A$  and  $Ax$  are not disjoint and so  $x \in H$ ; thus  $A \subset H$ . On the other hand if  $x \in H$  then  $A = Ax$  and so in particular, since  $\text{id}_G \in A$ , we have  $x \in A$  and so  $H \subset A$ . It follows that  $A = H$  is a subgroup of  $G$ . Observations equivalent to these may be found in Hardy-Littlewood [58].

Now let us think about the stability and robustness questions. To do this, let us introduce a parameter  $K \geq 1$ , and let us ask what may be said about those  $f \in \mathcal{F}$  for which  $v(f) \geq \frac{1}{K}$ . This includes both the stability question (where  $K \approx 1$ ) and the robustness question (where  $K$  is somewhat larger, for example  $K \sim 100$ ). To spell it out, we are asking for a description of the finitely-supported, symmetric probability measures  $f : G \rightarrow [0, \infty)$  for which

$$\|f * f\|_2^2 \geq \frac{1}{K} \|f\|_2^2. \quad (2.3)$$

To get a feel for this question, let us specialise to the case  $f(x) = \frac{1}{|A|} 1_A(x)$ , for some finite, symmetric set  $A \subset G$  containing the identity. We saw above that only this case is relevant for discussion of the equality question, and in fact the analysis of the stability and robustness questions may be reduced to this case by fairly routine technical arguments [7], [15, Appendix A]. In this case one may check that  $\|f\|_2^2 = |A|^{-1}$  and

$$\|f * f\|_2^2 = |A|^{-4} \#\{(a_1, a_2, a_3, a_4) \in A \times A \times A \times A : a_1 a_2 = a_3 a_4\}.$$

Thus (2.3) holds if and only if we have

$$|A|^{-3} \#\{(a_1, a_2, a_3, a_4) \in A \times A \times A \times A : a_1 a_2 = a_3 a_4\} \geq \frac{1}{K}. \quad (2.4)$$

The quantity on the left here is usually called the *multiplicative energy*  $E(A)$  of the set  $A$ . As can be seen, it records coincidences amongst products of elements of  $A$ . Young's inequality implies that  $E(A) \leq 1$ , and we showed above that equality occurs if and only if  $A$  is a subgroup. That  $E(A) \leq 1$  can in fact be established easily and directly by noting that if  $a_1 a_2 = a_3 a_4$  then  $a_4$  is uniquely determined by  $a_1, a_2$  and  $a_3$ .

When, then, does (2.4) hold? Here we split the discussion of the stability question ( $K \approx 1$ ) and the robustness question ( $K \gg 1$ ), making just a few remarks about the former. In the stability case it turns out that  $A$  must be “almost” a subgroup; in fact there is a subgroup  $H$  such that the symmetric difference of  $A$  and  $H$  is very small. Results of this type are certainly very interesting and may be dated to work of Freiman [29] and Fournier [28] amongst others. Among the diverse applications are the analysis of certain algorithms for sampling at random from finite groups [19, 22, 42] and the solution of the Dirac-Motzkin conjecture in combinatorial geometry connected with point-line configurations having few ordinary lines [51].

Our main focus here, however, is on the robustness regime  $K \gg 1$ , where the flavour and the applications are somewhat different. We begin by observing that

(2.4) is implied by a condition which is perhaps easier to understand, that of *small doubling*. We say that a set  $A \subset G$  has doubling at most  $K$  if

$$|A^2| \leq K|A|, \quad (2.5)$$

where  $A^2 = \{a_1 a_2 : a_1, a_2 \in A\}$ . To see that (2.5) implies (2.4), write  $r(x)$  for the number of representations of pairs  $(a_1, a_2) \in A \times A$  with  $a_1 a_2 = x$ . Then  $r(x) = 0$  for  $x \notin A^2$  and so by the Cauchy-Schwarz inequality we have

$$E(A) = \sum_x r(x)^2 \geq \frac{1}{|A^2|} \left( \sum_x r(x) \right)^2 = \frac{|A|^4}{|A^2|} \geq \frac{1}{K} |A|^3.$$

We have shown that (2.5) implies (2.4), and so if  $f(x) = \frac{1}{|A|} 1_A(x)$  for a symmetric set  $A$  satisfying (2.5) then indeed  $v(f) \geq \frac{1}{K}$ . Thus an analysis of the robustness question for Young's inequality necessarily involves studying sets  $A$  satisfying (2.5). It is not at all obvious that such a study is sufficient for that task, because we have not shown that (2.4) implies (2.5). In fact, it does not, as be easily seen by taking  $A$  to be  $H \cup X$ , where  $H$  is a subgroup of  $G$  and  $X$  is an arbitrary symmetric set of the same size, disjoint from  $H$ . Then (2.4) holds with  $K = 8$ , since we may take all quadruples  $(a_1, a_2, a_3, a_4)$  with  $a_1, \dots, a_4 \in H$  and  $a_1 a_2 = a_3 a_4$ . However, there is absolutely no reason to suppose that (2.5) holds, and indeed  $A^2$  contains  $X^2$  which could have size as large as  $c|X|^2$ . We leave it to the reader to provide an explicit example in a suitable group  $G$ . Remarkably, however, the large multiplicative energy condition (2.4) does imply a weak version of (2.5): specifically, (2.5) is true after passing from  $A$  to a large subset  $A'$  and replacing (2.5) by a somewhat weaker condition  $|A'^2| \leq K'|A'|$  with  $K' \sim K^{10}$ , say. This result is known as the Balog-Szemerédi-Gowers theorem, because in the case  $G$  abelian it was established by Gowers [35] in the course of his seminal work on Szemerédi's theorem, an earlier result of a qualitatively similar form but with the bound on  $K'$  being vastly weaker having previously been established by Balog and Szemerédi [2] by different means. It was shown by Tao [94] that the assumption that  $G$  is abelian could be dropped.

## 2.2 Approximate groups

We have discussed the relationship between the robustness question for Young's inequality and the study of finite sets  $A$  satisfying the small doubling condition  $|A^2| \leq K|A|$ . Since subgroups of  $G$  provide equality in Young's inequality, this provides some justification for thinking of such  $A$  as "approximate groups". Moreover, the small doubling condition visibly suggests that  $A$  is somehow almost closed under multiplication, surely a property we would expect from any sensible notion of an approximate group. As it turns out, it has been found convenient to introduce a slightly different but closely related notion.

**Definition 2.1** (Approximate group). Let  $A$  be a subset of a group  $G$ . Then we say that  $A$  is a  $K$ -approximate group if  $A$  is symmetric, contains the identity, and if  $A^2 \subset XA$  for some set  $X$  of size at most  $K$ .

This definition was introduced by Tao [94] and has certain advantages such as behaving well under homomorphisms, making sense for infinite sets  $A$  as well as finite ones, and immediately implying further conditions on  $A$  such the tripling bound  $|A^3| \leq K^2|A|$ .

Note that if  $A$  is a  $K$ -approximate group then  $A$  automatically satisfies the small doubling condition (2.5), and hence the large multiplicative energy condition (2.4). The reverse direction is less clear, and the situation is much the same as before: a set satisfying (2.5) need not be a  $K$ -approximate group, but there is a closely related set  $A'$  which is a  $K'$ -approximate group for some  $K' \sim K^{10}$ . This deduction is essentially due to Ruzsa, who laid the foundation for the whole theory in a series of works. For a precise statement and further references, §4 of [14] may be consulted.

### 2.3 Examples

We now give some examples of approximate groups. The first example is fairly trivial.

**Example 1.** If  $A \subset G$  is a subgroup then of course  $A$  is symmetric,  $\text{id}_G \in A$  and  $A^2 = A$ . Thus  $A$  is a 1-approximate group.

Thus far, we have not pointed out that there are in fact nontrivial examples of approximate groups. The simplest is a geometric progression.

**Example 2.** If  $P$  is the geometric progression

$$P = P(u; N) := \{u^n : 0 \leq n < N\}$$

for some element  $u \in G$  and if  $A = P \cup P^{-1}$  then  $A$  is a 2-approximate group. Indeed  $A = \{u^n : -N + 1 \leq n \leq N - 1\}$ ,  $A^2 = \{u^n : -2N + 2 \leq n \leq 2N - 2\}$  and so  $A^2 \subset XA$  where  $X = \{u^{N-1}, u^{-N+1}\}$ .

Less obviously, there are multidimensional generalisations of the preceding example.

**Example 3.** If  $P$  is the multidimensional geometric progression

$$P = P(u_1, \dots, u_d; N_1, \dots, N_d) := \{u_1^{n_1} u_2^{n_2} \dots u_d^{n_d} : 0 \leq n_i < N_i\}$$

for some *commuting* elements  $u_1, \dots, u_d \in G$  and integers  $N_1, \dots, N_d > 0$  and if  $A = P \cup P^{-1}$  then  $A$  is a  $2^d$ -approximate group. We leave the confirmation of this to the reader.

The commuting assumption was very important in the previous example (otherwise we cannot simplify a product  $u_1^{n_1} \dots u_d^{n_d} u_1^{n'_1} \dots u_d^{n'_d}$  to  $u_1^{n_1+n'_1} \dots u_d^{n_d+n'_d}$ ). However, it can be replaced by the weaker condition of nilpotence, as the following example shows.

**Example 4.** Let  $N_1, N_2, N_{1,2}$  be positive integers with  $N_{1,2} \geq N_1 N_2$ , let  $G = \begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}$  be the Heisenberg group and let  $A \subset G$  be the following set of matrices. Let

$$u_1 := \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad u_2 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix},$$

and take  $A = P \cup P^{-1}$  where  $P = P(u_1, u_2, [u_1, u_2]; N_1, N_2, N_{1,2})$  is the set

$$\{u_1^{n_1} u_2^{n_2} [u_1, u_2]^{n_{1,2}} : 0 \leq n_1 < N_1, 0 \leq n_2 < N_2, 0 \leq n_{1,2} < N_{1,2}\}.$$

Here,  $[u_1, u_2]$  is the commutator given by

$$[u_1, u_2] := u_1 u_2 u_1^{-1} u_2^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

It may be straightforwardly checked that

$$u_1^{n_1} u_2^{n_2} [u_1, u_2]^{n_{1,2}} \cdot u_1^{n'_1} u_2^{n'_2} [u_1, u_2]^{n'_{1,2}} = u_1^{n_1+n'_1} u_2^{n_2+n'_2} [u_1, u_2]^{n_{1,2}+n'_{1,2}-n'_1 n'_2}. \quad (2.6)$$

Hence

$$P^{-1} \subset \{u_1^{n_1} u_2^{n_2} [u_1, u_2]^{n_{1,2}} : -N_1 < n_1 \leq 0, -N_2 < n_2 \leq 0, -2N_{1,2} < n_{1,2} \leq 0\}$$

and

$$A^2 \subset \{u_1^{n_1} u_2^{n_2} [u_1, u_2]^{n_{1,2}} : |n_1| < 2N_1, |n_2| < 2N_2, |n_{1,2}| < 5N_{1,2}\}.$$

Now for any  $n'_1, n'_2, n'_{1,2}$  in (2.6) we may choose (unique) integers  $k_1, k_2, k_{1,2}$  such that

$$u_1^{k_1 N_1} u_2^{k_2 N_2} [u_1, u_2]^{k_{1,2} N_{1,2}} \cdot u_1^{n'_1} u_2^{n'_2} [u_1, u_2]^{n'_{1,2}} \in P.$$

Indeed we have  $k_1 = -\lfloor n'_1/N_1 \rfloor$ ,  $k_2 = -\lfloor n'_2/N_2 \rfloor$  and  $k_{1,2} = -\lfloor (n'_{1,2} - n'_1 k_2 N_2)/N'_{1,2} \rfloor$ .

Thus if  $u_1^{n'_1} u_2^{n'_2} [u_1, u_2]^{n'_{1,2}} \in A^2$  then  $|k_1| \leq 1$ ,  $|k_2| \leq 1$  and  $|k_{1,2}| \leq 6$ . Hence

$$A^2 \subset XP \subset XA,$$

where

$$X = \{u_1^{k_1 N_1} u_2^{k_2 N_2} [u_1, u_2]^{k_{1,2} N_{1,2}} : |k_1| \leq 1, |k_2| \leq 1, |k_{1,2}| \leq 6\}$$

is a set of size 117. That is,  $A$  is a 117-approximate group. (A smaller constant could be obtained with a more careful analysis.)

Example 4 is an example of a *nilprogression*. The key feature of the Heisenberg group  $G$  relevant to this example is the fact that it is nilpotent of class 2, which means that commutators of order 3 or higher are all equal to the identity, or equivalently that  $[u_1, u_2]$  commutes with everything else. Similar examples can be constructed in more general nilpotent groups of arbitrary class  $s$ , though the constant  $K$  (117 in Example 4) will generally grow with  $s$ . We will not give the details here, and refer the reader instead to [14, Definition 2.1]. The nilprogression in Example 4 is said to have rank 2 and class 2 (the rank being the number of generators  $u_i$  and the class being the nilpotency class of the group generated by  $u_1, u_2$ ).

Different instances of the above constructions may be combined to create new examples. For example, it is easy to see that the direct product of a  $K_1$ -approximate group and a  $K_2$ -approximate group is a  $(K_1 K_2)$ -approximate group. There are also combinations of the above examples which are not direct products, for example the following example of Helfgott [60].

**Example 5.** Let  $p$  be a large prime, let  $r, s, t \in \mathbb{F}_p$  be fixed generators of  $\mathbb{F}_p^*$ , let  $N_1, N_2, N_3$  be positive integers, and define  $A$  to be a set of  $3 \times 3$  matrices over  $\mathbb{F}_p$  as follows. Set  $A = HP \cup (HP)^{-1}$ , where

$$H := \left\{ \begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix} : x, y, z \in \mathbb{F}_p \right\}$$

and

$$P = P(u_1, u_2, u_3; N_1, N_2, N_3) := \{u_1^{n_1} u_2^{n_2} u_3^{n_3} : 0 \leq n_i < N_i\}$$

with

$$u_1 := \begin{pmatrix} r & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, u_2 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & s & 0 \\ 0 & 0 & 1 \end{pmatrix}, u_3 := \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & t \end{pmatrix}$$

for some  $r, s, t \in \mathbb{F}_p^*$  as in Example 3 above. It is quite easy to check that

$$A = \left\{ \begin{pmatrix} r^{n_1} & x & z \\ 0 & s^{n_2} & y \\ 0 & 0 & t^{n_3} \end{pmatrix} : x, y, z \in \mathbb{F}_p, -N_i < n_i < N_i \right\}$$

and hence

$$A^2 \subset \left\{ \begin{pmatrix} r^{n'_1} & x & z \\ 0 & s^{n'_2} & y \\ 0 & 0 & t^{n'_3} \end{pmatrix} : x, y, z \in \mathbb{F}_p, -2N_i < n'_i < 2N_i \right\},$$

from which it follows that  $A$  is an 8-approximate group.

In Example 5,  $HP$  was an example of a *coset nilprogression*, in this case of rank 3 and step 1. The general form of a coset progression is  $HP$  where  $P$  is a nilprogression (a notion we did not define in full generality) and  $H$  is a subgroup normal in the group  $\langle P \rangle$  generated by  $P$ . In fact, all five of our examples were of the form  $A = (HP) \cup (HP)^{-1}$  for some coset progression  $HP$  (in Examples 2, 3 and 4 the subgroup  $H$  was trivial, whilst in Example 1 the nilprogression  $P$  was trivial). Conversely, every  $A$  of this form is a  $K$ -approximate group, where  $K$  is bounded as a function of the rank  $r$  and the class  $s$  of  $P$ . Once again we refer the reader to [14] for more information.

#### 2.4 Theorems about approximate groups

Given the discussion of the last section, it is natural to ask whether every  $K$ -approximate group is of the form  $(HP) \cup (HP)^{-1}$  for some coset nilprogression  $HP$  (of rank and step bounded in terms of  $K$ ). The answer to this is, strictly speaking, negative, as the following example of a set  $A \subset \mathbb{Z}$  shows. Here, we use additive notation for the group operation on  $\mathbb{Z}$  and so our interest is in  $2A = A + A$  rather than  $A^2$ . Define  $A$  to be  $\{0\} \cup \bigcup_{j=1}^N \{2j - \varepsilon_j, -2j + \varepsilon_j\}$ , where the  $\varepsilon_j$  are independent  $\{0, 1\}$ -valued random variables. Then  $2A \subset [-4N, 4N]$ . However,  $\{-1, 0, 1\} + A \supset [-2N, 2N]$ , and so  $\{-2N, 2N\} + \{-1, 0, 1\} + A \supset [-4N, 4N] \supset 2A$ . It follows that  $A$  is a 6-approximate group. However, for a typical choice of the  $\varepsilon_j$ ,  $A$  does not have nearly so much structure as a progression (though it is *syndetic*, that is to say has bounded gaps, which is what makes this construction work).

However we do have the following recent result of Breuillard, Tao and the author [13].

**Theorem 2.2.** *Let  $A$  be a  $K$ -approximate subgroup of a group  $G$ . Then there is a coset nilprogression  $B = HP$  of rank and class bounded as functions of  $K$ , where such that  $|B| \leq K'|A|$  and there is a set  $X \subset G$  with  $|X| \leq K'$  such that  $A \subset (XB) \cap (BX)$ . Here,  $K'$  may be bounded as a function of  $K$  only.*

We say that  $A$  is  $K'$ -controlled by  $B$ . In the example preceding the theorem, we may take  $B = \{0, \dots, N-1\}$ . The reader is encouraged not to dwell too lengthily on the notion of “control” and read the above theorem as follows: every approximate group is roughly a coset nilprogression.

- For many specific types of group  $G$ , statements equivalent to Theorem 2.2 had previously been established, often with good quantitative control over the parameter  $K'$  as well as the rank and class. When  $G = \mathbb{Z}$ , this is essentially the celebrated Freiman-Ruzsa theorem [30, 80]. The general abelian case was handled by Ruzsa and the author [44], building on earlier work of Ruzsa [81]. Various matrix groups  $G$  were handled in work of (in chronological order) Elekes-Király [25], Chang [18] and Helfgott [59, 60], the latter handling  $\mathrm{SL}_2(k)$  and  $\mathrm{SL}_3(k)$  with  $k = \mathbb{F}_p$  or  $k = \mathbb{C}$ , amongst others.

- Hrushovski [65], in a very important 2009 breakthrough, dealt with  $G = \mathrm{GL}_n(\mathbb{C})$  (though with some dependence on  $n$ ). His argument was model-theoretic and a key ingredient of it was his “Lie model theorem”, also a key ingredient in the proof of Theorem 2.2.
- The proof of Theorem 2.2 additionally requires arguments related to the solution of Hilbert’s Fifth Problem (every locally compact group is locally an inverse limit of Lie groups), specifically lemmas due to Gleason from the 1950s. It also makes use of a lemma in additive combinatorics of a type developed by Sanders [84] and Croot-Sisask [21].
- Theorem 2.2 is in fact valid when  $G$  is a “local group” rather than a *bona fide* group. Moreover, it was necessary in [13] to work in this larger category, although Hrushovski and van den Dries have since managed to arrange the argument so that, at the expense of proving a slightly weaker result, one need only work in genuine groups.
- Theorem 2.2 rather easily implies Gromov’s famous theorem [56] on groups of polynomial growth. However, it does not really provide a new proof of Gromov’s theorem as all the deep ingredients Gromov developed (the notion of an asymptotic cone, and the application of the solution to Hilbert’s Fifth Problem) are also required here in some form.

Whilst Theorem 2.2 is definitive from the qualitative point of view, for many applications more quantitative statements are required. Unfortunately, a crucial use of ultrafilters in the proof of Theorem 2.2 means that no quantitative dependence of  $K'$  on  $K$  is currently known<sup>1</sup>. In the next section we will discuss perhaps the most substantial application of the theory of approximate groups so far, to the study of rapidly mixing random walks on groups (expanders), which find further application in the “affine sieve”. For these applications much more quantitative statements are required, but in more restricted settings.

A celebrated result of the type we have in mind is the theorem of Helfgott [59].

**Theorem 2.3** (Helfgott). *Let  $K \geq 2$ . Suppose that  $A \subset G$  is a  $K$ -approximate group, where  $G = \mathrm{PSL}_2(\mathbb{F}_p)$ , and that  $A$  generates  $G$ . Then either  $|A| \leq K^{C_2}$  or  $|A| \geq K^{-C_2}|G|$ , where  $C_2$  is an absolute constant.*

This theorem was generalised to  $\mathrm{PSL}_3(\mathbb{F}_p)$  by Helfgott in a subsequent paper [60], and then to  $\mathrm{PSL}_n(\mathbb{F}_p)$  (with  $C_2$  replaced by an exponent  $C_n$  depending on  $n$ ) and other finite simple groups of Lie type in independent works of Pyber-Szabó [79] and Breuillard, Tao and the author [11], the former paper containing a slightly more general result than the latter.

It is worth remarking that Helfgott’s arguments made substantial use of the theory of *approximate fields* or “sum-product theory”, in particular a result of Bourgain, Katz and Tao [10]. This is an important topic in arithmetic combinatorics and it has links to the theory of approximate groups as well as other substantial applications, perhaps most notably estimates for the additive Fourier transform of multiplicative subgroups of  $\mathbb{F}_p^\times$  due to Bourgain, Glibichuk and Konyagin [9]. The subsequent works [11, 79] do not make explicit use of this theory, and in fact it was noted in [11] that, conversely, results about approximate subgroups of  $\mathrm{SL}_2(k)$

---

<sup>1</sup>In principle one could be obtained by quantifier elimination but this would be a huge amount of effort and the bound would be desperately weak.



imply results about approximate subfields of  $k$ . Sadly we do not have the space to discuss these aspects any further here.

It is also of interest to note that [11] made use of an analogue for approximate groups of an argument of Larson and Pink [72], which gives a self-contained and relatively concise proof of certain statements which follow from the Classification of Finite Simple Groups (CFSG).

### 2.5 Applications

Several applications of Theorem 2.2 are given in the paper [13]. In addition to certain refinements of Gromov's theorem they include a result about the virtual nilpotence of the fundamental group of almost negatively-curved Riemannian manifolds, and a generalisation of a lemma of Margulis stating that the “almost stabiliser” of a point  $x$  in a finite-dimensional metric space  $X$  under the action of a discrete group of isometries is virtually nilpotent. Here, however, we wish to discuss an appealing application, due to Bourgain and Gamburd [7] of Helfgott's result, Theorem 2.3.

The result concerns a property of a generating set  $S$  of a finite group  $G$  known as *expansion*. This property has several equivalent characterisations, details of which may be found in the survey [62]. For our purposes here, however, it is convenient to define expansion in terms of the rapid mixing of the random walk on generators  $S \cup S^{-1}$ . For the sake of illustration suppose that  $|S| = 2$ , write  $S = \{a, b\}$ , and imagine  $G$  being quite large. Then we perform a random walk of  $m$  steps, the end result of which is a product  $x_m = g_1 \dots g_m$  where each  $g_i$  is selected independently at random from the set  $\{a, b, a^{-1}, b^{-1}\}$ . Note that if  $G = \mathbb{Z}^2$  and  $S = \{(1, 0), (0, 1)\}$  (and if additive notation is used) then this is precisely the classical random walk on the plane  $\mathbb{R}^2$ .

Now  $x_1$  takes values in a set of size 4, tiny in comparison to  $|G|$ , and by a trivial induction  $x_j$  takes values in a set of size at most  $4^j$  (in fact by an almost-as-trivial induction one may reduce this to  $4 \cdot 3^{j-1}$ ). Thus if  $j = c \log |G|$  for some small value of  $c$  then  $x_j$  takes values in a set of size at most  $|G|^{c'}$ , and in particular is nowhere near to equidistributed on  $G$ . However in certain situations it turns out to be the case that  $x_j$  is highly equidistributed not much later than this time, say for  $j \geq C \log |G|$ , for some  $C$ . By “highly-equidistributed” let us (slightly arbitrarily) say that we mean  $\mathbb{P}(x_j = g) = \frac{1}{|G|} + O(\frac{1}{|G|^{10}})$  for all  $g \in G$ . The situation just described is one possible definition of what it means for  $S$  to be an expander (the precise definition must include the parameter  $C$ ).

**Theorem 2.4** (Bourgain-Gamburd). *Let  $G = \mathrm{PSL}_2(\mathbb{F}_p)$  and suppose that  $S = \{(\begin{smallmatrix} 1 & 3 \\ 0 & 1 \end{smallmatrix}), (\begin{smallmatrix} 1 & 0 \\ 3 & 1 \end{smallmatrix})\}$ . Then the random walk on generating set  $S \cup S^{-1}$  becomes highly equidistributed in time at most  $C \log |G|$  for some absolute constant  $C$ , independent of  $p$ .*

The only reason we have put “3” in the matrices here is that the result was already known with “1” and “2” by different methods. Bourgain and Gamburd actually proved a far more general result, but we fix on this special case for the sake of illustration. To describe the proof, we note that the result can be reformulated in terms of convolution powers of  $\mu = \frac{1}{4}(\delta_a + \delta_b + \delta_{a^{-1}} + \delta_{b^{-1}})$ , that is to say the function  $\mu : G \rightarrow [0, \infty)$  taking values  $\frac{1}{4}$  at  $x = a, b, a^{-1}, b^{-1}$  and zero elsewhere. The probability that  $x_j = g$  is then precisely  $\mu^{(j)}(g)$ , where  $\mu^{(j)} = \mu * \dots * \mu$

and the convolution is repeated  $j$  times, and where we are defining  $\nu_1 * \nu_2(x) = \sum_y \nu_1(y) \nu_2(xy^{-1})$ . Note that  $\sum_x \mu^{(j)}(x) = 1$ .

We are interested in how quickly  $\mu^{(j)}$  tends towards the constant function  $\frac{1}{|G|}$ . To study this, we follow the progress of  $\mu^{(j)}$  in three stages:

- The early stage in which  $j \leq \frac{1}{10} \log |G|$ ;
- The middle stage in which  $\frac{1}{10} \log |G| \leq j \leq \frac{C}{10} \log |G|$ ;
- The end stage in which  $\frac{C}{10} \log |G| \leq j \leq C \log |G|$ .

The early stage is relatively easy to analyse. This is because the elements  $a, b$  behave as though they were generators of a free group, or in other words the random walk does not intersect itself nontrivially, so long as words of length at most  $\frac{1}{10} \log |G|$  are being considered. As a consequence, the size  $|\text{Supp}(\mu^{(j)})|$  of the support of  $\mu^{(j)}$  at the end of the early stage is somewhat large, of size at least about  $|G|^{0.01}$ , say. In fact the support is not quite the most sensible thing to look at, because  $\mu^{(j)}$  may take on several different values. A more nuanced quantity is  $\|\mu^{(j)}\|_2^{-2}$ , which we will call the *weighted support*. This would equal  $|\text{Supp}(\mu^{(j)})|$  if  $\mu^{(j)}$  did happen to be constant on its support, as is easily checked.

The theory of approximate groups is applied to analyse the middle stage. If  $j_1$  is the end of the early stage, we look at iterates  $\mu^{(j_1)}, \mu^{(2j_1)}, \dots, \mu^{(j_2)}$  where  $j_2 = 2^\ell j_1$  for some  $\ell$ . If  $\ell$  is somewhat large, the weighted support of  $\mu^{(2^t j_1)}$  cannot always increase substantially as we change  $t$  to  $t+1$ , and so there must be some  $t$  for which the weighted supports of  $\mu^{(2^t j_1)}$  and of  $\mu^{(2^{t+1} j_1)}$  are roughly the same. Since  $\mu^{(2^{t+1} j_1)} = \mu^{(2^t j_1)} * \mu^{(2^t j_1)}$ , the only way that this can happen is if  $f = \mu^{(2^t j_1)}$  satisfies (2.3) for some fairly small value of  $K$ , that is to say  $\|f * f\|_2^2 \geq \frac{1}{K} \|f\|_2^2$  or  $v(f) \geq \frac{1}{K}$ . This is precisely the robustness question for Young's inequality that we have been studying. As we discussed, this situation implies, very roughly speaking<sup>2</sup>, that  $f \sim \frac{1}{|A|} 1_A(x)$  where  $A$  is a  $K$ -approximate group. Here of course we are concerned with the particular case  $G = \text{PSL}_2(\mathbb{F}_p)$ , so by Helfgott's Theorem 2.3 there are three possibilities: (i)  $A$  is tiny, (ii)  $A$  is almost all of  $G$  and (iii)  $A$  does not generate  $G$ . Case (i) cannot occur, because at the end of the early stage the weighted support of  $\mu^{(j_1)}$  was quite large. It turns out that (iii) also cannot occur, because of the particular structure of proper subgroups of  $\text{PSL}_2(\mathbb{F}_p)$ : they are all soluble and so satisfy the law  $[[x_1, x_2], [x_3, x_4]] = \text{id}_G$ , quite at odds with the free behaviour exhibited during the early stage. We are left, then with possibility (ii), which implies that the weighted support of  $\mu^{(2^t j_1)}$  is almost  $|G|$ . By further applications of Young's inequality the same is true of  $\mu^{(2^\ell j_1)} = \mu^{(j_2)}$ . That is to say, at the end of the middle stage  $\mu^{(j_2)}$  fills out a large portion of  $G$  in a fairly uniform way.

The analysis of the end stage involves still different ideas – an application of representation theory having its origin in a paper of Sarnak and Xue [88]. The crucial input is the fact that all nontrivial representations of  $G = \text{PSL}_2(\mathbb{F}_p)$  have dimension at least  $\frac{1}{2}(p-1)$  and in particular at least  $|G|^c$  for some constant  $c$ . (In the language of Gowers [37],  $G$  is an example of a “quasirandom” group.) Further details may, of course, be found in the original paper [7].

---

<sup>2</sup>The  $\sim$  notation here hides quite a few technicalities.

The “Bourgain-Gamburd expansion machine” just described and modifications of it have found many further applications. One is the following variant of Theorem 2.4 due to Breuillard, Guralnick, Tao and the author [15].

**Theorem 2.5.** *Let  $G$  be any finite simple group of Lie type and suppose that  $S = \{a, b\}$  where  $a, b$  are chosen uniformly at random from  $G$ . Then, with probability at least  $1 - O(|G|^{-c})$ , the random walk on generating set  $S \cup S^{-1}$  becomes highly equidistributed in time at most  $C \log |G|$ . Here  $c, C > 0$  depend only on the rank of  $G$ .*

For example, this theorem holds with  $G = \mathrm{PSL}_n(\mathbb{F}_q)$  and with  $C$  depending only on  $n$  and not on  $q$ . The proof of this theorem relies on the Bourgain-Gamburd expansion machine but with the work of Pyber–Szabó [79] and Breuillard, Tao and the author [11] in place of Helfgott’s theorem. It also requires several other ingredients, including two different *ad hoc* analyses in two particular families of groups (the symplectic groups  $\mathrm{Sp}_4(k)$  in characteristic 3 and the triality Groups  ${}^3D_4(q)$ ). A different particular case, that in which  $G$  is a Suzuki group  $\mathrm{Sz}(q)$ , had been handled in an earlier paper [12] of the authors. This was of a certain amount of interest because it completed the proof of the following theorem of Lubotzky, Kassabov and Nikolov [67].

**Theorem 2.6.** *There are absolute constants  $k, C$  with the following property. For any nonabelian finite simple group  $G$ , there is a set  $S \subset G$  of size at most  $k$  such that the random walk on generating set  $S \cup S^{-1}$  becomes highly equidistributed in time at most  $C \log |G|$ .*

The proof of this theorem depends on CFSG and the most impressive ingredient is, in my view, Kassabov’s proof [66] in the case  $G = A_n$ . It appears to be unknown whether or not Theorem 2.5 holds uniformly for all finite simple groups  $G$  with  $C$  an absolute constant, even in (especially in?) the case  $G = A_n$ .

Perhaps of greater interest for applications than results such as Theorem 2.5, however, are generalisations of the original Bourgain-Gamburd theorem, where the groups under consideration range over a family such as  $G = \mathrm{PSL}_n(\mathbb{F}_p)$ ,  $p$  prime and the set  $S$  is obtained by reduction of a *fixed* set of integer matrices, rather than by random selection for each  $p$ . In the Bourgain-Gamburd theorem as stated above,  $S = \{(\begin{smallmatrix} 1 & 3 \\ 0 & 1 \end{smallmatrix}), (\begin{smallmatrix} 1 & 0 \\ 3 & 1 \end{smallmatrix})\}$ . The crucial property of this set of generators for rapid mixing of the random walk is that, considered as a subset of  $\mathrm{SL}_2(\mathbb{Z})$ , the subgroup they generate is *Zariski dense* (not contained in any proper algebraic subvariety). That this condition is sufficient was established by Bourgain-Gamburd for the family  $\mathrm{PSL}_2(\mathbb{F}_p)$ ,  $p$  prime. Varjú [97] obtained the same result for  $\mathrm{PSL}_n(\mathbb{F}_p)$ , and moreover for  $\mathrm{PSL}_n(\mathbb{Z}/q\mathbb{Z})$  where  $q$  is squarefree but may well be composite. (Such results had already been established in the case  $n = 2$  by Bourgain, Gamburd and Sarnak [8] by a more complicated method based in part on Helfgott’s arguments, necessitating in particular a foray into the tricky territory of *approximate subrings* of  $\mathbb{Z}/q\mathbb{Z}$ .) This last result is a crucial ingredient in the so-called *affine sieve* of Bourgain, Gamburd and Sarnak which finds almost primes in the matrix entries of orbits in matrix groups. Any serious discussion of this would take us too far afield, so we refer the reader to [82] for the state of the art and to the very nice exposition [87] for a (somewhat outdated) introduction. See also [41], again rather outdated.

## 2.6 Open questions

There are many open questions concerning the quantitative aspects of the theory described above. For example, no version of Theorem 2.2 in which the parameter  $K'$  is given quantitatively in terms of  $K$  is known, and nor does it seem prudent at this stage to speculate on what might be true in this regard. One tempting line of enquiry would be to look at Kleiner's alternative proof [68] of Gromov's theorem in the context of approximate groups, but this has not so far been successful.

Even in the case  $G = \mathbb{Z}$  there are unsolved problems connected with approximate groups. As previously noted, Theorem 2.2 in this case is due to Freiman [30] and Ruzsa [80]. In  $\mathbb{Z}$ , there are no interesting finite subgroups and, of course, all nilprogressions are automatically abelian progressions as in Example 3. Writing the group operation on  $\mathbb{Z}$  using addition as usual, the Freiman-Ruzsa theorem may be stated as follows.

**Theorem 2.7** (Freiman-Ruzsa). *Suppose that  $A \subset \mathbb{Z}$  is a  $K$ -approximate group, that is to say  $2A \subset X + A$  for some set  $X \subset \mathbb{Z}$  with  $|X| \leq K$ . Then there is a proper<sup>3</sup> progression  $P = P(u_1, \dots, u_d; N_1, \dots, N_d) := \{n_1 u_1 + \dots + n_d u_d : 0 \leq n_i < N_i\}$  which  $K'$ -controls  $A$ . Here,  $d$  and  $K'$  are bounded as functions of  $K$  only.*

The definition of “control” here is the same as in Theorem 2.2.

The optimal bounds on  $d$  and  $K'$  are not known. Following a sequence of developments by Chang [17] and Schoen [89], the state of the art is contained in a breakthrough paper of Sanders [85]. Sanders shows that we may take  $d \sim (\log K)^C$  and  $K' \sim e^{(\log K)^C}$  for some reasonable value of  $C$  (such as  $C = 4$ ). A key open question, known as the *Polynomial Freiman-Ruzsa conjecture*, asks whether one could in fact take  $d \sim \log K$  and  $K' \sim K^C$ . The bound  $d \sim \log K$  is significant as if  $P$  is a progression of this dimension then  $P \cup -P$  is itself a  $K^{C'}$ -approximate group. If one is prepared to sacrifice  $K'$  then bounds of this strength are known due to work of Freiman-Bilu [5] and Tao and the author [46]. For much greater depth on the quantitative issues surrounding Theorem 2.7, the recent survey of Sanders [86] may be consulted.

A solution to the Polynomial Freiman-Ruzsa conjecture ought to have serious applications in additive number theory – perhaps, for example, to questions about bases such as Waring's problem. However, no definite deductions of this type have so far been made.

Another abelian setting has attracted a lot of interest, and that is the case  $G = \mathbb{F}_2^\mathbb{Z}$ . In this group, where we have  $2 \cdot x = 0$  for every  $x$ , there are no interesting nilprogressions and one is left only with subgroups. Theorem 2.2 in this case is due to Ruzsa [81], and it may be stated as follows.

**Theorem 2.8** (Ruzsa). *Suppose that  $A \subset \mathbb{F}_2^\mathbb{Z}$  is a  $K$ -approximate group, that is to say  $2A \subset X + A$  for some set  $X \subset \mathbb{F}_2^\mathbb{Z}$  with  $|X| \leq K$ . Then there is a subgroup  $H \subset \mathbb{F}_2^\mathbb{Z}$  which  $K'$ -controls  $A$ . Here  $K'$  is bounded as a function of  $K$  only.*

The question of whether  $K'$  may be taken to be polynomial in  $K$  is also known as the Polynomial Freiman-Ruzsa conjecture, and it has attracted much attention. Ruzsa [81] attributes it to Katalin Marton. Once again the best results are due to Sanders [85], who shows that we may take  $K' \sim e^{(\log K)^C}$ . Ruzsa (unpublished, but

<sup>3</sup>This means that all the sums  $n_1 u_1 + \dots + n_d u_d$  under consideration are distinct.

see [39]) offers several equivalent formulations, of which the following is perhaps particularly appealing.

**Conjecture 2.9.** *Let  $V$  be a finite-dimensional vector space in characteristic 2. Suppose that  $f : V \rightarrow V$  satisfies the “approximate homomorphism” condition*

$$\{f(x+y) - f(x) - f(y) : x, y \in V\} \subset S.$$

*Then there is a linear map  $\tilde{f} : V \rightarrow V$  and a set  $\tilde{S}$  with  $|\tilde{S}| \ll |S|^C$  such that*

$$\{f(x) - \tilde{f}(x) : x \in V\} \subset \tilde{S}.$$

There is an extremely extensive literature on the closely-related notion of a *quasimorphism* in contexts arising in geometric group theory; see [70] for a brief introduction. At present there seems to be little connection between that context, where the concern is usually with quasimorphisms on infinite groups, and ours.

### 3 Approximate polynomials

#### 3.1 Gowers norms and polynomial phases

We turn now to the discussion of a different inequality. If  $f : \mathbb{Z} \rightarrow \mathbb{C}$  is a function and  $h \in \mathbb{Z}$  then we define the multiplicative derivative  $\Delta_h f$  by  $\Delta_h f(x) = f(x)f(x+h)$ . Let  $k \geq 2$  be a fixed integer, and suppose that  $N$  is large in terms of  $k$ . Write  $[N] = \{1, \dots, N\}$ . Then we define the *Gowers  $U^k[N]$ -norm* of  $f$  by

$$\|f\|_{U^k[N]} = \left( \mathbb{E}_{x, h_1, \dots, h_k} \Delta_{h_1} \dots \Delta_{h_k} f(x) \right)^{1/2^k}.$$

Here, the average  $\mathbb{E}$  is over all  $x, h_1, \dots, h_k$  for which  $x + \omega_1 h_1 + \dots + \omega_k h_k \in [N]$  for all  $\omega_i \in \{0, 1\}$ ; this means that the Gowers norm depends only on the values taken by  $f$  on  $[N]$ . In taking  $2^k$ th roots we make use of the not completely obvious fact that  $\mathbb{E}_{x, h_1, \dots, h_k} \Delta_{h_1} \dots \Delta_{h_k} f(x)$  is real and non-negative. This is not too hard to prove by induction: see for example [96]. The basic theory of Gowers norms was originally developed in [36].

The Gowers norms satisfy the following rather trivial inequality: if  $\mathcal{F}$  is the set of all functions  $f : [N] \rightarrow \mathbb{C}$  with  $\|f\|_\infty \leq 1$ ,  $v(f) = \|f\|_{U^k[N]}$ , then

$$v(f) \leq 1. \tag{3.1}$$

(The inequality is indeed trivial – bound every instance of  $f(\cdot)$  in the definition of the Gowers  $U^k$ -norm by 1).

When does equality occur, that is to say for which  $f$  do we have  $v(f) = 1$ ? For this to happen, we must have<sup>4</sup>

$$\Delta_{h_1} \dots \Delta_{h_k} f(x) = 1 \quad \text{for all } x, h_1, \dots, h_k. \tag{3.2}$$

This implies that  $|f(x)| = 1$  for all  $x$ , and so we may write  $f(x) = e^{2\pi i \phi(x)}$  for some phase function  $\phi : \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ . The condition (3.2) then becomes

$$\partial_{h_1} \dots \partial_{h_k} \phi(x) = 0 \quad \text{for all } x, h_1, \dots, h_k, \tag{3.3}$$

where  $\partial_h$  is the additive difference operator defined by  $\partial_h \psi(x) = \psi(x) - \psi(x+h)$ .

The condition (3.3) is satisfied if and only if  $\phi$  is a polynomial of degree at most  $k-1$ . The “if” direction of this assertion may be established by induction

<sup>4</sup>Here and in what follows we ignore the restriction that  $x + \omega_1 h_1 + \dots + \omega_k h_k \in [N]$ ; this has little bearing on the argument.

on the degree, since if  $\phi$  is a polynomial of degree  $d$  then, for fixed  $h$ ,  $\Delta_h \phi$  is a polynomial of degree  $d - 1$ . Then “only if” direction can then be established by taking  $h_1 = \dots = h_k = 1$  in (3.3), which tells us that  $\phi(x+k)$  is uniquely determined as a function of  $\phi(x), \phi(x+1), \dots, \phi(x+k-1)$ . Therefore  $\phi$  is uniquely determined by its values at  $0, 1, \dots, k-1$ , and hence coincides with the unique polynomial of degree at most  $k-1$  which agrees with it at those points.

The stability question, that is to say the characterisation of those  $f$  for which  $v(f) \geq 1 - o(1)$ , is already interesting. It turns out that  $f$  must be closely approximated by a polynomial phase  $e^{2\pi i \phi(x)}$ . A precise statement and proof of this result may be found in [24, Theorem 1.2]. The argument there is analogous to an earlier argument [1] in a finite field setting, which has applications to property testing in theoretical computer science.

As with Young’s inequality, however, our main focus here will be on the robustness question: for which  $f$  do we have  $v(f) \geq \frac{1}{K}$ ? This is known as the *inverse question for the Gowers norms*. When  $k = 2$ , all such  $f$  are at least somewhat related to exponentials of linear phases (the solutions to the equality question  $v(f) = 1$ ).

**Theorem 3.1.** *Suppose that  $f : [N] \rightarrow \mathbb{C}$  is a function with  $|f(x)| \leq 1$  for all  $x$ , and that  $\|f\|_{U^2[N]} \geq \frac{1}{K}$ . Then there is some  $\theta \in \mathbb{R}/\mathbb{Z}$  such that*

$$\frac{1}{N} \left| \sum_{x \in N} f(x) e^{-2\pi i \theta x} \right| \geq \frac{1}{K^2}.$$

The proof of this is an exercise in Fourier analysis, given in detail in [40, Proposition 8.2]. When  $k \geq 3$ , however, the situation is different. There are examples of functions  $f : [N] \rightarrow \mathbb{C}$  with  $|f(x)| \leq 1$  for all  $x$ ,  $\|f\|_{U^3[N]} \geq \frac{1}{K}$ , but for which

$$\frac{1}{N} \left| \sum_{x \in N} f(x) e^{-2\pi i \phi(x)} \right| \ll N^{-c} \quad (3.4)$$

for all quadratic phases  $\phi : \mathbb{Z} \rightarrow \mathbb{R}/\mathbb{Z}$ . It is actually rather easy to give an example of such a function, though considerably less easy to prove rigorously that it *is* an example: take  $f(x) = e^{2\pi i \alpha x \{\beta x\}}$ , with  $\alpha, \beta \in \mathbb{R}$  sufficiently irrational numbers such as  $\alpha = \sqrt{2}$  and  $\beta = \sqrt{3}$ . Here  $\{t\}$  denotes fractional part. The “reason” this function  $f$  has large  $U^3[N]$ -norm is that the phase  $\phi(x) = \alpha x \{\beta x\}$ , whilst it does not satisfy the derivative condition (3.3) exactly, does satisfy this condition for a positive proportion of  $x, h_1, h_2, h_3$ : in fact whenever  $\{\beta x\}, \{\beta h_1\}, \{\beta h_2\}, \{\beta h_3\} \in [-\frac{1}{10}, \frac{1}{10}]$ . Establishing (3.4) rigorously is quite tricky.

A more natural way to construct such functions is as *nilsequences*. These objects should be thought of as “higher-order characters” generalising the linear exponentials  $\Phi(n) = e^{2\pi i \theta n}$ . To explain this generalisation we write  $\Phi$  in the form

$$\Phi(n) = F(p(n)) \quad (3.5)$$

where

- $p(n) = T^n 0$ , where  $T : \mathbb{R} \rightarrow \mathbb{R}$  is the translation map  $Tx = x + \theta$ ;
- $F(x) = e^{2\pi i x}$ . Note that this function is  $\mathbb{Z}$ -periodic.

A nilsequence corresponds to a generalisation of this in which  $\mathbb{R}$  is replaced by a simply-connected nilpotent Lie group  $G$  and  $\mathbb{Z}$  is replaced by a lattice  $\Gamma \subset G$ . With this setup, a nilsequence is of the form (3.5) with

- $p(n) = T^n \text{id}_G$ , where  $T : G \rightarrow G$  is a *nilrotation*, that is to say a map of the form  $Tx = xg$  for some  $g \in G$ ;
- $F : G \rightarrow \mathbb{C}$  is smooth and  $\Gamma$ -*automorphic*, which means that  $F(\gamma x) = F(x)$  for all  $x \in G$  and  $\gamma \in \Gamma$ .

For example, we could take  $G$  to be the Heisenberg group  $\begin{pmatrix} 1 & \mathbb{R} & \mathbb{R} \\ 0 & 1 & \mathbb{R} \\ 0 & 0 & 1 \end{pmatrix}$  and  $\Gamma$  to be the lattice  $\begin{pmatrix} 1 & \mathbb{Z} & \mathbb{Z} \\ 0 & 1 & \mathbb{Z} \\ 0 & 0 & 1 \end{pmatrix}$ . In fact for various reasons one usually considers a generalisation of this in which  $p(n)$  is a *polynomial sequence* on the group  $G$ . We will not discuss this important issue here, save to remark that it leads to essentially the same concept in the end due to a lifting argument of Furstenberg [31, p. 31] (see also [54, Appendix C]). We say that  $\Phi$  is an  $s$ -step nilsequence if the underlying nilpotent group  $G$  has nilpotency class  $s$ , that is to say if the lower central series of  $G$  is

$$G = G_1 \supset G_2 \supset \cdots \supset G_s \supset G_{s+1} = \{\text{id}_G\},$$

with  $G_s$  nontrivial. For the Heisenberg group we have  $s = 2$ .

To give a specific example in the Heisenberg case, we need to specify  $g$  and an automorphic function  $F$ . The element  $g$  can of course be specified just by choosing a matrix. Nontrivial automorphic functions can be defined by hand (define  $F$  to be a smooth bump function supported on the interior a fundamental domain for  $\Gamma \backslash G$  and extend by automorphy). In the Heisenberg case there is a construction, pointed out in [64] for example, using the *Jacobi  $\theta$ -function*  $\theta(u, z) := \sum_n e^{\pi i z n^2 + 2\pi i n u}$  by defining  $F\left(\begin{pmatrix} 1 & x & z \\ 0 & 1 & y \\ 0 & 0 & 1 \end{pmatrix}\right) = e(z)e^{-\pi x^2}\theta(y + ix)$ .

Functions of the form  $e^{2\pi i \phi(x)}$  with  $\phi$  quadratic (that is to say, the functions for which  $\|f\|_{U^3[N]} = 1$ ) “morally” arise as nilsequences on the Heisenberg group by taking  $g = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & \alpha \\ 0 & 0 & 1 \end{pmatrix}$  and  $F(x, y, z) = e^{2\pi i(z - y[x])}$ . Indeed it may be checked by a computation that in this case we have  $\Phi(n) = e^{2\pi i \phi(n)}$  with  $\phi(n) = -\frac{1}{2}\alpha n(n+1)$ . The slight technical issue here is that, although  $F$  is automorphic (as may be confirmed by a computation) it is only piecewise smooth.

The following turns out to be true.

**Theorem 3.2.** *Suppose that  $\Phi(n)$  is an  $s$ -step nilsequence with<sup>5</sup>  $\|\Phi\|_2 = 1$ . Then  $\|\Phi\|_{U^{s+1}[N]} \geq \frac{1}{K}$ , where  $K$  is bounded above in terms of  $s$  and the “complexity” of  $\Phi$ .*

Giving a proper definition of the complexity is a rather tedious matter; it must take account of various parameters associated with  $G, \Gamma$  and the smoothness of the automorphic function  $F$ . The appendices of [49] go into considerable further detail.

The key to the proof is the observation that the multiplicative derivative  $\Delta_h \Phi(n) = \Phi(n)\overline{\Phi(n+h)}$  is an  $(s-1)$ -step nilsequence, which allows us to proceed inductively. In fact, this is not quite true, but it is true if the automorphic function  $F$  has the additional transformation property

$$F(g_s x) = \xi(g_s)F(x) \tag{3.6}$$

for every  $g_s$  in  $G_s$ , the last nontrivial subgroup in the lower central series of  $G$ , for some character  $\xi : G_s \rightarrow \mathbb{C}$  invariant under  $\Gamma$ . One may reduce to this case by a Fourier expansion on cosets of  $G_s$ . It is in effecting this Fourier expansion that the

<sup>5</sup>Here  $\|\Phi\|_2^2 = \frac{1}{N} \sum_{n \leq N} |\Phi(n)|^2$ . Some condition is needed to ensure that we do not have  $\Phi(n) = 0$  identically.

complexity of  $\Phi$ , and in particular the smoothness properties of  $F$ , comes into play. Suppose now that we do have the transformation property (3.6). For fixed  $h$  we have

$$\Delta_h \Phi(n) = F(T^n \text{id}_G) \overline{F(T^{n+h} \text{id}_G)} = \tilde{F}(T^n \text{id}_G)$$

where  $\tilde{F}(x) = F(x) \overline{F(T^h x)}$ . The function  $\tilde{F}$  is easily seen to be  $\Gamma$ -automorphic, and moreover it is invariant under  $G_s$ :

$$\tilde{F}(g_s x) = F(g_s x) \overline{F(T^n g_s x)} = \xi(g_s) F(x) \overline{\xi(g_s) F(T^n x)} = F(x) \overline{F(T^n x)} = \tilde{F}(x).$$

Here we used the fact that  $G_s$  is central in  $G$  to commute  $T$  with multiplication by  $g_s$ . As a consequence,  $\tilde{F}$  descends to an automorphic function on  $G/G_s$ , a nilpotent Lie group of class  $s-1$ . (Unfortunately the preceding discussion was actually quite a serious oversimplification, as in the definition of Gowers norm  $h$  is not fixed but can vary over  $[-N, N]$ . With the argument just described, various smoothness norms of  $\tilde{F}$  depend heavily on  $h$  and to get around this a more complicated construction is required. Such a construction is given in [49, Section 7].)

We have seen that functions supplying equality in (3.1), the inequality  $v(f) \leq 1$ , are polynomial phases. Theorem 3.2 states that nilsequences of step  $k-1$  and suitably bounded complexity are solutions to the corresponding robustness problem  $v(f) \geq \frac{1}{K}$ , and so we think of them as “approximate polynomials” (or, more accurately, approximate polynomial phases). The discussion of the previous paragraph, where we saw that the multiplicative derivatives of nilsequences (with an additional invariance property) are nilsequences of lower step, adds further weight to this philosophy.

It is very far from true that every solution to the robustness problem is a nilsequence. Indeed<sup>6</sup> if  $f : [N] \rightarrow \mathbb{C}$  and if  $\varepsilon : [N] \rightarrow \{-1, 1\}$  is a random  $\pm 1$ -valued function then almost surely  $v(f) \approx v(f + \varepsilon)$ . However, it is true that every solution is somewhat related to a nilsequence.

**Theorem 3.3.** *Suppose that  $f : [N] \rightarrow \mathbb{C}$  is a function with  $|f(n)| \leq 1$  and that  $\|f\|_{U^k[N]} \geq \frac{1}{K}$ . Then there is a  $(k-1)$ -step nilsequence  $\Phi(n)$  with  $|\Phi(n)| \leq 1$  for all  $n$  such that*

$$\frac{1}{N} \left| \sum_{n \in N} f(n) \overline{\Phi(n)} \right| \geq \frac{1}{K'},$$

where  $K'$  and the complexity of  $\Phi$  are bounded in terms of  $K$  and  $k$  only.

Note that Theorem 3.3 is a generalisation of Theorem 3.1, which was essentially the case  $k=2$ . This result is due to Tao, Ziegler and the author [54] and is known as the *Inverse Theorem for the Gowers norms*. A weaker “local” version of it was obtained by Gowers (in [35] for  $k=3$ , and in [36] for general  $k$ ). The case  $k=3$  was established by Tao and the author [45], and the case  $k=4$  by Tao, Ziegler and the author [52]. It should most certainly be mentioned that the relevance of nilpotent Lie groups in this general arena first became apparent in the context of ergodic theory in works of Conze, Lesigne, Furstenberg and Weiss [20, 32, 33, 34]. A result which may be thought of as an “ergodic analogue” of Theorem 3.3 was obtained by Host and Kra [63] (see also independent work of Ziegler [102]). The notion of nilsequence itself, as well as the word, was introduced by Bergelson, Host and Kra [4]. See [71] for a nice introduction to these connections.

<sup>6</sup>Passing from  $f$  to  $f + \varepsilon$  may destroy the property  $|f(x)| \leq 1$ , but we ignore this for the sake of illustration.



The conjecture to which Theorem 3.3 is a solution, together with potential applications of it to prime numbers, was formulated by Tao and the author [47] about four years before it was finally proved. The proof is unfortunately inordinately long and complicated. For a summary in about 20 pages, see [53]. An alternative approach has been developed by Szegedy [93] and Camarena-Szegedy [16], based in part on the work of Host and Kra, but these papers are not an easy read either.

The converse of Theorem 3.3 is also true, with appropriate changes to the constants. The proof is relatively straightforward and goes along very similar lines to the sketch of the proof of Theorem 3.2 we gave above: see [52, Appendix G] for further details.

Although we do not plan to discuss it much here, there has also been a good deal of work on finite field analogues of Theorem 3.3, which have applications in theoretical computer science. In addition to work by various subsets of the authors named above, we note that Samorodnitsky [83] established the case  $k = 3$  of Theorem 3.3 in the particularly interesting setting where  $[N]$  is replaced by a vector space in characteristic 2.

**Theorem 3.4.** *Suppose that  $V$  is a finite-dimensional vector space in characteristic 2 and that  $f : V \rightarrow \mathbb{C}$  is a function with  $|f(x)| \leq 1$  for all  $x \in V$ . Suppose that  $\|f\|_{U^3(V)} \geq \frac{1}{K}$ . Then there is a function  $\Phi : V \rightarrow \mathbb{C}$  of the form  $\Phi(x) = (-1)^{\psi(x)}$ , where  $\psi : V \rightarrow \mathbb{F}_2$  is a quadratic form, such that*

$$\frac{1}{|V|} \left| \sum_{x \in V} f(x) \Phi(x) \right| \geq \frac{1}{K'}.$$

Here  $K'$  is bounded in terms of  $K$  only.

The definition of the  $U^3(V)$ -norm is entirely analogous to that of the  $U^3[N]$  norm, except that the average  $\mathbb{E}$  is now simply taken over all  $x, h_1, h_2, h_3 \in V$ . Samorodnitsky obtained a bound of the form  $K' \sim e^{K^C}$ , but by combining his methods with the work of Sanders [85] one could improve this to  $K' \sim e^{(\log K)^C}$ . The similarity of these bounds to those stated in conjunction with Theorem 2.8 is no coincidence. Indeed a close relationship between the structure theory of approximate subgroups of  $\mathbb{F}_2^{\mathbb{Z}}$  and Theorem 3.4 was discovered by Tao and the author [48] and independently by Lovett [76]. In particular, it is known that the Polynomial Freiman-Ruzsa conjecture for finite fields, which is equivalent to Conjecture 2.9, is also equivalent to having a bound of shape  $K' \ll K^C$  in Theorem 3.4.

A similar equivalence between bounds in Theorem 2.7 and the case  $k = 3$  of Theorem 3.3 was developed in [48]: in other words the theories of approximate subgroups of  $\mathbb{Z}$  and of approximate quadratic polynomials are in a sense the same. I have often informally advanced the speculation that looking for a more effective proof of Theorem 3.3 may be a way of attacking the Polynomial Freiman-Ruzsa Conjecture, though without any convincing ideas about how this might be achieved.

### 3.2 Applications

The theory of Gowers norms as described in the last section was for the most part developed to study arithmetic progressions. Gowers himself was interested in Szemerédi's theorem, and Tao and the author were subsequently concerned with arithmetic progressions of primes. In [47] it was observed that the theory applied

to a fairly wide class of “linear” problems, including questions about linear configurations of primes. Since this theory was discussed<sup>7</sup> in my 2006 ICM lecture [40] and is described in the article of Ziegler in these *Proceedings*, we restrict ourselves to a very brief account.

The connection of the Gowers norms to linear configurations comes from results called *generalised von Neumann inequalities*, which have the form

$$|T(f_1, \dots, f_t)| \ll \inf_{i=1, \dots, t} \|f_i\|_{U^{s+1}[N]}. \quad (3.7)$$

Here,  $f_1, \dots, f_t : [N] \rightarrow [-1, 1]$  are functions and

$$T(f_1, \dots, f_t) = \mathbb{E}_{(n_1, \dots, n_d) \in S} f_1(\psi_1(n_1, \dots, n_d)) \dots f_t(\psi_t(n_1, \dots, n_d)),$$

where the  $\psi_i : \mathbb{Z}^d \rightarrow \mathbb{Z}$  are affine-linear forms and  $S$  is a “nice” set (for example a convex set). For any system of forms  $\psi_1, \dots, \psi_d$  which is not degenerate in a certain sense, there is a value of  $s$  for which (3.7) holds. For example, if  $d = 2$ ,  $t = 3$  and  $\psi_1(n_1, n_2) = n_1$ ,  $\psi_2(n_1, n_2) = n_1 + n_2$ ,  $\psi_3(n_1, n_2) = n_1 + 2n_2$  (3-term arithmetic progressions) then we may take  $k = 2$ , whilst if  $d = 2$ ,  $t = 4$  and  $\psi_1(n_1, n_2) = n_1$ ,  $\psi_2(n_1, n_2) = n_1 + n_2$ ,  $\psi_3(n_1, n_2) = n_1 + 2n_2$ ,  $\psi_4(n_1, n_2) = n_1 + 3n_2$  (4-term arithmetic progressions) then we may take  $k = 3$ . The degenerate configurations are those in which some two of the  $\psi_i$  have equal homogeneous part, up to scalar equivalence: thus for example we cannot take  $\psi_1(n_1) = n_1$  and  $\psi_2(n_1) = n_1 + 2$ . The proof of any generalised von Neumann inequality is conceptually quite easy, involving only several applications of the Cauchy-Schwarz inequality, but notationally a little unpleasant. A general form of (3.7) was obtained in [47, Appendix D]. Furthermore the inequality was established there under a weaker condition on the  $f_i$  than boundedness, namely that  $|f_i(x)| \leq \nu(x)$  pointwise for some “pseudorandom measure”  $\nu$ . This is crucial for applications to the primes.

Ignoring a few technicalities, the manner in which (3.7) is applied to the primes is as follows. For technical convenience the primes are weighted using the von Mangoldt function  $\Lambda$ , defined by  $\Lambda(n) = \log p$  if  $n = p^k$  is a prime power and  $\Lambda(n) = 0$  otherwise. We are interested in  $T(\Lambda, \dots, \Lambda)$ , which counts how often the linear forms  $\psi_1(n_1, \dots, n_d), \dots, \psi_t(n_1, \dots, n_d)$  all take prime values as  $(n_1, \dots, n_d)$  ranges over a set  $S$ . To estimate this we split  $\Lambda$  in a certain manner as

$$\Lambda = \Lambda^\sharp + \Lambda^\flat, \quad (3.8)$$

where  $\Lambda^\sharp$  is “structured” and  $\Lambda^\flat$  is “unstructured”. Since  $T$  is multilinear, we may split  $T(\Lambda, \dots, \Lambda)$  as a sum of  $T(\Lambda^\sharp, \dots, \Lambda^\sharp)$  plus  $2^t - 1$  other terms, each of which involves at least one copy of  $\Lambda^\flat$ . The first term provides the main term in the asymptotic formula for  $T(\Lambda, \dots, \Lambda)$ , and the aim is then to show that the other  $2^t - 1$  terms are all small. By (3.7), this may be accomplished if it can be shown that

$$\|\Lambda^\flat\|_{U^{s+1}[N]} = o(1).$$

By the inverse theorem for the Gowers norms, Theorem 3.3 (in the contrapositive), it is enough to establish that

$$\frac{1}{N} \left| \sum_{n \leq N} \Lambda^\flat(n) \overline{\Phi(n)} \right| = o(1) \quad (3.9)$$

---

<sup>7</sup>Naturally, however, this account is quite out of date and in particular predates the general case of Theorem 3.3.

for every  $s$ -step nilsequence  $\Phi(n)$  of bounded complexity. At least, this would be so were it not for the restriction  $|f(x)| \leq 1$  in Theorem 3.3: a large part of [47] is devoted to removing this restriction, showing that Theorem 3.3 implies a more general version of itself in which we only assume that  $|f(x)| \leq \nu(x)$  for some pseudorandom measure  $\nu$ .

The actual decomposition (3.8) we choose is based on the formula  $\Lambda(n) = \sum_{d|n} \mu(d) \log(n/d)$ , where  $\mu$  is the Möbius function. It transpires that the task of establishing (3.9) may be further reduced to establishing that

$$\frac{1}{N} \left| \sum_{n \leq N} \mu(n) \overline{\Phi(n)} \right| \ll_A \log^{-A} N \quad (3.10)$$

for every  $A > 0$ . This statement was formerly known as the “Möbius and nilsequences conjecture”, but it is now a theorem of Tao and the author [50]. Although the paper [50] is relatively short, it depends crucially on the much longer paper [49], in which various properties of nilsequences are established, in particular with regard to the distribution of finite orbit segments  $(T^n \text{id}_G)_{n \leq N}$  in  $\Gamma \backslash G$ . This work, like other material in this section, was motivated by earlier developments in the ergodic theory community, in particular work of Leon Green [55] and papers of Leibman of both an algebraic [73] and an ergodic-theoretic [74] nature.

### 3.3 Open questions

For me the key open question is to find the “right” proof of the inverse conjecture for the Gowers norms. At the moment the proofs are unsatisfactory on a conceptual level (the notion of a nilsequence is extremely natural, so it would be disappointing if it genuinely required 100+ pages to explain its role in Theorem 3.3). Furthermore, these proofs provide rather poor bounds for the complexity of the nilsequence  $\Phi$ , particularly when  $k \geq 4$  (in fact for  $k \geq 5$  the proofs provide no explicit bounds at all due to the use of ultrafilter arguments, though once again an explicit bound could in principle be extracted via quantifier elimination). As noted above it would be particularly interesting, in view of the link to approximate subgroups of  $\mathbb{Z}$ , to find a new approach to the inverse theorem when  $k = 3$ .

A more specific question is whether there is some smaller “natural” class of nilsequences. The space  $C^\infty(\Gamma \backslash G)$  of automorphic functions is extremely large, but we know for example that in the case  $\Gamma = \mathbb{Z}$ ,  $G = \mathbb{R}$  the exponentials  $e^{2\pi i x}$  have a special role. Eigenfunctions of Laplacians are one natural avenue of enquiry. Furthermore the space of all simply-connected nilpotent Lie groups  $G$  together with lattices  $\Gamma$  is also extremely large and complicated, and it may be natural to focus on some subclass (for example free nilpotent Lie groups).

## 4 Other directions

To conclude this article I want to mention a personal selection of a few other inequalities where the equality, stability and robustness questions may hide interesting algebraic or somewhat algebraic structure. In some cases there is at least a tenuous connection to the main sections of the article, and in others less so.

#### 4.1 Inverse questions for the large sieve

Let  $\mathcal{A}$  be a set of natural numbers with the property that  $|\mathcal{A} \pmod{p}| \leq \frac{1}{2}(p+1)$  for all sufficiently large primes  $p$ . The large sieve guarantees that  $|\mathcal{A} \cap [N]| \ll N^{1/2}$  for all  $N$ . This is sharp up to a multiplicative constant, as is shown by taking  $\mathcal{A}$  to be the set of squares (or the set of integer values of an arbitrary quadratic with rational coefficients).

It may well be the case that a very strong robustness assertion holds: if there is some  $K$  such that  $|\mathcal{A} \cap [N]| \geq \frac{1}{K}N^{1/2}$  for all sufficiently large  $N$  then  $\mathcal{A}$  is contained, up to a finite set, in the set of values of a rational quadratic. See [43] for evidence in this direction. This type of question was first raised by Helfgott and Venkatesh [61]; see also [99, 100].

#### 4.2 Point-line configurations

Let  $\mathcal{P} \subset \mathbb{R}^2$  be a set of  $n$  points, no four on a line<sup>8</sup>. Write  $T(\mathcal{P})$  for the number of pairs  $(x, y) \in \mathcal{P}$  of distinct points for which there is a third distinct point  $z \in \mathcal{P}$  on the line  $\overline{xy}$ . Trivially,  $T(\mathcal{P}) \leq n(n-1)$ . Less obviously, equality cannot occur: this follows from a famous result known as the Sylvester–Gallai theorem.

Almost-equality can occur: we can obtain  $T(\mathcal{P}) = n^2 - O(n)$  by taking  $\mathcal{P}$  to be a suitable set of points on a suitable cubic curve (for example a coset of a subgroup on an elliptic curve, although there are singular examples too). This was noted by Sylvester in the 1860s [92]. Conversely, it was recently shown by Tao and the author [51] that there is a strong converse to this statement.

It would be very interesting to have an understanding of those  $\mathcal{P}$  for which  $T(\mathcal{P}) = n^2(1 - o(1))$  (the stability question) or, more ambitiously,  $T(\mathcal{P}) \geq \frac{n^2}{K}$  (the robustness question). The paper [51] only covers the extreme end of the stability region. It is possible that cubic structure is responsible for all such  $\mathcal{P}$ . There are links here to the theory of approximate groups: for example, finite approximate subgroups of elliptic curve groups are a source of examples of such sets  $\mathcal{P}$ .

An interesting nontrivial result in higher dimensions is [3], motivated by applications in theoretical computer science.

#### 4.3 The Littlewood Problem

Suppose that  $A \subset \mathbb{Z}$  is a set of  $n$  integers. Then it was established 30 years ago by Konyagin [69] and McGehee-Pigno-Smith [77], answering a question of Littlewood [75], that

$$\int_0^1 \left| \sum_{a \in A} e^{2\pi i \theta a} \right| d\theta \gg \log n.$$

Earlier results had been obtained by Paul Cohen and others. This is sharp up to the constant, as is shown by taking  $A$  to be an arithmetic progression of length  $n$ . (In fact, this example may also provide the sharp constant, a conjecture known as the *Strong Littlewood Conjecture*.) Very little is known about the robustness question, that is to say about the structure of those  $A$  for which

$$\int_0^1 \left| \sum_{a \in A} e^{2\pi i \theta a} \right| d\theta \leq K \log n.$$

---

<sup>8</sup>This condition is included here for simplicity, but can probably be relaxed.

It is possible that such  $A$  are very close to being unions of a few arithmetic progressions. If so, this would have applications to questions in combinatorial number theory about sum-free sets due to a connection established by Bourgain [6]. For some partial results and a further discussion, see [78].

#### 4.4 No-three-in-a-line

Let  $p$  be an odd prime, and suppose that  $A \subset \text{PG}(2, p)$  is a set containing no three distinct points in a line<sup>9</sup>. (Here,  $\text{PG}(2, p)$  is the 2-dimensional projective space over  $\mathbb{F}_p$ , thus  $|\text{PG}(2, p)| = p^2 + p + 1$ ). It is very easy to see that  $|A| \leq p + 2$  and an exercise to show that  $|A| \leq p + 1$ . Equality occurs when  $A$  is a conic. Remarkably, a celebrated result of Segre [90] shows that in fact equality occurs *only* when  $A$  is a conic.

The stability question was resolved by Voloch [98], building upon remarkable work of Segre. Voloch shows that any  $A$  with no three-in-a-line and  $|A| \geq \frac{44}{45}p$  is contained in a conic. This argument is quite deep, depending on an application of the polynomial method [95] as well as bounds of Stöhr and Voloch [91] about counting points on high degree curves.

The robustness question, that is to say the classification of those  $A$  with  $|A| \geq \frac{1}{K}p$ , is very interesting. There are examples coming from cubic curves, such as  $A = \{(x : x^3 : 1) : 0 < x < p/3\}$ . So far as I am aware there is no example in the literature to contradict the possibility that all sets  $A \subset \text{PG}(2, p)$  with no-three-on-a-line and  $|A| \geq \frac{1}{K}p$  have all but  $o(p)$  of their points lying on a curve of degree at most 3. So far as I am aware no-one has explicitly conjectured this either, so perhaps I shall take this opportunity to do so.

There is a superficial link to a notorious problem of Dudeney [23] about whether there is a set  $A$  of  $2N$  points on the grid  $[N] \times [N]$  with no three in a line. There are many fewer colinear triples in  $[N] \times [N]$  than in  $\mathbb{Z}/p\mathbb{Z} \times \mathbb{Z}/p\mathbb{Z}$  for  $p \sim N$ , however, so the study of sets  $A$  such as this is likely to be even harder than the problem discussed above. Nonetheless, the best-known examples (with  $|A| \sim 3N/2$ , see [57]) are given by very algebraic constructions. It seems likely that the answer to Dudeney's question is negative.

#### 4.5 Sidon sets

Suppose that  $A \subset [N]$  is a set with the property that all pairwise sums  $x + y$  with  $x, y \in A$  are distinct, apart from the obvious coincidences  $x + y = y + x$ . Such a set  $A$  is called a Sidon set. It is very easy to see that  $|A| \ll \sqrt{N}$ , and with more care (an argument of Erdős and Turán) one may show that  $|A| \leq (1 + o(1))\sqrt{N}$ . There are examples of Sidon sets  $A$  with  $|A| = (1 - o(1))\sqrt{N}$ , all constructed in a highly algebraic manner using finite fields. There are different variants due to Bose, Ruzsa and Singer. It is possible that the stability question (that is, the classification of those Sidon set  $A$  with  $|A| = (1 - o(1))\sqrt{N}$ ) has a satisfactory answer, but there is no obvious guess, based on the known examples, as to what it might be. The robustness question, that is to say the classification of those  $A$  with  $|A| \geq \frac{1}{K}\sqrt{N}$ , is of course even more difficult. A discussion of it was had on the blog of Tim Gowers [38]. In commenting on that discussion, Terence Tao raised the possibility that an answer to this question could lead to progress on a famous and old problem of

<sup>9</sup>Such sets are called “arcs” in the literature, which is extremely extensive.

Erdős, namely to determine if there is an additive basis  $\mathcal{A}$  of the natural numbers of order 2 (i.e.  $\mathcal{A} + \mathcal{A} = \mathbb{N}$ ) with an absolute bound on the number of representations of  $x$  as a sum of two elements of  $\mathcal{A}$ .

## 5 Acknowledgements

I would like to thank Sean Eberhard, Bryna Kra, Freddie Manners, Peter Sarnak and Terence Tao for comments on a draft of this article. I wish to thank the last of these for our extensive collaboration over the last decade, which has so far led to 30 joint papers. I also thank my other coauthors, three of whom are speaking at this congress. It is only because of these mathematicians that I have the opportunity to present these topics at the 2014 ICM.

## References

- [1] N. Alon, T. Kaufman, M. Krivelevich, S. Litsyn and D. Ron, *Testing low-degree polynomials over  $\text{GF}(2)$* , RANDOM-APPROX 2003, 188–199.
- [2] A. Balog and E. Szemerédi, *A statistical theorem of set addition*, Combinatorica **14** (1994), 263–268.
- [3] B. Barak, Z. Dvir, A. Yehudayoff and A. Wigderson, *Rank bounds for design matrices with applications to combinatorial geometry and locally correctable codes*, In Proceedings of the 43rd annual ACM symposium on Theory of computing, STOC '11, 519–528.
- [4] V. Bergelson, B. Host and B. Kra, *Multiple recurrence and nilsequences*, (with an appendix by I. Ruzsa), Invent. Math. **160**, no. 2 (2005), 261–303.
- [5] Y. Bilu, *Structure of sets with small sumset*, Structure Theory of Set Addition, Astérisque **258** (1999), 77–108.
- [6] J. Bourgain, *Estimates related to sumfree subsets of sets of integers*, Israel J. Math. **97** (1997), 71–92.
- [7] J. Bourgain and A. Gamburd, *Uniform expansion bounds for Cayley graphs of  $\text{SL}_2(\mathbb{F}_p)$* , Ann. Math. **167** (2008), no. 2, 625–642.
- [8] J. Bourgain, A. Gamburd and P. Sarnak, *Affine linear sieve, expanders and sum-product*, Invent. Math. **179** (2010), no. 3, 559–644.
- [9] J. Bourgain, A. Glibichuk and S. Konyagin, *Estimates for the number of sums and products and for exponential sums in fields of prime order*, J. London Math. Soc. **73** (2006), no. 2, 380–398.
- [10] J. Bourgain, N. Katz, and T. C. Tao. *A sum-product estimate in finite fields, and applications*, Geom. Funct. Anal. (GAFA) **14** (2004), no. 1, 27–57.
- [11] E. Breuillard, B. J. Green, and T. C. Tao, *Approximate subgroups of linear groups*, Geom. Funct. Anal. (GAFA) **21** (2011), no. 4, 774–819.
- [12] E. Breuillard, B. J. Green and T. C. Tao, *Suzuki groups as expanders*, Groups Geom. Dyn. **5** (2011), no. 2, 281–299.
- [13] E. Breuillard, B. J. Green and T. C. Tao, *The structure of approximate groups*, Publ. Math. IHES **116** (2012), no. 1, 115–221.
- [14] E. Breuillard, B. J. Green and T. C. Tao, *Small Doubling in Groups*, Erdős Centennial, Bolyai Society Mathematical Studies **25** (2013), 129–151.
- [15] E. Breuillard, B. J. Green, R. Guralnick and T. C. Tao, *Expansion in Finite Simple Groups of Lie Type*, to appear in J. European Math. Soc.
- [16] O. A. Camarena and B. Szegedy, *Nilspaces, nilmanifolds and their automorphisms*, 2010 preprint. <http://arxiv.org/abs/1009.3825>.
- [17] M.-C. Chang, *A polynomial bound in Freiman’s theorem*, Duke Math. J. **113** (2002), no. 3, 399–419.
- [18] M.-C. Chang, *Product theorems in  $\text{SL}_2$  and  $\text{SL}_3$* , J. Inst. Math. Jussieu **7** (2008), no. 1, 1–25.
- [19] G. Cooperman, *Towards a practical, theoretically sound algorithm for random generation in finite groups*, unpublished. arXiv:math/0205203.
- [20] J. P. Conze and E. Lesigne, *Sur un théorème ergodique pour des mesures diagonales*, C. R. Acad. Sci. Paris **306** (1988), 491–493.

- [21] E. Croot and O. Sisask, *A probabilistic technique for finding almost-periods of convolutions*, Geom. Funct. Anal. (GAFA) **20** (2010), no. 6, 1367–1396.
- [22] J. D. Dixon, *Generating random elements in finite groups*, Electronic J. Comb. **15** (2008), R94.
- [23] H. E. Dudeney, *Amusements in Mathematics*, Nelson, Edinburgh, 1917.
- [24] T. Eisner and T. C. Tao, *Large values of the Gowers-Host-Kra seminorms*, J. Anal. Math. **117** (2012), 133–186.
- [25] G. Elekes and Z. Király, *On the combinatorics of projective mappings*, J. Algebraic Combin. **14** (2001), no. 3, 183–197.
- [26] A. Figalli and D. Jerison *Quantitative stability for sumsets in  $\mathbb{R}^n$* , 2013 preprint.
- [27] A. Figalli and D. Jerison, *Quantitative stability for the Brunn-Minkowski inequality*, 2013 preprint.
- [28] J. J. F. Fournier, *Sharpness in Young’s inequality for convolution*, Pacific J. Math. **72** (1977), no. 2, 383–397.
- [29] G. A. Freiman, *Groups and the inverse problems of additive number theory*, In Number-theoretic studies in the Markov spectrum and in the structural theory of set addition (Russian), 175–183. Kalinin. Gos. Univ., Moscow, 1973.
- [30] G. A. Freiman, *Foundations of a structural theory of set addition*, American Mathematical Society, Providence, R. I., 1973. Translated from the Russian, Translations of Mathematical Monographs, Vol. 37.
- [31] H. Furstenberg, *Recurrence in ergodic theory and combinatorial number theory*, Princeton University Press 1981.
- [32] H. Furstenberg, *Nonconventional ergodic averages*, The legacy of John von Neumann (Hempstead, NY, 1988), Proc. Sympos. Pure. Math. **50** (1990), 43–56, AMS (Providence, RI).
- [33] H. Furstenberg, *From the Erdős-Turán conjecture to ergodic theory – the contribution of combinatorial number theory to dynamics*, in Paul Erdős and his mathematics, I (Budapest, 1999), 261–277, Bolyai Soc. Math. Stud. **11**, János Bolyai Math. Soc., Budapest, 2002.
- [34] H. Furstenberg, and 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), 193–227, Ohio State Univ. Math. Res. Inst. Publ., **5**, de Gruyter, Berlin, 1996.
- [35] W. T. Gowers, *A new proof of Szemerédi’s theorem for progressions of length four*, Geom. Funct. Anal. (GAFA) **8** (1998), no. 3, 529–551.
- [36] W. T. Gowers, *A new proof of Szemerédi’s theorem*, Geom. Funct. Anal. (GAFA) **11** (2001), no. 3, 465–588.
- [37] W. T. Gowers, *Quasirandom groups*, Combin. Probab. Comput. **17** (2008), no. 3, 363–387.
- [38] W. T. Gowers, *What are dense Sidon subsets of  $\{1, 2, \dots, n\}$  like?*, blog post. Available at [gowers.wordpress.com/2012/07/13/what-are-dense-sidon-subsets-of-1-2-n-like/](http://gowers.wordpress.com/2012/07/13/what-are-dense-sidon-subsets-of-1-2-n-like/)
- [39] B. J. Green, *Notes on the Polynomial Freiman-Ruzsa Conjecture*, unpublished. Available at [people.maths.ox.ac.uk/greenbj/papers/PFR.pdf](http://people.maths.ox.ac.uk/greenbj/papers/PFR.pdf)
- [40] B. J. Green, *Generalising the Hardy-Littlewood method for primes*, International Congress of Mathematicians. Vol. II, 373–399, Eur. Math. Soc., Zurich, 2006.
- [41] B. J. Green, *Approximate groups and their applications: work of Bourgain, Gamburd, Helfgott and Sarnak*, 25 pages, Current Events Bulletin of the AMS, 2010.
- [42] B. J. Green, *Barbados lecture notes*, 2010. Transcript available at <http://www.cs.mcgill.ca/~denis/additive-lectures-v2.pdf>
- [43] B. J. Green and A. J. Harper, *Inverse questions for the large sieve*, to appear in Geom. Funct. Anal. (GAFA).
- [44] B. J. Green and I. Z. Ruzsa, *Freiman’s theorem in an arbitrary abelian group*, J. Lond. Math. Soc. **75** (2007), no. 2, 163–175.
- [45] B. J. Green and T. C. Tao, *An inverse theorem for the Gowers  $U^3$ -norm, with applications*, Proc. Edinburgh Math. Soc. **51**, no. 1, 71–153.
- [46] B. J. Green and T. C. Tao, *Compressions, convex geometry and the Freiman-Bilu theorem*, Quart. J. Math. (Oxford) **57** (2006), no. 4, 495–504.
- [47] B. J. Green and T. C. Tao, *Linear equations in primes*, Ann. Math **171** (2010), no. 3, 1753–1850.
- [48] B. J. Green and T. C. Tao, *An equivalence between inverse sumset theorems and inverse conjectures for the  $U^3$ -norm*, Math. Proc. Camb. Phil. Soc. **149** (2010), no. 1, 1–19.

- [49] B. J. Green and T. C. Tao, *The quantitative behaviour of polynomial orbits on nilmanifolds*, Ann. Math. **175** (2012), no. 2, 465–540.
- [50] B. J. Green and T. C. Tao, *The Möbius function is strongly orthogonal to nilsequences*, Ann. Math. **175** (2012), no. 2, 541–566.
- [51] B. J. Green and T. C. Tao, *On sets defining few ordinary lines*, Disc. Comp. Geom. **50** (2013), no. 2, 409–468.
- [52] B. J. Green, T. C. Tao and T. Ziegler, *An inverse theorem for the Gowers  $U^4[N]$  norm*, Glasgow Math. J. **53** (2011), 1–50.
- [53] B. J. Green, T. C. Tao and T. Ziegler, *An inverse theorem for the Gowers  $U^{s+1}[N]$ -norm*, announcement, Electronic Research Announcements **18** (2011), 69–90.
- [54] B. J. Green, T. C. Tao and T. Ziegler, *An inverse theorem for the Gowers  $U^{s+1}[N]$ -norm*, Ann. Math **176** (2012), no. 2, 1231–1372.
- [55] L. W. Green, *Spectra of nilflows*, Bull. Amer. Math. Soc. **67** (1961) 414–415.
- [56] M. Gromov, *Groups of polynomial growth and expanding maps*, Publ. Math. IHES **53** (1981), 53–73.
- [57] R. R. Hall, T. H. Jackson, A. Sudbery and K. Wild, *Some advances in the no-three-in-line problem*, J. Combinatorial Theory Ser. A **18** (1975), 336–341.
- [58] G. H. Hardy and J. E. Littlewood, *Some new properties of Fourier constants*, Math. Ann. **97** (1927), 159–209.
- [59] H. A. Helfgott, *Growth and generation in  $SL_2(\mathbb{Z}/p\mathbb{Z})$* , Ann. Math. **167** (2008), 601–623.
- [60] H. A. Helfgott, *Growth in  $SL_3(\mathbb{Z}/p\mathbb{Z})$* , J. Eur. Math. Soc. **13** (2011), no. 3, 761–851.
- [61] H. A. Helfgott and A. Venkatesh, *How small must ill-distributed sets be?*, in Analytic Number Theory: Essays in honour of Klaus Roth, 224–234, Cambridge Univ. Press, Cambridge, 2009.
- [62] S. Hoory, N. Linial, and A. Wigderson, *Expander graphs and their applications*, Bull. Amer. Math. Soc. (N.S.) **43** (2006), no. 4, 439–561.
- [63] B. Host and B. Kra, *Nonconventional ergodic averages and nilmanifolds*, Ann. Math. **161** (2005), no. 1, 397–488.
- [64] B. Host and B. Kra, *Analysis of two step nilsequences*, Annales de l’institut Fourier **58** (2008), no. 5, 1407–1453.
- [65] E. Hrushovski, *Stable group theory and approximate subgroups*, J. Amer. Math. Soc. **25** (2012), no. 1, 189–243.
- [66] M. Kassabov, *Symmetric groups and expander graphs*, Invent. Math. **170** (2007), no. 2, 327–354.
- [67] M. Kassabov, A. Lubotzky and N. Nikolov, *Finite simple groups as expanders*, Proc. Nat. Acad. Sci. **103** (2006), no. 16, 6116–6119.
- [68] B. Kleiner, *A new proof of Gromov’s theorem on groups of polynomial growth*, J. Amer. Math. Soc. **23** (2010), no. 3, 815–829.
- [69] S. V. Konyagin, *On the Littlewood problem*, Izv. Akad. Nauk SSSR Ser. Mat. **45** (1981), no. 2, 243–265.
- [70] D. Kotschick, *What is a quasimorphism?* Notices Amer. Math. Soc. **51** (2004), no. 2, 208–209.
- [71] B. Kra, *From combinatorics to ergodic theory and back again*, Proceedings of International Congress of Mathematicians, Madrid 2006, volume III, 57–76.
- [72] M. J. Larsen and R. Pink, *Finite subgroups of algebraic groups*, J. Amer. Math. Soc. **24** (2011), no. 4, 1105–1158.
- [73] A. Leibman, *Polynomial sequences in groups*, Journal of Algebra **201** (1998), 189–206.
- [74] A. Leibman, *Pointwise convergence of ergodic averages for polynomial sequences of translations on a nilmanifold*, Ergodic Th. Dyn. Systems **25** (2005), no. 1, 201–213.
- [75] J. E. Littlewood, *Some problems in real and complex analysis*, Heath Mathematical Monographs, D. C. Heath and co., Lexington, MA 1968.
- [76] S. Lovett, *Equivalence of polynomial conjectures in additive combinatorics*, Combinatorica **32** (2012), no. 5, 607–618.
- [77] O. C. McGehee, L. Pigno and B. Smith, *Hardy’s inequality and the  $L^1$ -norm of exponential sums*, Ann. Math. **113** (1981), no. 3, 613–618.
- [78] G. Petridis, *The  $L^1$ -norm of exponential sums in  $\mathbb{Z}^d$* , Math. Proc. Camb. Phil. Soc. **154** (2013), no. 3, 381–392.
- [79] L. Pyber and E. Szabó, *Growth in finite simple groups of Lie type of bounded rank*, preprint (2010). arXiv:1005.1858.



- [80] I. Z. Ruzsa, *Sums of finite sets, Number theory (New York, 1991–1995)*, 281–293. Springer, New York, 1996.
- [81] I. Z. Ruzsa, *An analog of Freiman’s theorem in groups*, in *Structure Theorem of Set Addition*, Astérisque **258** (1999) 323–326.
- [82] A. Salehi-Golsefidy and P. Sarnak, *Affine sieve*, J. Amer. Math. Soc. **26** (2013), no. 4, 1085–1105.
- [83] A. Samorodnitsky, *Low-degree tests at large distances*, STOC 2007.
- [84] T. Sanders, *On a nonabelian Balog-Szemerédi-type lemma*, J. Aust. Math. Soc. **89** (2010), no. 1, 127–132.
- [85] T. Sanders, *On the Bogolyubov-Ruzsa lemma*, Anal. PDE **5** (2012), no. 3, 627–655.
- [86] T. Sanders, *The structure theory of set addition revisited*, Bull. Amer. Math. Soc. (N.S.) **50** (2013), no. 1, 93–127.
- [87] P. Sarnak, *Equidistribution and primes*, available at <http://web.math.princeton.edu/sarnak/EquidPrimes.pdf>
- [88] P. Sarnak and X. X. Xue, *Bounds for multiplicities of automorphic representations*, Duke Math. J. **64** (1991), no. 1, 207–227.
- [89] T. Schoen, *Near optimal bounds in Freiman’s theorem*, Duke Math. J. **158** (2011), no. 1, 1–12.
- [90] B. Segre, *Ovals in a finite projective plane*, Canadian J. Math. **7** (1955), 414–416.
- [91] K.-O.- Stöhr and J. F. Voloch, *Weierstrass points and curves over finite fields*, Proc. London Math. Soc. (3) **52** (1986), no. 1, 1–19.
- [92] J. Sylvester, *Mathematical Question 2571*. Educational Times, February 1868.
- [93] B. Szegedy, *On higher-order Fourier analysis*, 2012 preprint. <http://arxiv.org/abs/1203.2260>.
- [94] T. C. Tao, *Product set estimates for non-commutative groups*, Combinatorica **28** (2008), no. 5, 547–594.
- [95] T. C. Tao, *Algebraic combinatorial geometry: the polynomial method in arithmetic combinatorics, incidence combinatorics, and number theory*, preprint 2013. <http://arxiv.org/abs/1310.6482>
- [96] T. C. Tao and V. H. Vu, *Additive Combinatorics*, Cambridge Stud. Adv. Math. **105** (2006).
- [97] P. P. Varjú, *Expansion in  $SL_d(O_K/I)$ ,  $I$  square-free*, J. Eur. Math. Soc. **14** (2012), no. 1, 273–305.
- [98] J. F. Voloch, *Arcs in projective planes over prime fields*, J. Geom. **38** (1990), no. 1–2, 198–200.
- [99] M. N. Walsh, *The inverse sieve problem in high dimensions*, Duke Math. J. **161** (2012), no. 10, 2001–2022.
- [100] M. N. Walsh, *The algebraicity of ill-distributed sets*, preprint. Available at <http://arxiv.org/abs/1307.0259>
- [101] W. H. Young, *On the multiplication of successions of Fourier constants*, Proc. Roy. Soc. London Ser. A **87** (1912), 331–339.
- [102] T. Ziegler, *Universal Characteristic Factors and Furstenberg Averages*, J. Amer. Math. Soc. **20** (2007), 53–97.

MATHEMATICAL INSTITUTE, RADCLIFFE OBSERVATORY QUARTER, WOODSTOCK ROAD, OXFORD OX2 6GG, ENGLAND

*E-mail address:* `ben.green@maths.ox.ac.uk`