Spectral structure of sets of integers

Ben Green[1]

Abstract

Let A be a small subset of a finite abelian group, and let R be the set of points at which its Fourier transform is large. A result of Chang states that R has a great deal of additive structure. We give a statement and proof of this result and an example which shows that it is sharp. We also discuss some of the applications of it which have so far been discovered. Finally we discuss some related open questions.

1. Introduction, notation and definitions. Harmonic analysis has been used to great effect in additive number theory for over 150 years. In this article we will look at one specific theme which has received attention of late. This is the principle that the large values of the Fourier transform of a small set have a great deal of structure.

We begin by introducing a small amount of notation which is necessary for the discussion. Throughout this paper N will be a large prime number and we will write $\mathbb{Z}_N$ for the additive group[2] of residues modulo N. If $E = \{e_1, \ldots, e_L\} \subseteq \mathbb{Z}_N$ we write $\mathrm{Span}(E)$ for the set of all sums $s(\varepsilon) = \sum_j \varepsilon_j e_j$ with $\varepsilon_j \in \{-1, 0, 1\}$. We will write $\omega_N^x = e^{2\pi i x/N}$. Often the subscript N will be suppressed, as the value of N will be clear from the context. If $f : \mathbb{Z}_N \to \mathbb{C}$ is a function and $r \in \mathbb{Z}_N$ then we define the Fourier transform of f at r by

$$\widehat{f}(r) \;=\; \sum_x f(x)\omega^{rx}.$$

We will adopt the convenient notational practice of identifying sets with their characteristic functions.

2. Chang's structure theorem. In a recent paper [5] of Chang the following result is stated[3].

Theorem 1 (Chang) *Let $\rho, \alpha \in [0,1]$, Let $A \subseteq \mathbb{Z}_N$ be a set of size αN and let $R \subseteq \mathbb{Z}_N$ be the set of all r for which $|\widehat{A}(r)| \geq \rho|A|$. Then there is a set $E \subseteq \mathbb{Z}_N$ with $|E| \ll \rho^{-2}\log\left(\frac{1}{\alpha}\right)$ such that $R \subseteq \mathrm{Span}(E)$.*

[1]The author is a Fellow of Trinity College, Cambridge. Address: Trinity College, Cambridge CB2 1TQ, England.

[2]Much of what we have to say can be generalised to arbitrary finite abelian groups. However in this article we will eschew such generality and discuss instead the group $\mathbb{Z}_N$ and, occasionally, the group $\mathbb{Z}_2^n$.

[3]Chang's paper seems to be the first place where this result is explicitly stated. However, similar ideas can be found in an earlier paper of Bourgain [4], and the whole circle of ideas perhaps originated with Rudin [14]. We will discuss Rudin's inequality later in the paper.

It is convenient to give a name to the situation covered by this theorem. Thus if $A, R \subseteq \mathbb{Z}_N$ and if $\rho \in (0,1)$ then we say that A is ρ-large at R if $|\widehat{A}(r)| \geq \rho |A|$ for all $r \in R$.

Theorem 1 is an extremely interesting result. Parseval's theorem implies that the set R has size at most $\rho^{-2}\alpha^{-1}$, but for small α this is much bigger than the size of E guaranteed by Chang's result. Theorem 1 may thus be viewed as saying that the "large spectrum" of a small set is very highly structured.

There are already two rather different applications of this result in combinatorial number theory. The first, in Chang's original paper [5], concerns Freiman's theorem on sets with small sumset. The second, due to the author [7], concerns arithmetic progressions in sumsets. We will discuss this application in §6.

In [5] Theorem 1 is derived from an inequality of Rudin. We will describe a proof of this result in the next two sections. A rather different proof was shown to us by I.Z. Ruzsa (personal communication), an account of which may be found in [9] (2). In §5 we give the deduction of Theorem 1.

3. An inequality of Rudin. The main sources for this discussion were [10] and [14]. Let us begin by stating the inequality of Rudin that interests us. We say that a set $\Lambda = \{\lambda_1, \ldots, \lambda_m\} \subseteq \mathbb{Z}_N$ is *dissociated*[4] if the only solution to the equation

$$\varepsilon_1\lambda_1 + \cdots + \varepsilon_m\lambda_m \;=\; \varepsilon_1'\lambda_1 + \cdots + \varepsilon_m'\lambda_m$$

with $\varepsilon_j, \varepsilon_j' \in \{-1, 0, 1\}$ is the trivial solution $\varepsilon_j = \varepsilon_j' = 0$.

In the statement of Rudin's inequality, Λ will be assumed to be dissociated and we will regard Λ and $\mathbb{Z}_N$ as finite measure spaces (M_1, μ_1) and (M_2, μ_2) respectively. μ_1 will be the counting measure, so that $\mu_1(M_1) = |\Lambda|$, while μ_2 will be the *normalised* counting measure, which means that $\mu_2(M_2) = 1$. Write $B(M_i)$ for the space of functions on M_i.

Proposition 1 (Rudin) *Let $T : B(M_1) \to B(M_2)$ be the linear map which sends a sequence $(a_n)_{n\in\Lambda} \in B(M_1)$ to the function $f(x) = \sum_{n\in\Lambda} a_n\omega^{nx}$. Then for any $p > 2$ we have the bound*

$$\|T\|_{2\to p} \;\leq\; 12\sqrt{p}$$

on the L^2–L^p norm of the operator T.

Written out in full, this means that

$$\|f\|_p^p \;=\; N^{-1}\sum_x \left|\sum_{n\in\Lambda} a_n\omega^{nx}\right|^p \;\leq\; (144p)^{p/2}\left(\sum_{n\in\Lambda}|a_n|^2\right)^{p/2}.$$

[4] The reader should be aware that various slightly different definitions will be encountered in the literature.

The formulation we have used in Proposition 1 is, perhaps, more suggestive.

Observe that $\left(\sum_{n\in\Lambda}|a_n|^2\right)^{1/2}$ is equal to $\|f\|_2$. The inequality may, therefore, be interpreted as a statement to the effect that the L^2 and L^p norms of a function whose spectrum is dissociated are comparable.

In the next few paragraphs we show that Rudin's inequality is true, on average, for modified versions of f in which the a_n have been subjected to random and independent changes of sign. This may seem like a curious thing to do, so we offer some motivation at the end of the section.

Suppose then that X_j, $j\in\Lambda$ are independent Bernoulli random variables taking values in $\{\pm 1\}$ and let us consider the random function

$$X(x) \;=\; \sum_{n\in\Lambda} a_n X_n \omega^{nx}.$$

A sensible way to estimate $\mathbb{E}\|X\|_p^p$ is to write

$$\mathbb{E}\|X\|_p^p \;=\; N^{-1}\sum_{x\in\mathbb{Z}_N}\int_0^\infty \mathbb{P}\left(|X(x)|\geq t^{1/p}\right)\,dt, \tag{1}$$

recalling the availability of certain *large deviation inequalities* associated with the names of Bernstein, Chernoff and Hoeffding. The following is a typical example:

Proposition 2 *Let $Z_1,\ldots,Z_n$ be independent complex-valued random variables with zero means and with $|Z_i|\leq a_i$ for all $i=1,\ldots,n$. Let t be a positive real number. Then*

$$\mathbb{P}\left(|Z_1+\cdots+Z_n|\geq t\right) \;\leq\; 2e^{-t^2/4\sum|a_i|^2}.$$

See, for example, [9] (1). □

Substituting into (1) gives

$$\mathbb{E}\|X\|_p^p \;\leq\; 2\int_0^\infty e^{-t^{2/p}/4\sum|a_i|^2},$$

an expression which may be evaluated explicitly as

$$2^{p+1}\Gamma\left(\frac{p+2}{2}\right)\left(\sum|a_i|^2\right)^{p/2}.$$

A short calculation using a sharp form of Stirling's formula then yields

$$\mathbb{E}\|X\|_p^p \;\leq\; (6\sqrt{p})^p\left(\sum|a_i|^2\right)^{p/2}. \tag{2}$$

This is all very well, but there is no reason to suppose that the behaviour of f should be linked in any way to that of the random function X. The dissociativity of Λ is exactly what provides such a link, a fact that we shall endeavour to explain now.

We begin with the observation that the L^p norm of $f(x)$ is the same as that of

$$f(x+\theta) \ = \ \sum_{n\in\Lambda} a_n \omega^{n\theta}\omega^{nx}$$

for any $\theta \in \mathbb{Z}_N$ that we may care to select. Suppose that for any choice of a sign function $\varepsilon : \Lambda \to \{\pm 1\}$ we could find a θ with $\omega^{n\theta} \approx \varepsilon_n$ for all $n \in \Lambda$ (we will not be precise about what we mean by the approximate symbol $\approx$ here). Now (2) implies that there is a specific choice of ε for which

$$\left\| \sum_n a_n \varepsilon_n \omega^{nx} \right\|_p \ \leq \ 6\sqrt{p} \left(\sum_n |a_n|^2 \right)^{1/2}. \tag{3}$$

Selecting an appropriate θ would then allow us to recover an inequality of the desired form for f. Now whether or not one can find such a θ is related to issues of simultaneous diophantine approximation. Observe that if there is a "small" linear relation amongst the elements of Λ - say, for example, $\{5, 7, 12\} \subseteq \Lambda$ - then such a θ need not exist. One can prove using Fourier analysis that this is necessary and sufficient; that is to say, if there are no small linear relations then θ can always be found, whatever the choice of signs ε_n. The phrase "no small linear relations" turns out to mean that Λ is linearly independent over a set such as $\{-D, -D+1, \ldots, D\}$ where $D \sim |\Lambda|$. Unfortunately this is a stronger condition than just dissociativity, but when it does hold f models the L^p behaviour of the randomised sum X very closely. It turns out however that dissociativity is exactly what we need to make a different approach to the comparison of f and X work.

4. Riesz products and Young's inequality. It is convenient to have a notation for twisted versions of f like those we encountered in (3). If $\varepsilon : \Lambda \to \{\pm 1\}$ is a sign function then write

$$f_\varepsilon \ = \ \sum_{n\in\Lambda} a_n \varepsilon_n \omega^{nx}.$$

Write $p_\varepsilon(x)$ for the *Riesz product*

$$p_\varepsilon(x) \ = \ 2\prod_{n\in\Lambda} \left(1 + \tfrac{\varepsilon_n}{2}\left(\omega^{nx} + \omega^{-nx}\right)\right).$$

Claim 1 *We have* $f = f_\varepsilon * p_\varepsilon$.

Proof of claim. This can be established by a fairly straightforward computation. We have

$$f_\varepsilon * p_\varepsilon(x) \;=\; 2N^{-1} \sum_y \sum_{m\in\Lambda} a_m \varepsilon_m \omega^{m(x-y)} \prod_{n\in\Lambda} \left(1 + \tfrac{\varepsilon_n}{2}\left(\omega^{ny} + \omega^{-ny}\right)\right). \tag{4}$$

Multiplying out the product and changing the order of summation, one is confronted with a weighted sum of terms of the form

$$\sum_y \omega^{(n_1+\cdots+n_r-n_1'-\cdots-n_s'-m)y}, \tag{5}$$

where the n_i, n_i' are distinct elements of Λ and $m \in \Lambda$. The dissociativity of Λ implies that such a sum is zero unless $r = 1$, $s = 0$ and $m = n_1$, in which case it equals N. It is easy to see that the weight attached to (5) in this case (in the expanded out version of (4)) is

$$N^{-1} a_{n_1} \varepsilon_{n_1}^2 \omega^{n_1 x} \;=\; N^{-1} a_{n_1} \omega^{n_1 x},$$

and the claim follows quickly. □

Now the Riesz product p_ε is non-negative, and so $\|p_\varepsilon\|_1$ is simply $N^{-1}\sum_x p_\varepsilon(x)$. This sum may easily be calculated by expanding out another product and using dissociativity, and it turns out that $\|p_\varepsilon\|_1 = 2$. Thus by Young's inequality and the claim we have

$$\begin{aligned} \|f\|_p &= \|f_\varepsilon * p_\varepsilon\|_p \\ &\le \|f_\varepsilon\|_p \|p_\varepsilon\|_1 \\ &= 2\|f_\varepsilon\|_p \end{aligned} \tag{6}$$

for any $p \ge 2$ and any choice of sign function ε. Now (2) implies that there is a specific choice of ε for which

$$\|f_\varepsilon\|_p \;\le\; 6\sqrt{p} \left(\sum_n |a_n|^2\right)^{1/2}.$$

Thus

$$\|f\|_p \;\le\; 12\sqrt{p} \left(\sum_n |a_n|^2\right)^{1/2},$$

and Proposition 1 follows immediately. □

5. Completion of the proof of Chang's theorem. In this section we derive Theorem 1 from Proposition 1. It turns out that the dual form of Proposition 1 is easier to work with in this context. This takes the form

$$\|T^*\|_{p'\to 2} \;\le\; 12\sqrt{p}, \tag{7}$$

where p' is the dual exponent of p. Here $T^* : B(M_2) \to B(M_1)$ is the adjoint of T, which is easily seen to be given by

$$T^* f(n) \;=\; N^{-1} \sum_x f(x) \omega^{nx}$$

for $n \in \Lambda$.

Now recall that we are interested in a set $A \subseteq \mathbb{Z}_N$ with cardinality αN, and we have written R for the set of all $r \in \mathbb{Z}_N$ for which $|\widehat{A}(r)| \geq \rho |A|$. We wish to show that R has lots of structure, and we do this by proving that it does not contain a very large unstructured subset. To this end let Λ be a *maximal* dissociated subset of R, and apply (7) with $p = \log(1/\alpha)$ and f equal to the characteristic function of A. It is easy to check that

$$\|T^* A\|_2 \;=\; N^{-1} \left(\sum_{n \in \Lambda} |\widehat{A}(n)|^2 \right)^{1/2} \;\geq\; \rho \alpha |\Lambda|^{1/2}$$

and that

$$\|A\|_{p'} \;=\; \alpha^{1/p'} \;=\; \alpha^{1-1/p} \;\leq\; e\alpha.$$

It follows immediately that

$$|\Lambda| \;\ll\; \rho^{-2} \log(1/\alpha).$$

Thus Λ, some maximal dissociated subset of R, is rather small. The maximality implies that the addition of any new $r \in R$ will spoil the dissociativity property. It is easy to see that this implies that each r is expressible as $\sum_j \eta_j \lambda_j$, where $\eta \in \{-2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2\}$, and Theorem 1 follows on taking $E = \frac{1}{2}\Lambda \cup \Lambda \cup 2\Lambda$. □

6. Chang's theorem and progressions in sumsets. In this section we discuss the paper [7]. At various points we will use the function $\| \, . \, \| : \mathbb{Z}_N \to \mathbb{R}$ defined as follows. If x is a residue class modulo N, pick a representative $\overline{x}$ for x from the interval $\{-(N-1)/2, \ldots, (N-1)/2\}$. Set $\|x\| = |\overline{x}/N|$.

The main result of [7] is the following improvement of a result of Bourgain [4].

Theorem 2 *Let $C, D \subseteq \mathbb{Z}_N$ have cardinalities γN and δN respectively. Then there is an absolute constant $c > 0$ such that $C + D$ contains an AP of length at least*

$$\exp\left(c\left((\gamma\delta \log N)^{1/2} - \log\log N\right)\right).$$

This looks a little technical. It is perhaps easier to think of γ and δ as being fixed positive reals: then the theorem says that for large N the sumset $C+D$ contains a progression of length $e^{c'\sqrt{\log N}}$.

The first step of the argument involves the introduction of a concept that we called, in [7], *hereditary non-uniformity* (HNU). Roughly speaking, a set $A \subseteq \mathbb{Z}_N$ was said to be HNU if every subset $B \subseteq A$ has a large Fourier coefficient. As pointed out to us by Gowers (personal communication, and see also [16]), this is not quite the "right" definition. It is more natural to consider, instead of subsets of A, arbitrary functions supported on A. Before stating the lemma which explains this, we introduce two very temporary pieces of notation. Let A be a subset of $\mathbb{Z}_N$, write A° for its complement and let c be a positive real. We say that A has property $P(c)$ if, for any function f supported on A, we have

$$\sup_{r\neq 0} |\widehat{f}(r)| \ \geq \ c \left| \sum_x f(x) \right|. \tag{8}$$

We say that A has property $Q(c)$ if there is a function g supported on A° for which

$$c \sum_{r\neq 0} |\widehat{g}(r)| \ \leq \ \left| \sum_x g(x) \right|. \tag{9}$$

Lemma 1 *The properties $P(c)$ and $Q(c)$ coincide.*

Proof. We begin by proving that $P(c) \Rightarrow Q(c)$, which is the easier of the two implications. It is also the only part of the lemma which is actually used in proving Theorem 2. Suppose then that A has the property $Q(c)$, and let g be a function supported on A° and satisfying (9). If f is any function supported on A then we have $\sum_x f(x)g(x) = 0$, which implies that $\sum_r \widehat{f}(r)\widehat{g}(r) = 0$. By the triangle inequalty, this gives

$$\sup_{r\neq 0} |\widehat{f}(r)| \sum_{r\neq 0} |\widehat{g}(r)| \ \geq \ \left| \sum_x f(x) \right| \left| \sum_x g(x) \right|.$$

Thus indeed A has property $P(c)$.

To prove that $P(c) \Rightarrow Q(c)$ we use the (finite-dimensional) Hahn-Banach thorem. Suppose that A has property $P(c)$. Write X for the space of all $\mathbb{C}$-valued functions on $\mathbb{Z}_N$, and define a seminorm $\gamma : X \to \mathbb{R}_{\geq 0}$ by $\gamma(f) = c^{-1} \sup_{r\neq 0} |\widehat{f}(r)|$ (this has all the properties of a norm, except that $\gamma(\mathbf{1}) = 0$). Let Y be the space spanned by the complex-valued functions on A° and the constant function $\mathbf{1}$, and define a linear functional $T : Y \to \mathbb{R}$ by

$$T(f + \lambda \mathbf{1}) \ = \ \sum_x f(x). \tag{10}$$

for all f supported on A° and $\lambda \in \mathbb{C}$. Since A has property $P(c)$, this satisfies

$$|Tf| \;\leq\; \gamma(f)$$

for all $f \in Y$. By the Hahn-Banach theorem we may extend T to a functional T' on all of X which satisfies the same bound,

$$|T'f| \;\leq\; \gamma(f). \tag{11}$$

This functional will be of the form $T'f = \langle f, \psi \rangle$ for some function $\psi : \mathbb{Z}_N \to \mathbb{C}$, and it is clear from (10) that $\psi(x) = 1$ for all $x \in A^\circ$. We claim that the function $g = \psi - 1$ satisfies (9). To see this, observe that by (11) we have, for any function $f : \mathbb{Z}_N \to \mathbb{R}$, the bound

$$\sum_x f(x)\overline{\psi(x)} \;\leq\; c^{-1} \sup_{r \neq 0} |\widehat{f}(r)|. \tag{12}$$

Take f to be the following function, defined by specifying its Fourier transform:

$$\widehat{f}(r) \;=\; \begin{cases} N \exp(i \arg \widehat{\psi}(r)) & r \neq 0 \\ 0 & r = 0. \end{cases}$$

The left-hand side of (12) is then just $\sum_{r \neq 0} |\widehat{\psi}(r)|$, which is equal to $\sum_{r \neq 0} |\widehat{g}(r)|$. Thus

$$\text{LHS of (12)} \;=\; \sum_{r \neq 0} |\widehat{g}(r)|.$$

On the other hand $\sup_{r \neq 0} |\widehat{f}(r)|$ is at most N. Furthermore, since $\sum \psi(x) = T'\mathbf{1} = 0$, we have $|\sum g(x)| = N$. Thus

$$\text{RHS of (12)} \;\leq\; c^{-1} \left| \sum g(x) \right|.$$

The proof is complete. □

We shall say that a set A is c-hereditarily non-uniform (HNU) if it has the property $P(c)$ (or $Q(c)$). Note once again that this notion is rather stronger than that used in the paper [7].

Using the easy direction of Lemma 1, we can demystify the connection between sumsets and HNU sets.

Lemma 2 *Let $C, D \subseteq \mathbb{Z}_N$ have $|C| = \gamma N$ and $|D| = \delta N$. Let A be the complement of $C + D$. Then A is $\sqrt{\gamma\delta}$-HNU.*

Proof. All one has to do is check that $C+D$ has property $Q(\sqrt{\gamma\delta})$. This is easy; by Parseval's identity and the Cauchy-Schwarz inequality one can satisfy (9) by taking $g = C * D$. □

For the remainder of the discussion we will assume for simplicity that $\gamma = \delta = \frac{1}{4}$; this does not simplify the argument, but keeps the number of unspecified variables down. The heart of [7] is the following proposition, which in combination with Lemma 2 leads quickly to Theorem 2.

Proposition 3 *Suppose that A is $\frac{1}{4}$-HNU. Then A°, the complement of A, contains an AP of length at least $e^{\sqrt{\log N}/128}$.*

The proof of this goes roughly as follows. Suppose that A is HNU. Then every $B \subseteq A$ must be non-uniform in the sense of having a large Fourier coefficient. Take B to be as close to uniform as possible among all subsets of A having a certain size. That is, fix $\beta \in \mathbb{R}$ and let $B \subseteq A$ have

$$\sup_{r \neq 0} |\widehat{B}(r)|$$

minimal subject to $|B| = \lfloor \beta N \rfloor$. If the value of this minimum is $\eta|B|$ then we have $\eta \geq 1/4$ from the definition of HNU. The value of β gets chosen at the end of the proof to optimise the argument; it turns out that a sensible choice is $\beta = e^{-\sqrt{\log N}/64}$.

The idea is now that we try to modify the set B to give a new subset $B' \subseteq A$ of the same size but which is more uniform than B. Of course this is impossible, so there must be something wrong with any such modification technique that we might care to write down.

One way of modifying B is as follows. Choose small random subsets $Y \subseteq B$ and $X \subseteq \mathbb{Z}_N$ of the same size $t = 2^{18} \log N$. Form the (multi)set $B_0 = (B \setminus Y) \cup X$. What are the Fourier coefficients of B_1? Applying a Bernstein-type inequality similar to that in Proposition 2, it is not hard to see that with positive probability $\widehat{Y}(r) \approx t\widehat{B}(r)/|B|$ and that $\widehat{X}(r)$ is small compared with t for all $r \neq 0$. Picking specific X and Y for which these rough statements hold, we see that

$$|\widehat{B}_0(r)| \;\leq\; |\widehat{B}(r)| \tag{13}$$

for all $r \neq 0$, which certainly implies that

$$\sup_{r \neq 0} |\widehat{B}_0(r)| \;\leq\; \sup_{r \neq 0} |\widehat{B}(r)|.$$

Naturally this does not violate the extremal property of B, because B_0 need not be a subset of A. However we can try and modify it by changing the elements $x_1, \ldots, x_t$ of X. Let us try changing x_1 to $x_1 + h_1$ to give a set B_1. Then

$$\widehat{B}_1(r) \;=\; \widehat{B}_0(r) + \omega^{rx_1}\left(\omega^{rh_1} - 1\right).$$

Now if $\|rh_1\|$ were small for all r then this would differ insignificantly from $\widehat{B}_0(r)$. If we performed t such operations, changing each x_j to $x_j + h_j$ in turn to give sets $B_2, \dots, B_t$, then we might still have $|\widehat{B}_t(r)| \leq |\widehat{B}(r)|$. Sadly there is no h_1 with this property. However for many r we are not at all concerned about changing $\widehat{B}_0(r)$ quite substantially. Indeed if $|\widehat{B}(r)| \leq \eta|B|/2$ then, by (13), we also have $|\widehat{B}_0(r)| \leq \eta|B|/2$. Hence after t arbitrary modifications we would still have

$$|\widehat{B}_t(r)| \ < \ \eta|B|$$

provided that $2t < \eta|B|/2$. This holds by a huge margin because t is so small, and so we come to the following key realisation. Let R be the set of $r \in \mathbb{Z}_N \setminus 0$ for which $|\widehat{B}(r)| \geq \eta|B|/2$. If we can find $h_1, \dots, h_t$ such that $\|rh_j\|$ is small for all $r \in R$, and for which the modified set B_t is a subset of A, then we will have violated the extremality of B.

Thus we are interested in H, the set of all h for which $\|rh\|$ is small for all $r \in R$. At this point we invoke Theorem 1, which tells us that R lies in $\mathrm{Span}(E)$ for some set E of cardinality $\ll \eta^{-2}\log(1/\beta)$. Note that in this setting Theorem 1 is extremely powerful as β is so small; a straightforward application of Parseval's theorem would be hopeless. Using a classical application of the pigeonhole principle due to Dirichlet it is easy to see that there is an arithmetic progression P of length $\sim N^{c/\log(1/\beta)}$ such that $\|eh\|$ is small for all $e \in E$ and $h \in P$. The fact that $R \subseteq \mathrm{Span}(E)$ tells us that $P \subseteq H$, provided that the different occurences of the word "small" are replaced by appropriate numerical values.

The above shows that any set B_t formed by replacing x_j with $x_j + h_j$ $(j = 1, \dots, t)$, $h_j \in P$, has $\sup_{r \neq 0} |\widehat{B}_t(r)| < \eta|B|$. By the extremal property of B, there can be no choice of the h_j for which $B_t \subseteq A$. Roughly speaking it seems reasonable that this can only be the case if some progression $x_j + P$ has very small intersection with A, which in turn forces A° to contain a long AP. Turning this into a rigourous statement requires a couple of further tricks, for which we refer the reader to [7]. Details aside, we have completed the proof of Proposition 3 and hence of Theorem 3. □

Let us make a few remarks about the use of Theorem 1 here. We wanted to say something about the set H of all $h \in \mathbb{Z}_N$ such that $\|rh\| \leq \varepsilon$, say, for all $r \in R$. Such a set is usually called a *Bohr neighbourhood* and denoted by $B(R, \varepsilon)$ in honour of mathematician and Danish football legend Harald Bohr. By Dirichlet's argument such a set will contain an AP of length at least $\varepsilon N^{1/|R|}$. Suppose, however, we know that $R \subseteq \mathrm{Span}(E)$. Then it is easy to see that

$$B\left(E, \varepsilon/|E|\right) \ \subseteq \ B(R, \varepsilon),$$

so that $B(R, \varepsilon)$ contains an AP of length at least $\varepsilon N^{1/|E|}/|E|$. If $|E|$ is much less than $|R|$ then this represents a significant improvement. Chang's theorem, as applied to the proof of Theorem 2, put us in exactly such a situation. A similar situation arises in the proof [5] of

Freiman's theorem for which Theorem 1 was originally intended.

7. Miscellaneous further remarks. To conclude this article we collect together a number of items related to what we have discussed.

i. Chang's theorem is sharp. Let us begin by mentioning that Chang's theorem is in a sense best possible. The following theorem from [8] illustrates this (the reader may care to recall the definition of ρ-large):

Theorem 3 (Chang's theorem is sharp) *Let α, ρ be positive real numbers satisfying $\alpha \leq 1/8$, $\rho \leq 1/32$ and*

$$\rho^{-2} \log(1/\alpha) \ \leq \ \frac{\log N}{\log \log N}. \tag{14}$$

Then there is a set $A \subseteq \mathbb{Z}_N$ with $|A| = \lfloor \alpha N \rfloor$ which is ρ-large at R, where R is not contained in $\mathrm{Span}(E)$ for any set E with $|E| \leq 2^{-12}\rho^{-2}\log(1/\alpha)$.

ii. Sumsets in $\mathbb{F}_2^n$. It is becoming increasingly apparent that for many problems concerning integers it is advantageous to start by thinking about the corresponding problems in $\mathbb{F}_2^n$, where arguments are typically much cleaner. Example of this are Freiman's theorem (compare [5] with [17]) and Roth's theorem on 3-term arithmetic progressions (here one should work in $\mathbb{F}_3^n$). The same is true of the problem of locating structures in sumsets which we considered in §6. Indeed, one can adapt the method of [7] as described in that section to prove the following result about sumsets in $\mathbb{F}_2^n$.

Theorem 4 *Suppose that γ, δ are real numbers with $\gamma\delta \geq 1/\sqrt{n}$, and that C, D are subsets of $\mathbb{F}_2^n$ with cardinalities γN and δN, where $N = 2^n$. Then $C + D$ contains a translate of some subspace of $\mathbb{F}_2^n$ having dimension at least $\gamma\delta n/80$.*

The details were given in [9] (3). In place of Rudin's inequality one may use a celebrated inequality of Beckner [2].

iii. Upper bounds on progressions in sumsets. We have not yet said anything about whether Theorem 2 is at all sharp. In other words, might it be the case that if C, D are large subsets of $\mathbb{Z}_N$ then $C + D$ contains an arithmetic progression of length substantially longer than $e^{\sqrt{\log N}}$? A curious feature of this problem, which makes it different from many problems in combinatorics, is that the extremal examples are neither random nor particularly regular. Indeed, if C, D are large random subsets of $\mathbb{Z}_N$ then $C + D = \mathbb{Z}_N$, whereas if C, D are arithmetic progressions then obviously $C + D$ also contains a long progression. What is needed is something rather different, and this was provided by Ruzsa [15].

Proposition 4 (Ruzsa) *For any $\varepsilon > 0$ there is a set $C \subseteq \mathbb{Z}_N$ with $|C| > \left(\frac{1}{2} - \varepsilon\right) N$ but such that $C + C$ does not contain an AP of length $e^{(\log N)^{2/3+\varepsilon}}$.*

Ruzsa calls his examples *niveau sets*. After learning about Proposition 4 one might think that Theorem 2 is not far short of the truth. My opinion is that this is somewhat of an illusion. The reason for this is that Ruzsa's argument, when adapted to $\mathbb{F}_2^n$, gives a bound which differs substantially from that implied by Theorem 4. Let us give the argument here, because it takes a very simple form[5].

Proposition 5 (Niveau sets in $\mathbb{F}_2^n$) *There is a set $C \subseteq \mathbb{F}_2^n$ with $|C| > N/4$, but such that $C + C$ does not contain a translate of any subspace with dimension more than $n - \sqrt{n}$.*

Proof. Let C be the set of all vectors $x \in \mathbb{F}_2^n$ with at least $n/2 + \sqrt{n}/2$ ones with respect to the standard basis. By the central limit theorem the number of ones in a random vector $(x_1, \ldots, x_n)$ is roughly normally distributed with mean $n/2$ and standard deviation $\sqrt{n}/2$, and so for large n we have $\gamma \geq 1/4$. Now any vector $x \in C + C$ must have at least $\sqrt{n}$ zeros. Using this fact, we shall prove that $C + C$ meets all translates of all $(n - \lfloor \sqrt{n} \rfloor)$-dimensional subspaces. Indeed, write $d = \lfloor \sqrt{n} \rfloor$ and suppose that U is a translate of some subspace of dimension $n = d$. U can be written as

$$U \ = \ \{a_0 + \lambda_1 a_1 + \cdots + \lambda_{n-d} a_{n-d} \ : \ \lambda_i \in \mathbb{F}_2\}\,,$$

where the a_i are linearly independent. Write a_i in component form as $(a_i^{(j)})_{j=1}^n$. The column rank of the matrix (a_{ij}) is $n - d$, and hence so is the row rank. Without loss of generality, suppose that the first $n - d$ rows $(a_1^{(j)}, \ldots, a_{n-d}^{(j)})$, $j = 1, \ldots, n - d$, are linearly independent. Then we can solve the $n - d$ equations

$$a_0^{(j)} + \lambda_1 a_1^{(j)} + \cdots + \lambda_{n-d} a_{n-d}^{(j)} \ = \ 1$$

for the λ_i, giving a vector in U with no more than d zeros.

iv. Spectral structure of large sets. Chang's theorem tell's us nothing useful about the structure of R, the set of points at which $|\widehat{A}(r)| \geq \rho |A|$, where $A \subseteq \mathbb{Z}_N$ has cardinality $\lfloor N/2 \rfloor$. It turns out that in fact nothing can be said over and above the trivial bound $|R| \ll \rho^{-2}$ coming from Parseval's identity. A result in this direction was proved in [8], but it later came to the author's attention that better results follow from earlier approaches of de Leeuw, Kahane and Katznelson [6] and Nazarov [13]. Nazarov's argument is nicely described in the article [1], from which one can extract the following result.

Theorem 5 (Nazarov) *Let α_r, $r \in \mathbb{Z}_N$, be positive reals satisfying $\sum_r \alpha_r^2 \leq N/1600$. Then there is a function $f : \mathbb{Z}_N \to [0, 1]$ such that $|f| = \sum_x f(x) = N/2$, and so that*

$$|\widehat{f}(r)| \ \geq \ \alpha_r |f|$$

for all $r \in \mathbb{Z}_N$.

[5]Ruzsa's paper is, by contrast, something of a *tour de force*. This is because in $\mathbb{Z}_N$ one does not have the notion of independence, and one is forced to work with the weaker notion of dissociativity instead. In the context of these constructions, this presents a significant barrier.

Roughly speaking, this beautiful result says that the only information one can infer concerning the large spectrum of a large subset of $\mathbb{Z}_N$ comes directly from Parseval's theorem. [6]

v. L^1-norms of exponential sums. Lemma 1 introduced the notion of a set supporting a function whose Fourier transform has small L^1-norm. In this section we mention a couple of open problems relating to sets such that the Fourier transform of the set itself has small L^1 norm.

Problem 1 (Strong Littlewood conjecture) *Let $\lambda_1, \ldots, \lambda_N$ be distinct positive integers. Is it true that*

$$\left\| \sum_{n=1}^{N} e(\lambda_n \theta) \right\|_1 \geq \left\| \sum_{n=1}^{N} e(n\theta) \right\|_1 \quad ?$$

It was shown that this holds if one replaces $\geq$ by $\gg$. This result, proved independently by Konyagin [11] and McGehee, Pigno and Smith [12], solved a famous conjecture which had been known as Littlewood's conjecture.

Problem 2 (Chowla's cosine problem) *Let $\lambda_1, \ldots, \lambda_N$ be distinct positive integers. How large can*

$$\min_{\theta \in [0,1)} \sum_{n=1}^{N} e(\lambda_n \theta)$$

be?

Improving an approach of Bourgain [3], Ruzsa [18] has shown that the minimum cannot be greater than $-e^{-c\sqrt{\log N}}$. The truth may be more like $-\sqrt{N}$.

References

[1] Ball, K, *Convex geometry and functional analysis*, Handbook of the geometry of Banach spaces, Vol. I, 161–194, North-Holland, Amsterdam, 2001.

[2] Beckner, W, *Inequalities in Fourier analysis*. Ann. Math(2) **102** (1975), no. 1, 159–182.

[3] Bourgain, J, *Sur le minimum d'une somme de cosinus*, Acta Arith. **45** (1986), no. 4, 381–389.

[6] Admittedly, the function f is not quite the characteristic function of a set. One can use f to create a set A by picking each $x \in \mathbb{Z}_N$ to lie in A independently at random with probability $f(x)$. A theorem of Spencer [19] then allows one to conclude that $\widehat{A}$ is very similar to $\widehat{f}$. However, I do not see how to get Theorem 5 with f equal to the characteristic function of a set using only this observation.

[4] Bourgain, J, *On arithmetic progressions in sums of sets of integers*, in *A tribute to Paul Erdős*, CUP 1990.

[5] Chang, M.C, *Polynomial bounds for Freimans's theorem*, Duke Math. J. **113** (2002), no. 3., 399–419.

[6] de Leeuw, K; Katznelson, Y; Kahane, J. P. *Sur les coefficients de Fourier des fonctions continues*, C. R. Acad. Sci. Paris Sér. A-B **285** (1977), no. 16, A1001–A1003.

[7] Green, B.J, *Arithmetic progressions in sumsets*, GAFA **12** (2002) 584–597.

[8] Green, B.J, *Some constructions in the inverse spectral theory of cyclic groups*, Comb. Prob. Comp. **12** (2003), no. 2, 127–138.

[9] Green, B.J, *Various unpublished lecture notes*,
1. Large deviation inequalities
2. Edinburgh lecture notes on Freiman's theorem
3. Restriction and Kakeya phenomena (Part III course, Cambridge 2002).

[10] López, J. M. and Ross, K. A, *Sidon sets*, Lecture Notes in Pure and Applied Mathematics, Vol. 13. Marcel Dekker, Inc., New York, 1975.

[11] Konyagin, S.V, *On the Littlewood problem*, Izv. Akad. Nauk. SSSR Ser. Mat. **45** (1981), no. 2, 243–265, 463.

[12] McGehee, O.C, Pigno, L, and Smith, B, *Hardy's inequality and the L^1-norm of exponential sums*, Ann. Math (2) **113** (1981), no. 3., 613–618.

[13] Nazarov, F. L, *The Bang solution of the coefficient problem*, Algebra i Analiz **9** (1997), no. 2, 272–287; English translation in St. Petersburg Math. J. **9** (1998), no. 2, 407–419.

[14] Rudin, W, *Fourier analysis on groups*, Wiley 1990 (reprint of the 1962 original).

[15] Ruzsa, I.Z, *Arithmetic progressions in sumsets*, Acta Arith. **60** (1991), no.2, 191 – 202.

[16] Ruzsa, I.Z, *Connections between the uniform distribution of a sequence and its differences*, Topics in classical number theory, Vol. I, II (Budapest, 1981), 1419–1443, Colloq. Math. Soc. János Bolyai, **34**, North-Holland, Amsterdam, 1984.

[17] Ruzsa, I.Z, *An analog of Freiman's theorem in groups*, Structure theory of set addition, Astérisque **258** (1999), 323–326.

[18] Ruzsa, I.Z, *On negative values of cosine sums*, submitted to Acta Arithmetica.

[19] Spencer, J, *Six standard deviations suffice*, Trans. Amer. Math. Soc. **289** (1985), no. 2, 679–706.