On Arithmetic Progressions in Sums of Sets of Integers

Jean Bourgain

An exposition by Ben Green, May 2000

Contents

1 Introduction

This is an exposition of yet another brilliant paper by Bourgain, "Arithmetic Progressions in Sums of Sets of Integers" [1]. In this amazingly short paper Bourgain proves the following result.

Theorem 1 *Let $A, B \subseteq \{1, \dots, N\}$ be sets with $|A| = \alpha N$ and $|B| = \beta N$. Then there is a constant $c = c(\alpha, \beta)$ such that the sumset $A + B$ contains an arithmetic progression of length $e^{c(\log n)^{1/3}}$.*

The result itself is clearly very interesting, but more interesting still is the method of proof. Much inspiration is taken from the theory of thin sets in harmonic analysis, and indeed it is almost impossible to understand this paper without delving into some texts on that subject such as [3] or [4]. Even then the necessary results can only be extracted with some difficulty, and one of our main objectives in writing this exposition is to give an account accessible to those with a background in additive number theory.

We begin by giving an overview of the proof of Theorem 1. The first thing to appreciate is the fundamental difference between this result and the result of Roth on arithmetic progressions of length 3. To prove Roth's result the basic idea is to pass to a subprogression on which our set has increased density, and then iterate the argument. Such an approach cannot work here for the very obvious reason that a subset of $A + B$ is unlikely to be a set of form $A' + B'$, and so it is hard to see what an inductive hypothesis might be. Instead we must locate a long arithmetic progression straight away.

It is possible to conceive of a variety of ways in which we might find such a progression, and Bourgain uses the following clever approach. We begin by making the problem modular, so that harmonic analysis can be done more easily. To do this regard A and B as subsets of $\mathbb{Z}_p$, where $p > 2N$ is a prime number. Let $f(x) = A * B(x)$, where we are using convolutions in the conventional way

(so that $A * B(x) = \sum_y A(y)B(x-y) = |(a,b) \in A \times B : a+b=x|$). Observe that, because of the choice of p, the modular version of f is the same as the $\mathbb{Z}$-version. Let K be a positive integer. If we could find x, y for which

$$\max_{0<k<K} |f(x+ky) - f(x)| \ < \ f(x),$$

then $f(x+ky)$ would be positive for all $x = 0, 1, \ldots, K-1$ and we would have a modular arithmetic progression of length K lying in $A+B$. It is easy to see that we can find a genuine (i.e. $\mathbb{Z}$-) progression of length $\frac{1}{2}\sqrt{K}$ from this.

How are we going to find such an x and y? It seems reasonable that we are going to need to look at the arithmetic structure of $A+B$ more closely to find y, but we might well get away with averaging over x. That is to say, we shall look for y such that

$$\sum_x \max_{0<k<K} |f(x+ky) - f(x)| \ < \ \sum_x f(x).$$

If we can show, for $K \sim e^{c(\log n)^{1/3}}$, that such a y can be found then Theorem 1 will be proved.

As everyone knows, one of the best ways of isolating arithmetic information about a function is by looking at its Fourier coefficients. This is particularly true when we are looking for some sort of bias along an arithmetic progression as is the case here. Examples of such techniques are the classical result of Roth mentioned earlier, and the improvements of that result due to Heath-Brown [2] and Szemerédi [6]. In fact, readers familiar with these works will notice a number of similarities with the current paper.

Roth, Heath-Brown and Szemerédi all look for large Fourier Coefficents $\hat{f}(r)(r \in R)$ and then procced to show that there is a bias of f along a progression with common difference d, where $rd \pmod p$ is small for all $r \in R$. The main difference between the approaches is that Roth only considers the case $|R| = 1$. Such an approach is of course extremely natural because *if* f is the characteristic function of a small arithmetic progression with common difference d then $\hat{f}(r)$ is large precisely when $rd \pmod p$ is small.

Bourgain goes a step further. Suppose that the large Fourier Coefficients of f are $\hat{f}(r)(r \in R)$ (for some suitable definition of "large"). Suppose for a moment that f is the characteristic function of a short AP with common difference d. Then R is the set of r for which $rd \pmod p$ is small, and is itself an arithmetic progression. One can concieve of making the jump from this observation to the following rough statement, where f is now $A * B$ again.

$(*)$ If we are looking for a bias of f along an arithmetic progression, then the "relevant" large Fourier coefficients $\hat{f}(r)$ ought to be those with r lying in some set R which has lots of structure.

What Bourgain does, then, is take all the large Fourier coefficients of f, and then isolate a highly structured subset of them. This subset will be used to find an appropriate arithmetic progression on which f is biased. This procedure is carried out in such a way that the other large Fourier coefficients are extremely unstructured, and their contribution has the status of an error term. It is

in dealing with the unstructured contribution that we must appeal to the ideas of [3] and [4]. The exact notion of "unstructured" that we shall use is the concept of *dissociativity*, which we discuss at length now.

2 Dissociated Sets

At this point we take a fairly lengthy detour into the world of dissociated sets. Fix an integer N. Then a set $E \subseteq \mathbb{Z}_N$ is said to be *dissociated* if no x has more than one representation as $\sum_i \epsilon_i x_i$, where the x_i are distinct elements of E and $\epsilon_i \in \{-1, 1\}$. Although this is a purely additive number theory definition, the concept of dissociated set arises quite naturally in the study of sparse trigonometric series. See, for example, [3] (Chapter 5) or [4].

We note that [1] has a slightly weaker notion of dissociated set (that is to say not all of Bourgain's dissociated sets are dissociated in our sense). We have not been able to prove the results that follow for this weaker notion. However it turns out that a small modification to a later part of [1] enables one to get away with our rather more restrictive definition.

In the following we work in $\mathbb{Z}_N$. In particular the hat symbol refers to Fourier transforms in that group, so that if $f : \mathbb{Z} \to \mathbb{C}$ is a function then we set

$$\hat{f}(r) \;=\; \sum_x f(x)\omega^{-rx},$$

where $\omega = e^{2\pi i/N}$.

From now on we fix a dissociated set $E \subseteq \mathbb{Z}_N$. Let $f : \mathbb{Z} \to \mathbb{C}$ be a function supported on E, and for any positive integer s write

$$f^{\oplus s}(x) \;=\; \sum_{\substack{(x_1,\ldots,x_s)\in E^s \\ x_1+x_2+\cdots+x_s=x}} f(x_1)\ldots f(x_s)$$

for the s-fold convolution of f with itself. It is easy to check that the Fourier transform $(f*g)\hat{}(r)$ is simply $\hat{f}(r)\hat{g}(r)$, and hence that $f^{\oplus s}$ has transform $\hat{f}(r)^s$. Now suppose it were the case that every $x \in \mathbb{Z}_N$ had at most one representation in the form $x_1 + \cdots + x_s$ ($x_i \in E$), at least up to reordering of the summands. Then we would have, using Parseval's identity,

$$\begin{aligned}
\|\hat{f}\|_{2s}^{2s} &= \sum_x \left|f^{\oplus s}(x)\right|^2 \\
&\leq s! \sum_x \sum_{\substack{(x_1,\ldots,x_s)\in E^s \\ x_1+x_2+\cdots+x_s=x}} |f(x_1)|^2\ldots|f(x_s)|^2 \\
&= s!\|\hat{f}\|_2^{2s}.
\end{aligned}$$

This immediately gives the rather strong-looking inequality

$$\|\hat{f}\|_{2s} \;\leq\; \sqrt{s}\|\hat{f}\|_2. \qquad (1)$$

Alas that derivation was based on the (in general) false supposition that no x has two essentially distinct representations in the form $x_1+\cdots+x_s$. Such a statement does not follow from the fact that E is dissociated, because these representations are allowed to have repeated summands, a situation which is not covered by the definition of dissociativity. One still feels, however, that a statement similar to (1) ought to be true. A reason for this is that dissociativity implies that no x can have two different "good" representations as $x_1 + \cdots + x_s$ with the x_i all distinct, but if E is at all large then most representations are good.

A possible approach, then, would be to prove that the contribution to $\|\hat{f}\|_{2s}$ from "bad" representations is small. I cannot make such an approach give more than the weaker bound $\|\hat{f}\|_{2s} \ll s\|\hat{f}\|_2$ and I believe that there is a good reason for this, namely that it is hard to make such an approach use any of the information about sums $x_1 + \cdots + x_r$ $(r > s)$ that dissociativity gives us. A similar situation is considered in Ruzsa [5], pages 270–271. In that paper Ruzsa considers sets in which no integer has two representations as $x_1 + \cdots + x_s$ with $x_1, \ldots, x_s$ distinct and s fixed. There is no interesting information about sums $x_1 + \cdots + x_r$ with $r > s$, and Ruzsa gets a bound of shape $\|\hat{f}\|_{2s} \ll s\|\hat{f}\|_2$. It is interesting to observe that such a bound still seems to give non-trivial information along the lines of Theorem 1, but with the exponent 1/3 replaced by 1/4.

We are going to prove that a considerably cleverer approach will yield an inequality comparable to (1).

Theorem 2 *Let $E \subseteq \mathbb{Z}_N$ be dissociated. Let $f : \mathbb{Z}_N \to \mathbb{C}$ be supported on E. Then we have, for all real numbers $p \geq 2$, the inequality*

$$\|\hat{f}\|_p \ \leq \ 2\sqrt{p}\|\hat{f}\|_2.$$

In [1] there appears a similar inequality in which the factor 2 is replaced by 10. This would suggest to me that one can work with Bourgain's weaker definition of dissociativity, but that I have not been clever enough to do so. The first step towards a proof of Theorem 2 is the following standard Lemma.

Lemma 3 (Young's Inequality) *Let m be a positive integer, and let $f, g : \mathbb{Z}_N \to \mathbb{C}$ be two functions. Then*

$$\|f * g\|_m \ \leq \ N\|f\|_1\|g\|_m.$$

Proof We have

$$\begin{aligned}
N\|f*g\|_m &= \sum_x \left|\sum_y f(y)g(x-y)\right|^m \\
&\leq \sum_x \sum_{y_1} \cdots \sum_{y_m} |f(y_1)|\dots|f(y_m)||g(x-y_1)|\dots|g(x-y_m)| \\
&= \sum_{y_1} \cdots \sum_{y_m} |f(y_1)|\dots|f(y_m)| \sum_x |g(x-y_1)|\dots|g(x-y_m)| \\
&\leq \sum_{y_1} \cdots \sum_{y_m} |f(y_1)|\dots|f(y_m)| \left(\sum_x |g(x-y_1)|^m\right)^{1/m} \dots \left(\sum_x |g(x-y_m)|^m\right)^{1/m} \\
&= N^m\|f\|_1^m \cdot N\|g\|_m^m.
\end{aligned}$$

The result follows immediately. □

Now we come to the key part of the proof of Theorem 2. Our strategy will be as follows. We shall consider "twists" of f by functions $\epsilon : E \to \mathbb{T}$. Writing $f_\epsilon(x) = f(x)\epsilon(x)$ we shall prove in Proposition 4 that, for any positive integer m, one has $\|\hat{f}\|_m \leq 2\|\hat{f}_\epsilon\|_m$. This is a rather surprising result which depends heavily on the dissociativity of E. Next, in Proposition 6, we shall show that one can choose ϵ such that the function f_ϵ is well-behaved and, in particular, satisfies $\|\hat{f}_\epsilon\|_m \ll \sqrt{m}\|\hat{f}_\epsilon\|_2 = \sqrt{m}\|\hat{f}\|_2$. This is considerably less surprising because one feels that ϵ can be chosen so that contributions from the "bad" representations cancel out.

Proposition 4 *Let m be a positive integer, and let $\epsilon : E \to \mathbb{T}$. Define $f_\epsilon(x) = f(x)\epsilon(x)$ as above. Then*

$$\|\hat{f}\|_m \leq 2\|\hat{f}_\epsilon\|_m.$$

Proof It is far from clear to me how one is supposed to come up with a proof like this. Define

$$p_\epsilon(x) \;=\; 2\prod_{r\in E}\left(\tfrac{1}{2}\overline{\epsilon(r)}\omega^{-rx} + 1 + \tfrac{1}{2}\epsilon(r)\omega^{rx}\right).$$

This object is, as the reader may appreciate, some kind of Riesz product. The first thing to observe is that p_ϵ is a product of non-negative real factors, and so $\|p_\epsilon\|_1$ is just $\frac{1}{N}\sum_x p_\epsilon(x)$. If one expands out the product for p_ϵ in full one gets a sum of terms of the form

$$2^{1-k}\epsilon(r_1)\dots\epsilon(r_l)\overline{\epsilon(r_{l+1})}\dots\overline{\epsilon(r_k)}\omega^{(r_1+\dots+r_l-r_{l+1}-\dots-r_k)x},$$

where the r_i are distinct. When one sums over x all such terms disappear except those with $r_1+\dots+r_l-r_{l+1}-\dots-r_k = 0$. Since E is dissociated there is only one such term, corresponding to $k = l = 0$. It follows immediately that $\|p_\epsilon\|_1 = 2$.

Very similar considerations allow us to find $\hat{p_\epsilon}(r)$ for certain r. Suppose that $r \in E$. Then $\hat{p_\epsilon}(r)$ is a sum of terms of the form

$$2^{1-k}\epsilon(r_1)\dots\epsilon(r_l)\overline{\epsilon(r_{l+1})}\dots\overline{\epsilon(r_k)}\omega^{(r_1+\dots+r_l-r_{l+1}-\dots-r_k-r)x}.$$

When we sum over x the only terms that remain are those with $r_1+\cdots+r_l-r_{l+1}-\cdots-r_k=r$. The dissociativity of E tells us that this can only happen if $l=k=1$ and $r_1=r$, and it follows quickly that $\hat{p_\epsilon}(r)=N\epsilon(r)$. This information can be used to establish the following interesting result.

Claim 5 *We have, for all r,*

$$\hat{f_\epsilon} * p_\epsilon \;=\; N\hat{f}.$$

Proof We check this by taking Fourier transforms of both sides. The Fourier transfrom of $\hat{f_\epsilon} * p_\epsilon$ is $\hat{\hat{f}}_\epsilon(x)\hat{p_\epsilon}(x)$, and it is easy to check that

$$\hat{\hat{f}}_\epsilon(x) \;=\; Nf(-x)\epsilon(-x).$$

This is zero unless $x \in -E$, and for such values we can say what $\hat{p_\epsilon}$ is: indeed since p_ϵ is real we have, when $x \in -E$, that

$$\hat{p_\epsilon}(x) \;=\; \overline{\hat{p_\epsilon}(-x)} \;=\; N\overline{\epsilon(-x)}.$$

Putting these remarks together gives

$$(\hat{f_\epsilon} * p_\epsilon)\hat{}\,(x) \;=\; N^2 f(-x)$$

for all x. It is a simple matter to check that this equals the Fourier transform of $N\hat{f}$, thereby establishing the claim. □

It is now an easy matter to prove Proposition 4, for using Young's Inequality we have

$$N\|\hat{f}\|_m \;=\; \|\hat{f_\epsilon} * p_\epsilon\|_m \;\leq\; N\|\hat{f_\epsilon}\|_m\|p_\epsilon\|_1 \;=\; 2N\|\hat{f_\epsilon}\|_m.$$

We turn now to the issue of finding an $\epsilon : E \to \mathbb{T}$ for which $\|\hat{f_\epsilon}\|_m$ is not large, and we hope the reader will agree that it would be a surprise if we were not to do this by averaging over a suitable set of ϵ. We present two arguments of which the second is a little more complicated, but contains some interesting points.

Proposition 6 *Let $\Omega = \mathbb{T}^{|E|}$ be the group of all possible maps ϵ. Let Ω have normalised Haar measure μ, and let s be a positive integer. Then*

$$\int_\Omega \|\hat{f_\epsilon}\|_{2s}^{2s}\, d\mu \;\leq\; s!\|\hat{f}\|_2^{2s}.$$

Proof We have

$$|\hat{f_\epsilon}(x)|^s \;=\; \sum_m \left(\sum_{r_1+\cdots+r_s=m} f(r_1)\dots f(r_s)\epsilon(r_1)\dots\epsilon(r_s) \right) \omega^{-mx}$$

and so, by Parseval's identity,

$$\begin{aligned} \|\hat{f_\epsilon}\|_{2s}^{2s} &= \sum_m \left| \sum_{r_1+\cdots+r_s=m} f(r_1)\dots f(r_s)\epsilon(r_1)\dots\epsilon(r_s) \right|^2 \\ &= \sum_m \sum_{r_1+\cdots+r_s=m} \sum_{t_1+\cdots+t_s=m} f(r_1)\dots f(r_s)\overline{f(t_1)}\dots\overline{f(t_s)}\epsilon(r_1)\dots\epsilon(r_s)\overline{\epsilon(t_1)}\dots\overline{\epsilon(t_s)}. \end{aligned}$$

Now it is quite easy to check that if $r_1, \dots, r_s$ and $t_1, \dots, t_s$ are fixed then

$$\int_\Omega \epsilon(r_1) \dots \epsilon(r_s) \overline{\epsilon(t_1)} \dots \overline{\epsilon(t_s)} \, d\mu$$

is zero unless the r_i are a rearrangement of the t_i, in which case it equals 1. It follows from (2) and the Cauchy-Schwarz inequality that

$$\begin{aligned} \int_\Omega \|\hat{f}_\epsilon\|_{2s}^{2s} \, d\mu &\leq s! \sum_m \sum_{r_1 + \dots + r_s = m} |f(r_1)|^2 \dots |f(r_s)|^2 \\ &= s! \|\hat{f}\|_2^{2s}. \end{aligned}$$

This completes the proof of the proposition. □

In a moment we shall give an alternative derivation of something like Proposition 6, in which it is only necessary to consider maps $\epsilon : E \to \{-1, 1\}$. First, however, we complete the proof of Theorem 2. It is immediate from Propositions 4 and 6 that, for any positive integer s, we have

$$\|\hat{f}\|_{2s} \leq 2(s!)^{1/2s} \|\hat{f}\|_2 \leq 2\sqrt{s} \|\hat{f}\|_2.$$

Now if $p \geq 2$ is any real number then we certainly have $\| \ \|_p \leq \| \ \|_{2s}$ where $s = \lceil p/2 \rceil \leq (p+2)/2$. Hence

$$\|\hat{f}\|_p \leq 2\sqrt{\frac{p+2}{2}} \|\hat{f}\|_2 \leq 2\sqrt{p} \|\hat{f}\|_2.$$

This completes the proof of Theorem 2. □

We now give an alternative proof of Proposition 6 using only maps $\epsilon : E \to \{-1, 1\}$. This proof comes from the same circle of ideas as Khintchin's Inequality, as discussed in [7]. Let Ω be the probability space consisting of all $\epsilon : E \to \{-1, 1\}$, chosen with equal probability. Our aim is to show that $\mathbb{E}\|\hat{f}_\epsilon\|_p$ is not too large. If we try to copy the above proof it becomes necessary to show that

$$\mathbb{E}\epsilon(r_1) \dots \epsilon(r_s) \overline{\epsilon(t_1)} \dots \overline{\epsilon(t_s)} = 0$$

unless the r_i are a rearrangement of the t_i. Unfortunately this is simply false, and we must proceed in a different way. Let X_r denote the random variable

$$X_r(\epsilon) = \sum_x f(x) \epsilon(x) \omega^{-rx},$$

so that

$$\|\hat{f}_\epsilon\|_p = \left(\frac{\sum_r X_r^p}{N} \right)^{1/p}. \tag{2}$$

Now X_r is itself a sum of N independent random variables $Y_i^{(r)}$, where $Y_i^{(r)}$ takes the values $\pm f(x)\omega^{-rx}$ with equal probability. In such a situation the following inequality of Hoeffding tells us that X_r is strongly concentrated about its mean, 0.

Theorem 7 (Hoeffding's Inequality) *Let $Z_1, Z_2, \dots, Z_n$ be independent random variables such that $a_i \le Z_i \le b_i$ for all i and some real numbers a_i, b_i. Write $\overline{Z} = (Z_1 + \cdots + Z_n)/n$ and $\mu = \mathbb{E}\overline{Z}$. Let $t > 0$. Then*

$$\mathbb{P}\left(\overline{Z} - \mu \ge t\right) \ \le \ e^{-2n^2t^2/\sum_i (b_i - a_i)^2}.$$

Applying this in the present situation gives, using Parseval's Identity,

$$\mathbb{P}\left(|X_r| > \lambda\right) \ \le \ e^{\lambda^2/2\|\hat{f}\|_2^2}.$$

It follows immediately that

$$\mathbb{P}\left(|X_r|^p > \lambda\right) \ \le \ 2e^{-\lambda^{2/p}/\|\hat{f}\|_2^2},$$

and so

$$\begin{aligned}
\mathbb{E}|X_r|^p &= \int_0^\infty \mathbb{P}\left(|X_r|^p > \lambda\right)\, d\lambda \\
&\le 2\int_0^\infty e^{-\lambda^{2/p}/\|\hat{f}\|_2^2}\, d\lambda \\
&= 2\|\hat{f}\|_2^p \cdot \frac{p}{2}\int_0^\infty u^{\frac{p}{2}-1}e^{-u}\, du \\
&= p\Gamma(p/2)\|\hat{f}\|_2^p.
\end{aligned}$$

From (2) we have that $\mathbb{E}\|\hat{f}_\epsilon\|_p^p \le p\Gamma(p/2)\|\hat{f}\|_2^p$. This is, as claimed, rather similar to Proposition 6.

In [1] Theorem 2 is applied through the following corollary. Before stating and proving this, we observe that Theorem 2 has a rather straightforward dual formulation. This states that if $f : \mathbb{Z}_N \to \mathbb{C}$ is a function with $\operatorname{Supp}\hat{f} \subseteq E$, where E is dissociated, then

$$\|f\|_p \ \le \ 2\sqrt{p}\|f\|_2$$

for all real numbers $p \ge 2$. We shall call functions of this form "Fourier-dissociated".

Corollary 8 *Let $f : \mathbb{Z}_N \to \mathbb{C}$ be Fourier-dissociated, and let I be a subset of $\mathbb{Z}_N$. For $i \in I$ denote by f_i the translate of f by i, so that $f_i(x) = f(i + x)$. Then*

$$\left\|\max_{i\in I}|f_i|\right\|_2 \ \le \ 5\sqrt{\log|I|}\|f\|_2.$$

Proof This is best done without words. We have, for any $p \geq 2$,

$$\begin{aligned}
\left\| \max_{i \in I} |f_i| \right\|_2 &= \frac{1}{N} \sum_x \max_{i \in I} |f(x+i)|^2 \\
&\leq \frac{1}{N} \sum_x \left(\sum_{i \in I} |f(x+i)|^p \right)^{2/p} \\
&\leq N^{-2/p} \left(\sum_x \sum_{i \in I} |f(x+i)|^p \right)^{2/p} \\
&= |I|^{2/p} \|f\|_p^2 \\
&\leq 4p|I|^{2/p} \|f\|_2^2.
\end{aligned}$$

Take $p = 2\log|I|$ and we get that this is at most $8e \log|I| \|f\|_2^2$, which concludes the proof. □

If the reader desires a few words on this then we might add that considering $\mathcal{L}^p$ norms for large p is an obvious tactic given that we are interested in an $\mathcal{L}^\infty$ object (i.e. $\max_{i \in I} |f_i|$). Of more relevance, perhaps, is some explanation of what this corollary is saying. In some vague sense, Fourier-dissociated functions have very little additive structure. Thus it is not unreasonable to expect that their translates should be rather regularly behaved (this would certainly not be the case if f were supported on some arithmetic progression, for example).

3 Proof of Theorem 1

The reader who has survived that exposition of dissociativity may wish, at this juncture, to read the introductory outline once more. Suppose that $A, B \subseteq \mathbb{Z}_N$, where N is prime (N has taken the place of the prime p that we used earlier, because I like my $\mathbb{Z}$'s to be subscripted with N's, and also so that we can keep the notation of §2). Suppose that $|A| = \alpha N$ and $|B| = \beta N$, and let $f = A * B$.

Lemma 9 *We have* $\sum_x f(x) = N^2\alpha\beta$ *and*

$$\sum_r |\hat{f}(r)| \leq N^2 \alpha^{1/2} \beta^{1/2}.$$

Proof The first statement is obvious. The second is an easy consequence of Cauchy-Schwarz and Parseval's identity:

$$\begin{aligned}
\sum_r |\hat{f}(r)| &= \sum_r |\hat{A}(r)||\hat{B}(r)| \\
&\leq \left(\sum_r |\hat{A}(r)|^2 \right)^{1/2} \left(\sum_r |\hat{B}(r)|^2 \right)^{1/2} \\
&= N^2 \alpha^{1/2} \beta^{1/2}.
\end{aligned}$$

That concludes the proof of the lemma. □

Although rather trivial, this lemma is the reason the argument works. Indeed the reader may care to check that we use nothing more about the sumset $A+B$ than that $\sum_r |\hat{f}(r)|$ is small compared to $\|f\|_1$.

Recall from the introduction that we are trying to find $K \sim e^{c(\log n)^{1/3}}$ and $y \in \mathbb{Z}_N$ such that

$$\sum_x \max_{0<k<K} |f(x+ky) - f(x)| \ < \ \sum_x f(x). \tag{3}$$

Since this inequality is homogeneous in f, we can renormalise f and it turns out to be convenient to do so. We shall abandon the notion that $f = A * B$, and work only with the facts that

$$\sum_r |\hat{f}(r)| \ \leq \ 1 \tag{4}$$

and

$$\sum_x f(x) \ = \ \sqrt{\alpha\beta}. \tag{5}$$

Recall that our plan was to identify some structure in the set of large Fourier coefficients $\hat{f}(r)(r \in R)$. As a first step, let ν be an integer to be chosen later and set

$$E_s \ = \ \left\{ r : 2^{-s} \leq |\hat{f}(r)| \leq 2^{-s+1} \right\}$$

for $s = 1, \ldots, \nu$ and

$$E_{\text{small}} \ = \ \left\{ r : |\hat{f}(r)| \leq 2^{-v} \right\}.$$

We have divided up into powers of two because it gives us increased control which turns out to be crucial (and we can be fairly safe in the knowledge that the eventual bound will not be affected too greatly). The next lemma describes our method for dividing an arbitrary set into a highly-structured set and some totally unstructed sets.

Lemma 10 *Let $X \subseteq \mathbb{Z}_N$ and let l be a positive integer. Then we can decompose X as*

$$X \ = \ \bigcup_{i=1}^{m} X^{(i)} \cup X^{(\text{st})},$$

where each $X^{(i)}$ is a dissociated set of size l and $X^{(\text{st})}$ is highly structured. By this we mean that there is a set $H = \{h_1, \ldots, h_k\}$, $k \leq l$, such that all elements of $X^{(\text{st})}$ can be written in the form

$$x \ = \ \sum_{i=1}^{k} \eta_i h_i$$

with $\eta_i \in \{-2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2\}$.

Proof We proceed by induction on l, the result being trivial for $l = 0, 1$. Suppose that we have achieved a decomposition of the required type for some $l \geq 1$. Let $X^{(1)} = \{x_1^{(1)}, \ldots, x_l^{(1)}\}$. We try to find an element of $X \setminus X^{(1)}$ that can be added to $X^{(1)}$ so that the resulting set is still dissociated. The only way in which we could fail to locate such an element would be if everything in $X \setminus X^{(1)}$ could be written in the form

$$\sum_{i=1}^{l} \eta_i x_i^{(1)} \tag{6}$$

with $\eta_i \in \{-2, -1, -\frac{1}{2}, 0, \frac{1}{2}, 1, 2\}$, in which case the result is proved. Indeed suppose that, for every $y \in X \setminus X^{(1)}$, there exists a t with two different representations as $\sum_i \epsilon_i x_i^{(1)} + \epsilon y$ with $\epsilon_i, \epsilon \in \{-1, 1\}$. The ϵ's in these two representations must be different or else dissociativity of $X^{(1)}$ would be contradicted, and this allows us, by eliminating t, to express y in the form (6).

If we succeed in adding an element to $X^{(1)}$ then we try our luck with $X^{(2)}$, and a similar analysis applies. If we succeed in adding an element to each of $X^{(1)}, \ldots, X^{(m)}$ then we are particularly happy, because then we may take the new $X^{(\mathrm{st})}$ to be just what is left of the old one. □

Now we are ready for our grand "cutting" of the Fourier transform. We apply the decomposition of Lemma 10 to each of the sets E_s defined above, getting

$$E_s \;=\; \bigcup_i E_s^{(i)} \cup E_s^{(\mathrm{st})}$$

for $s = 1, \ldots, \nu$.

We use this to decompose f as

$$\begin{aligned} f &= \sum_{s=1}^{\nu} \sum_i f_s^{(i)} + \sum_{s=1}^{\nu} f_s^{(\mathrm{st})} + f_{\mathrm{small}} \\ &= g_1 + g_2 + g_3. \end{aligned} \tag{7}$$

The notation here will be clear if we say that $f_s^{(i)}$ has Fourier coeffiecents $\hat{f}(r)$ for $r \in E_s^{(i)}$, and 0 for all other r. One can check that $f_s^{(i)}$ is given by the formula

$$f_s^{(i)}(x) \;=\; \frac{1}{N} \sum_{r \in E_s^{(i)}} \hat{f}(r) \omega^{rx},$$

and that similar results hold for the other functions featuring in (7). One has

$$\sum_x \max_{0<k<K} |f(x+ky) - f(x)| \;\leq\; \sum_{i=1}^{3} \sum_x \max_{0<k<K} |g_i(x+ky) - g_i(x)|. \tag{8}$$

Our aim is, the reader will recall, to show that y can be chosen so that this is less than $\sum_x f(x) = \sqrt{\alpha\beta}$. We begin by bounding the two "error terms", by which we mean the terms involving g_1 and g_3. These estimates will hold regardless of the value of y; we will then show that y can be chosen to make the term with g_2 small as well.

3.1 Estimation of the g_1 term.

In this subsection we shall prove the following lemma.

Lemma 11

$$\sum_x \max_{0<k<K} |g_1(x+ky) - g_1(x)| \ \leq \ 20\sqrt{\frac{\log K}{l}}.$$

Proof This is the only place in which the results of §2 will be used, and even here we only require Corollary 8. We start by observing that

$$\max_{0<k<K} |g_1(x+ky) - g_1(x)| \ \leq \ 2 \max_{0\leq k<K} |g_1(x+ky)|.$$

However we have

$$\begin{aligned}
\left\| \max_{0\leq k<K} |g_1(x+ky)| \right\|_1 &\leq \sum_{s=1}^{\nu} \sum_i \left\| \max_{0\leq k<K} |f_s^{(i)}(x+ky)| \right\|_1 \\
&\leq \sum_{s=1}^{\nu} \sum_i \left\| \max_{0\leq k<K} |f_s^{(i)}(x+ky)| \right\|_2 \\
&\leq 5\sqrt{\log K} \sum_{s=1}^{\nu} \sum_i \|f_s^{(i)}\|_2 \\
&= \frac{5\sqrt{\log K}}{N} \sum_{s=1}^{\nu} \sum_i \left(\sum_{r\in E_s^{(i)}} |\hat{f}(r)|^2 \right)^{1/2},
\end{aligned}$$

where we have applied Corollary 8 in deriving the third line from the second. However

$$\begin{aligned}
\left(\sum_{r\in E_s^{(i)}} |\hat{f}(r)|^2 \right)^{1/2} &\leq 2^{-s+1}\sqrt{l} \\
&\leq \frac{2}{\sqrt{l}} \sum_{r\in E_s^{(i)}} |\hat{f}(r)|,
\end{aligned}$$

since $|E_s^{(i)}| = l$ and $2^{-s} \leq |\hat{f}(r)| \leq 2^{-s+1}$ for $r \in E_s^{(i)}$. It follows that

$$\begin{aligned}
\max_{0<k<K} |g_1(x+ky) - g_1(x)| &\leq 2N \left\| \max_{0\leq k<K} |g_1(x+ky)| \right\|_1 \\
&\leq 20\sqrt{\frac{\log K}{l}} \sum_{s=1}^{\nu} \sum_i \sum_{r\in E_s^{(i)}} |\hat{f}(r)| \\
&\leq 20\sqrt{\frac{\log K}{l}}.
\end{aligned}$$

This proves the lemma. □

3.2 Estimation of the g_3 term.

Lemma 12

$$\sum_x \max_{0<k<K} |g_3(x+ky) - g_3(x)| \;\leq\; 2\sqrt{2^{-\nu}K}.$$

Proof No words are required. We have

$$\begin{aligned}
\sum_x \max_{0<k<K} |g_3(x+ky) - g_3(x)| &\leq 2\sum_x \max_{0\leq k<K} |g_3(x+ky)| \\
&= \frac{2}{N}\sum_x \max_{0\leq k<K} \left|\sum_{r\in E_{\rm small}} \hat{f}(r)\omega^{r(x+ky)}\right| \\
&\leq \frac{2}{N^{1/2}}\left(\sum_x \max_{0\leq k<K} \left|\sum_{r\in E_{\rm small}} \hat{f}(r)\omega^{r(x+ky)}\right|^2\right)^{1/2} \\
&\leq \frac{2}{N^{1/2}}\left(\sum_x \sum_{0\leq k<K} \left|\sum_{r\in E_{\rm small}} \hat{f}(r)\omega^{r(x+ky)}\right|^2\right)^{1/2} \\
&= 2\sqrt{K}\left(\sum_{r\in E_{\rm small}} |\hat{f}(r)|^2\right)^{1/2} \\
&\leq 2\sqrt{2^{-\nu}K},
\end{aligned}$$

which proves the lemma. □

3.3 Estimation of the g_2 term.

From now on we write $B = \bigcup_{s=1}^{\nu} E_s^{(\rm st)}$, the union of all our highly-structured sets of large Fourier coefficients. We have arrived at what is in some sense the heart of the proof of Theorem 1. The previous two sections have consisted in estimating what are, morally, error terms and our inequalities were valid for all y. We now come to actually picking a value of y which makes

$$\sum_x \max_{0<k<K} |g_2(x+ky) - g_2(x)|$$

small. Our main tool is an argument of Dirichlet on simultaneous approximation, and in this respect Bourgain's paper bears a substantial resemblance to the papers [2] and [6] on progressions of length 3. We start with a chain of straightforward inequalities.

$$\begin{aligned}
\sum_x \max_{0<k<K} |g_2(x+ky) - g_2(x)| &= \frac{1}{N}\sum_x \max_{0<k\leq K} \left|\sum_{n\in B} \hat{f}(n)\left(\omega^{nky}-1\right)\omega^{nx}\right| \\
&\leq \max_{\substack{n\in B\\ 0<k\leq K}} \left|\omega^{nky}-1\right| \\
&\leq K \max_{n\in B} |1-\omega^{ny}|. \qquad (9)
\end{aligned}$$

Now B, being the union of the sets $E_s^{(\mathrm{st})}$, has the following property. Any $b \in B$ can be written as $\sum \eta_i t_i$ where

$$\eta_i \in \{-2, -1, -\tfrac{1}{2}, 0, \tfrac{1}{2}, 1, 2\},$$

at most l of the η_i are non-zero and the t_i come from some set T of size at most $l\nu$. This trivially implies that B also has the following property, which is easier to work with: any $b \in B$ can be written as $\sum \epsilon_i u_i$ where $\epsilon_i \in \{-1, 0, 1\}$, at most l of the ϵ_i are non-zero and the u_i come from a set U of size at most $3l\nu$.

We have then

$$K \max_{n \in B} |1 - \omega^{ny}| \;\leq\; Kl \max_{n \in U} |1 - \omega^{ny}|\,.$$

Let the elements of U be $u_1, \ldots, u_q$ where $q \leq 3l\nu$. An easy application of Dirichlet's argument (which amounts to an application of the pigeonhole principle) shows that if $N > \delta^{-q}$ then there is some y for which the fractional parts $\{u_i y/N\}$, $i = 1, \ldots, q$, are all at most δ in magnitude. For such a y one then has, by an easy computation, that

$$\max_{n \in U} |1 - \omega^{ny}| \;\leq\; 2\pi\delta.$$

Taking $\delta = N^{-1/3l\nu}$ and recalling (9), we have

$$\sum_x \max_{0<k<K} |g_2(x + ky) - g_2(x)| \;\leq\; 7KlN^{-1/3l\nu}$$

for a suitable y.

3.4 Putting everything together.

From (3), (8) and the last three subsections it suffices to show that, for $K \sim e^{c(\log N)^{1/3}}$, there are choices of l and ν for which

$$20\sqrt{\frac{\log K}{l}} + 2\sqrt{2^{-\nu}K} + 7KlN^{-1/3lv} \;<\; \sqrt{\alpha\beta}.$$

One can check that

$$K = e^{\frac{1}{100}(\alpha\beta \log N)^{1/3}}$$

works for $N > N(\alpha, \beta)$. A more precise result can be obtained if desired (which might well be the case, if one is considering sets whose density tends to 0 slowly with N), but we leave this to the reader. One gets a constant c for which one can take

$$K = e^{c\left((\alpha\beta \log N)^{1/3} - \log\log N\right)}.$$

We remark that the fact that we are considering modular arithmetic progressions rather than genuine ones makes almost no difference to the bound we have obtained. This follows from the observation, made at the start of our discussion, that a modular progression of length L contains a genuine progression of length at least $\frac{1}{2}\sqrt{L}$. Of even less consequence is the fact that the N we have been using is actually the least prime greater than twice the original N, as we stated at the very beginning of the proof.

References

[1] Bourgain, J. *On arithmetic progressions in sums of sets of integers*, in *A Tribute to Paul Erdös*, CUP 1990.

[2] Heath-Brown, D.R. *Integer Sets Containing No Arithmetic Progressions*, J. London Math. Soc (2) **35** (1987) 385 – 394.

[3] López, J.M and Ross, K.A. *Sidon Sets*, Lecture Notes in Pure and Applied Mathematics **13**, Marcel Dekker (New York) 1975.

[4] Rudin, W. *Fourier Analysis on Groups*, Wiley 1990 (Reprint of the original 1962 edition).

[5] Ruzsa, I.Z. *Solving a linear equation in a set of integers I*, Acta Arith. **65** (1993) 259–282.

[6] Szemerédi, E. *Integer Sets Containing No Arithmetic Progressions*, Acta. Math. Hung. **56** (1990) 155 – 158.

[7] Tao, T. *Lecture Notes from a course on the Restriction and Kakeya Conjectures*, available at http://www.math.ucla.edu/ tao.