DENSITY INCREMENT METHODS IN ADDITIVE COMBINATORICS

THOMAS F. BLOOM

CHAPTER 1

Lecture One

1. What is the density increment method?

The density increment method is one of the oldest and most useful tools in additive combinatorics. It is an efficient way of turning the 'structure vs. randomness' heuristic into an actual proof. The general idea behind the method is:

(1) We want to prove some property of a set which is perhaps quite sparse inside some global structure.
(2) Show that this property holds if the set behaves randomly.
(3) Show that this property holds if the set is very dense inside the global structure.
(4) Show that if a set does not behave randomly then there is some local structure whose interection with the original set is larger than we expect (a 'density increment').
(5) Then throw away the rest of the universe and treat this local structure as a global structure. Iterate.
(6) Since the density cannot increase forever, at some point we must either behave randomly, or are very dense. In either case we have the desired property inside some subset of the original set, and deduce it about the original set.

This is perhaps a little too abstract to be useful, so I'll explain what happens in perhaps the most famous application of the density increment argument: finding three-term arithmetic progressions.

Question 1. What is the size of the largest $A \subseteq \{1, \ldots, N\}$ that does not contain a three-term arithmetic progression? (i.e. $x, x+d, x+2d$ with $d \neq 0$, or equivalently, a solution to $x + y = 2z$ with $x \neq y$).

This is a very old question in additive combinatorics, and we'll study it in some depth. It was first considered (in print at least) by Erdős and Turán in 1936, who proved that we must have $|A| \leq (\frac{4}{9} + o(1))N$. They conjectured that $|A| = o(N)$, and a convincing reason for this conjecture is given by the structure vs. randomness heuristic.

Firstly, what happens if $A \subseteq \{1, \ldots, N\}$ is a random set, say selecting each element with probability p? There are $\asymp N^2$ many 3APs in $\{1, \ldots, N\}$ and each one survives to be a 3AP in A with probability p^3. Therefore we expect $\gg p^3 N^2$ many 3APs and this is ≥ 1 if $p \gg N^{-2/3}$. Heuristically then if A is a random set of size $\gg N^{1/3}$ then we might expect there to be 3APs. So certainly a random set of size $\gg N$ should contain 3APs.

At the other extreme, what happens if A is very structured, like being an arithmetic progression itself? Well then clearly A contains lots of 3APs. Note that, importantly, this is true whichever progression A is, as there are no 'local obstructions'. This should be compared to the similar problem of finding solutions to $x+y=z$ – here there are obvious obstructions and we can take A to be the set of odd numbers, a set with density $1/2$ without solutions.

So if A is random or very structured then we expect to find 3APs. Of course there are many sets which are neither, but the fact that in both extremes it's true that 'if $|A| \gg N$ then A contains 3APs' with lots of room to spare (for random sets $|A| \gg N^{1/3}$ suffices, for structured sets $|A| \geq 3$ suffices) suggests that we might be able to interpolate between both extremes to cover all sets. This interpolation is the density increment method - given an arbitrary set either it's random or we can push it more towards the structured end.

The conjecture of Erdős and Turán was proved by Roth in 1953, in what is perhaps the birth of the density increment method.

Theorem 1 (Roth 1953)**.** *For any $\delta > 0$ there exists $N \ll_\delta 1$ such that if $A \subseteq \{1, \ldots, N\}$ has $|A| \geq \delta N$ then A contains a non-trivial three-term arithmetic progression.*

In fact Roth proved the explicit estimate that if $A \subseteq \{1, \ldots, N\}$ contains no 3APs then

$$|A| \ll \frac{N}{\log\log N}.$$

For this theorem we can summarise the density increment method as follows:

(1) (the random part) if $A \subseteq \{1, \ldots, N\}$ has $|A| \gg N$ and $\sum_{n\in A} e(n\theta) = o(|A|)$ for every $\theta \in (1/N, 1)$ (where $e(x) = e^{2\pi i x}$) then A contains non-trivial three-term arithmetic progressions.
(2) (the density increment part) if $\sum_{n\in A} e(n\theta) \gg |A|$ then A has larger than expected density on some arithmetic progression $P \subseteq \{1, \ldots, N\}$.
(3) Now repeat, viewing P (say of length M) as a dilated/translated copy of $\{1, \ldots, M\}$.

Roth's bound has been improved a number of times. In this course I'll discuss in particular the following two landmark results, both of which use the density increment method.

Theorem 2 (Bourgain 1999)**.** *If $A \subseteq \{1, \ldots, N\}$ contains no 3APs then*

$$|A| \ll \frac{N}{(\log N)^{1/2-o(1)}}.$$

The following result is a very recent breakthrough, which came as a big surprise. For context, the best known lower bound (due to Behrend 1946) has a set A without 3APs of size

$$|A| \gg \frac{N}{\exp(c(\log N)^{1/2})}$$

for some constant $c > 0$.

Theorem 3 (Kelley-Meka 2023)**.** *If $A \subseteq \{1, \ldots, N\}$ contains no 3APs then*

$$|A| \ll \frac{N}{\exp(c(\log N)^{1/12})}$$

for some constant $c > 0$.

I'll sketch the Kelley-Meka approach, but will prove in detail the Bourgain result – Bourgain's proof is a masterclass in the proper use of the density increment method and Bohr sets, and displays a lot of important technical details that are glossed over in other treatments.

The rough plan for this course is to cover the following:

1. Meshulam's bound and the finite field model (3APs in $\mathbb{F}_p^n$).
2. Bohr sets and basic theory.
3. Bourgain's bound.
4. Chang's lemma.
5. Sketch of the idea behind Bateman-Katz.
6. Sketch of the Kelley-Meka proof.
7. Application of density increment to colouring problems.
8. Energy increment method.

2. Preliminaries

Asymptotic notation. We write $f(x) = O(g(x))$ if there exists some constant $C > 0$ such that $|f(x)| \leq C\,|g(x)|$ for all sufficiently large x. We will also use the Vinogradov notation $f \ll g$ to denote the same thing (so that $f = O(g)$ and $f \ll g$ are equivalent). Occasionally we will use subscript notation to denote dependence of the constants. For example, $f \ll_\delta g$ means there exists some constant $C(\delta)$ depending on δ such that $|f(x)| \leq C(\delta)\,|g(x)|$ for all sufficiently large x (where sufficiently large may also depend on δ). We may write $O(f)$ to denote some unspecified function g which satisfies $g = O(f)$ (for example, one can say $(x+h)^2 = x^2 + O_h(x)$).

We write $f = o(g)$ if $\lim_{x\to\infty} \frac{f(x)}{g(x)} = 0$. We will also write $f \asymp g$ to mean $f \ll g \ll f$. We also write $f \lesssim g$ to mean $f \leq (\log X)^{O(1)} g$ where X is some parameter usually clear from context.

Functions. We will usually adopt an analytic point of view, in particular often viewing sets $A \subseteq G$ as their indicator function

$$1_A(x) = \begin{cases} 1 & \text{if } x \in A \text{ and} \\ 0 & \text{otherwise.} \end{cases}$$

We define the convolution of two functions $f, g : G \to \mathbb{C}$ by

$$f * g(x) = \sum_{y\in G} f(y)g(x-y) = \sum_{y+z=x} f(y)g(z).$$

It's convenient to define the difference convolution by

$$f \circ g(x) = \sum_{z\in G} f(x+z)\overline{g(z)} = \sum_{y-z=x} f(y)\overline{g(z)}.$$

We also define the inner product by

$$\langle f, g\rangle = \sum_x f(x)\overline{g(x)}.$$

We note here the trivial, but useful, adjoint property, that

$$\langle f * g, h\rangle = \langle f, h \circ g\rangle.$$

Indeed, this is nothing more than an analytic expression of the triviality

$$x + y = z \quad \text{if and only if} \quad x = z - y.$$

Dyadic pigeonholing. Suppose we have a function $f : X \to [\delta, \Delta]$. The idea of dyadic pigeonholing (sometimes called 'layer-cake decomposition' in harmonic analysis) is to divide the domain into level sets of the shape $\{x : 2^k \leq f(x) < 2^{k+1}\}$. There are $O(\log(\Delta/\delta))$ many such level sets we need to consider, which for many applications is a negligible factor. Ignoring this factor therefore we can essentially assume the function is constant up to a multiplicative factor. This is very useful for both proofs and heuristics, and can be used as a more informative version of the Cauchy-Schwarz inequality (or Hölder's inequality). For example if $f : X \to \{1, \ldots, M\}$ then we know by Cauchy-Schwarz that

$$\sum_{x \in X} f(x)^2 \geq \lvert X \rvert^{-1} \left(\sum f(x) \right)^2.$$

On the other hand, by dyadic pigeonholing, there exists some set $S \subseteq X$ and an integer K such that $f(x) \asymp K$ on S and

$$\lvert S \rvert \gtrsim K^{-1} \sum f(x).$$

It follows that

$$\sum_{x \in S} f(x)^2 \gg K^2 \lvert S \rvert \gg K \sum f(x).$$

Since $\lvert S \rvert \leq \lvert X \rvert$ we know that $K \gg \lvert X \rvert^{-1} \sum f(x)$, and so this recovers the Cauchy-Schwarz conclusion (a little weaker by a factor of $O(\log M)$). If f is roughly constant on X then $S = X$ and we get nothing new. If not, however, then we have done better than Cauchy-Schwarz by having $\sum_{x \in S} f(x)^2$ large with a much smaller S.

Of course there are various equivalent ways to do this – the observation that Cauchy-Schwarz can be improved if the function is not constant can be formalised in various ways. I find, however, this coarse dyadic pigeonhole trick to be the easiest to think about.

Fourier analysis. For any finite abelian group G, we can consider its dual group $\widehat{G}$ of characters, which are homomorphisms $\gamma : G \to \mathbb{C}$. The set of characters can be made into a group, with the group operation given by pointwise multiplication, so that $(\gamma \cdot \lambda)(x) = \gamma(x)\lambda(x)$. We will use $\mathbf{1}$ to denote the trivial character, the identity of $\widehat{G}$. We will always use lower-case Greek letters to denote characters, and will use additive notation for the group operation in both G and $\widehat{G}$.

Lemma 1. *If G is a finite abelian group then $\widehat{G}$ is isomorphic to G. (In particular it is also a finite abelian group, and is of the same order.)*

For example, if $G = \mathbb{F}_p^n$, then for any $\gamma \in \mathbb{F}_p^n$ we have an associated character

$$\gamma(x) := e(\gamma \cdot x/p),$$

with $e(x) = e^{2\pi i x}$. Similarly, if $G = \mathbb{Z}/N\mathbb{Z}$, any $\gamma \in \mathbb{Z}/N\mathbb{Z}$ yields a character by

$$\gamma(x) = e(\gamma x/N)$$

(where we think of γ and x as integers in $\{1, \ldots, N\}$, for example).

We will adopt the convention that when talking about G we will use the 'counting measure', i.e. unnormalised sums. When dealing with $\widehat{G}$, we will use the 'probability measure', which is just a sum but normalised by dividing through by the size of the group. (There are good philosophical reasons for this: it is known that the dual operation turns discrete groups (which naturally have the counting measure) into compact groups (which naturally have a probability measure), and vice versa. As G is finite, it is both compact and discrete, so one could use either the counting or probability measure, and both are defensible positions. If we decide to prioritise that G is discrete, in using the counting measure, then it is natural to view $\widehat{G}$ as a compact group above all else, hence the probability measure.)

Thus the natural inner product for functions on G is

$$\langle f, g\rangle = \sum_{x\in G} f(x)\overline{g(x)}.$$

When dealing with $\widehat{G}$ it is convenient to introduce new notation that hides the normalising factor – convention in this area is to use expectation notation. In this context it has nothing to do with probability, but is defined as

$$\mathop{\mathbb{E}}_{\gamma\in\widehat{G}} f(\gamma) = \frac{1}{|G|}\sum_{\gamma\in\widehat{G}} f(\gamma).$$

Use of the expectation notation is widespread in additive combinatorics, and is a very convenient way of sweeping normalising factors under the rug. In general, one should just view it as a sum, and check at the end that the normalising factors of $1/|G|$ go where they should.

Definition 1. For any $f : G \to \mathbb{C}$ we define the Fourier transform of f to be the function $\widehat{f} : \widehat{G} \to \mathbb{C}$ defined by

$$\widehat{f}(\gamma) = \langle f, \gamma\rangle = \sum_{x\in G} f(x)\overline{\gamma(x)} = \sum_{x} f(x)\gamma(-x).$$

Lemma 2 (Parseval's identity)**.** *For any $f, g : G \to \mathbb{C}$,*

$$\langle f, g\rangle = \langle \widehat{f}, \widehat{g}\rangle.$$

In particular, $\|f\|_2 = \|\widehat{f}\|_2$ for any function $f : G \to \mathbb{C}$.

Proof. This is simply writing out the definitions and rearranging (remember all sums are finite, so no delicate analytical issues arise), and using orthogonality:

$$\begin{aligned}\langle f, g\rangle &= \sum_{x\in G} f(x)\overline{g(x)}\\ &= \sum_{x,y\in G} f(x)\overline{g(y)} \mathop{\mathbb{E}}_{\gamma\in\widehat{G}} \gamma(y-x)\\ &= \mathop{\mathbb{E}}_{\gamma\in\widehat{G}} \left(\sum_{x\in G} f(x)\gamma(-x)\right)\left(\sum_{y\in G}\overline{g(y)\gamma(-y)}\right)\\ &= \langle \widehat{f}, \widehat{g}\rangle.\end{aligned}$$

□

Lemma 3 (Diagonalising convolution)**.** *For any $f, g : G \to \mathbb{C}$,*

$$\widehat{f * g} = \widehat{f} \cdot \widehat{g}$$

and

$$\widehat{f \circ g} = \widehat{f} \cdot \overline{\widehat{g}}.$$

Proof. By definition, for any $\gamma \in \widehat{G}$,

$$\widehat{f * g}(\gamma) = \sum_{x,y \in G} f(x)g(y)\overline{\gamma(x+y)}.$$

Since $\gamma(x+y) = \gamma(x)\gamma(y)$ this sum factorises and we're done. The other claim is proved in a similar fashion:

$$\widehat{f \circ g}(\gamma) = \sum_{x,y \in G} f(x)g(y)\overline{\gamma(x-y)} = \left(\sum_{x \in G} f(x)\overline{\gamma(x)}\right)\left(\sum_{y \in G} g(y)\gamma(y)\right).$$

□

In particular, for example, if $A \subseteq G$ then

$$\widehat{1_A \circ 1_A} = |\widehat{1_A}|^2,$$

and so the Fourier transform of $1_A \circ 1_A$ is always a non-negative real number. This is much more convenient that the Fourier transform of $1_A * 1_A$, which may take complex values. This is one reason why it is often more convenient to work with $\circ$ than $*$.

Finally, we remark that the Fourier transform is invertible, in the following sense.

Lemma 4. *For any $f : G \to \mathbb{C}$ and any $x \in G$,*

$$f(x) = \mathop{\mathbb{E}}_{\gamma} \widehat{f}(\gamma)\gamma(x).$$

The proof is a simple exercise in orthogonality (or follows directly from Parseval's identity).

CHAPTER 2

Lecture Two

3. Roth's theorem in $\mathbb{F}_p^n$

The most interesting setting for additive combinatorics is usually the integers – but the questions make sense over any abelian group. It is often much simpler to study these questions over a simpler abelian group such as $\mathbb{F}_p^n$, where p is some fixed small prime. This is known as the 'finite field model' (there are now no less than three surveys on this, due to Ben Green, Julia Wolf, and Sarah Peluse).

The point is that $\mathbb{F}_p^n$ behaves like a very rigid/structured version of the integers, so many technical difficulties are smoothed over. Assuming the kind of proof techniques we use are robust enough (e.g. Fourier analysis) the hope is that when we go to the integers and replace exact structure by approximate structure a form of the same proof, with the same basic ideas but more technical workarounds, will go through.

Question 2. Let $p \geq 3$ be a fixed prime. What is the size of the largest $A \subseteq \mathbb{F}_p^n$ that does not contain a three-term arithmetic progression? (i.e. $x, x+d, x+2d$ with $d \neq 0$, or equivalently, a solution to $x + y = 2z$ with $x \neq y$).

We remark here that it is much more convenient to count *all* solutions to $x+y = 2z$, even those 'trivial' ones with $x = y = z$. Generally, from now on, when I say 3AP I will include such trivial ones.

Before the proof, we'll give a big picture sketch. We are given a set $A \subseteq \mathbb{F}_p^n$, and all we know about it is its size. Suppose A has no non-trivial 3APs. We want to show $\alpha = |A|/p^n$ is small. The idea of the density increment process is to show that either:

(1) A has $\gg \alpha^3 p^{2n}$ many 3APs (the 'random' case), and hence
   (a) A is very small, $\alpha \ll p^{-n/2}$, and done, or
   (b) A has non-trivial 3APs, contradiction,
   or,
(2) A must be structured in the following weak sense: it is not well-distributed across different cosets. In particular, there is a large (coset of a) subspace $W \leq \mathbb{F}_p^n$ on which the density of A is large.

But then we zoom in on A intersect this coset. Translate the coset so that it's also a subspace. There are still no 3APs, since 3APs are translation invariant. So we now have a large subset of W without 3APs. Do it all over again! We can't carry on in the second case forever, since the density can never go past 1. So at some point exit in the very small case.

In this section we'll make the above sketch rigorous. Fourier analysis will be essential in the structured case, and we will use it to find the subspace on which A has increased density.

Our goal is to prove the following estimate.

Theorem 4 (Meshulam)**.** *If $A \subseteq \mathbb{F}_p^n$ has no non-trivial three-term arithmetic progressions then*

$$|A| \ll_p \frac{p^n}{n}.$$

(In particular, $|A|/p^n \to 0$ as $n \to \infty$.)

(Note that, writing $N = p^n$ for the size of the group, this upper bound looks like $|A| \ll N/\log N$, which is better than Bourgain's upper bound of $|A| \ll N/(\log N)^{1/2+o(1)}$ – but Bourgain's bound works for the harder setting of $\mathbb{Z}/N\mathbb{Z}$!)

Our main tool is the following lemma, which says that if A has no 3APs then either A is small, or there is a large density increment.

Lemma 5. *Let V be an n-dimensional vector space over $\mathbb{F}_p$, and let $A \subseteq V$ be a subset of density $\alpha = |A|/p^n$. Suppose that A has no non-trivial three-term arithmetic progressions. Then either*

1. *$|A| \leq (2p^n)^{1/2}$, or*
2. *there is a subspace $V' \leq V$ of codimension 1 and $x \in V$ such that*

$$\frac{|(A-x)\cap V'|}{|V'|} \geq (1+\tfrac{1}{4}\alpha)\alpha.$$

Before proving this, we will show how to use it iteratively in a density increment fashion to prove Meshulam's theorem. There are various different ways to phrase this. We find using the language of maximality the most straightforward.

Proof of Theorem 4. Let $A \subseteq \mathbb{F}_p^n$ be a fixed set of density $\alpha > 0$ without non-trivial 3APs. Our goal is to show that $\alpha \ll p^n/n$. If $\alpha \leq p^{-n/4}$ then we're done, so suppose that $\alpha > p^{-n/4}$. Also, note that it suffices to prove the bound for large n, since for small n we can just use the trivial $|A| \leq p^n$ and change the hidden constant in $\ll_p$ accordingly.

Let $k \geq 0$ be maximal such that the following holds. There is a sequence of sets $A_0, \ldots, A_k$ and associated vector spaces $V_0, \ldots, V_k$ such that

1. $A_0 = A$ and $V_0 = \mathbb{F}_p^n$,
2. $A_i \subseteq V_i$,
3. A_i has no non-trivial three-term arithmetic progressions,
4. if $\alpha_i = |A_i|/|V_i|$ then
$$\alpha_{i+1} \geq (1+\alpha_i/4)\alpha_i,$$
5. $|V_{i+1}| \geq |V_i|/p$.

How large can k be? Well, simple induction shows that

$$\alpha_i \geq (1+\alpha/4)^i\alpha \geq (1+i\alpha/4)\alpha.$$

In particular, after $\lceil 4/\alpha \rceil$ many steps, $\alpha_i \geq 2\alpha$. After another $\lceil 4/2\alpha \rceil$ many steps, $\alpha_i \geq 4\alpha$, and so on. In the end, after

$$\sum_{i=0}^{r} \lceil 4/2^i\alpha \rceil$$

many steps, the density is $\geq 2^r\alpha$ – but since trivially the density is ≤ 1, this forces $r \ll \log(1/\alpha)$. So

$$k \leq \sum_{i=0}^{O(\log(1/\alpha))} \lceil 4/2^i\alpha \rceil \ll \sum_{i=0}^{\infty} (4/2^i\alpha) + O(\log(1/\alpha)) \ll \alpha^{-1}.$$

In particular, we can assume that $k \leq n/10$, or else $\alpha^{-1} \gg n$, and so $\alpha \ll 1/n$ as required.

Now let's see what Lemma 5 tells us, applied to $A_k \subseteq V_k$. By maximality of k, the second condition of Lemma 5 can't hold: otherwise we could let $V_{k+1} = V'$ and $A_{k+1} = A - x$. Therefore the first condition must hold, and so

$$|A_k| = \alpha_k |V_k| \ll |V_k|^{1/2}.$$

Hence

$$p^{-n/4} \leq \alpha \leq \alpha_k \ll |V_k|^{-1/2}.$$

But by induction $|V_k| \geq p^{n-k} \geq p^{9n/10}$, and hence

$$p^{-n/4} \ll p^{-9n/20},$$

which is a contradiction for large enough n. □

To complete the proof of Meshulam's theorem, or Roth's theorem in $\mathbb{F}_p^n$, it remains to prove Lemma 5. The strategy is the following:

(1) Write the difference between the actual number of 3APs in A and the 'expected' number of 3APs in a set of the same density as an inner product involving 1_A and the balanced function $1_A - \alpha$.
(2) If A has no non-trivial 3APs, and is not too large, then this difference is large in absolute value.
(3) Apply Parseval's identity, to convert this inner product into one involving the Fourier transform of 1_A and $1_A - \alpha$.
(4) Deduce from the largeness of this inner product that there is some $\gamma \neq \mathbf{1}$ at which the Fourier transform of $1_A - \alpha$ is large.
(5) Show that if V' is the subspace which is orthogonal to γ, which has codimension 1, then the large Fourier coefficient from the previous point creates a density increment on some coset of V'.

Proof of Lemma 5. We will think of V as just $\mathbb{F}_p^n$, and all Fourier transforms, sums, and so on, will be taken over this group. The number of 3APs in A can be written as

$$\sum_{x,y,z\in A} 1_{x+y=2z} = \sum_{x,y\in A}\sum_{w\in 2\cdot A} 1_{x+y=w} = \sum_{w\in 2\cdot A} 1_A * 1_A(w) = \langle 1_A * 1_A, 1_{2\cdot A}\rangle.$$

Here we are using the obvious notation $2\cdot A = \{2a : a \in A\}$ – note that since $\mathbb{F}_p^n$ is a group of odd order $g \mapsto 2g$ is a bijection, and in particular $|2\cdot A| = |A|$.

We will now compare this to the amount of 3APs we 'expect' to see in A. The most convenient way to do this is to consider the same inner product with 1_A replaced by $\alpha 1_G$ – that is, the constant function on $\mathbb{F}_p^n$ which maps every element to α. This can be viewed as the first-order approximation to A, which agrees with it in density, in the sense that $\sum_x 1_A(x) = |A| = \alpha p^n = \sum_x \alpha 1_G(x)$. As a constant function on the entirety of G, it is much easier to count 3APs weighted by this function, even if we only replace one copy of 1_A by $\alpha 1_G$:

$$\begin{aligned}\langle \alpha 1_G * 1_A, 1_{2\cdot A}\rangle &= \alpha\langle 1_G * 1_A, 1_{2\cdot A}\rangle\\ &= \alpha\langle 1_G, 1_{2\cdot A}\circ 1_A\rangle\\ &= \alpha\sum_{x\in G}\left(\sum_{a,b\in A} 1_{2a-b=x}\right)\\ &= \alpha\left|A\right|^2\\ &= \alpha^3 p^{2n}.\end{aligned}$$

This is, recall, the number of 3APs we expect from A if it were a random set of density α. To compare the difference between the actual count and the expected count, we take the difference: let $f_A = 1_A - \alpha 1_G$ be the 'balanced function'. Then, using the fact that the number of 3APs in A is just $|A|$ (since only the trivial ones with $d = 0$ appear), we have

$$\langle f_A * 1_A, 1_{2\cdot A}\rangle = \langle 1_A * 1_A, 1_{2\cdot A}\rangle - \langle \alpha 1_G * 1_A, 1_{2\cdot A}\rangle = |A| - \alpha^3 p^{2n} = \alpha p^n(1-\alpha^2 N).$$

In particular, if the first case does not hold, then $1 - \alpha^2 N \leq -\frac{1}{2}\alpha^2 p^n$, and so

$$|\langle f_A * 1_A, 1_{2\cdot A}\rangle| \geq \tfrac{1}{2}\alpha^3 p^{2n}.$$

We now write the left-hand side using Fourier analysis: Parseval's idenity and the fact that the Fourier transform diagonalises convolution yields

$$\langle f_A * 1_A, 1_{2\cdot A}\rangle = \langle \widehat{f_A}\cdot\widehat{1_A}, \widehat{1_{2\cdot A}}\rangle.$$

Writing out the definition of the inner product and using the triangle inequality, we therefore get

$$\mathop{\mathbb{E}}_{\gamma}\left|\widehat{f_A}(\gamma)\right|\left|\widehat{1_A}(\gamma)\right|\left|\widehat{1_{2\cdot A}}(\gamma)\right| \geq \tfrac{1}{2}\alpha^3 p^{2n}. \tag{1}$$

We now make two observations about the left-hand side: the first is that the trivial character $\gamma = \mathbf{1}$ makes no contribution, since

$$\widehat{f_A}(\mathbf{1}) = \sum_x f_A(x) = \sum_x 1_A(x) - \alpha 1_G(x) = |A| - \alpha p^n = 0.$$

Secondly, we use the Cauchy-Schwarz inequality and Parseval's identity to see that

$$\begin{aligned}\mathop{\mathbb{E}}_{\gamma}\left|\widehat{1_A}(\gamma)\right|\left|\widehat{1_{2\cdot A}}(\gamma)\right| &\leq \left(\mathop{\mathbb{E}}_{\gamma}\left|\widehat{1_A}(\gamma)\right|^2\right)^{1/2}\left(\mathop{\mathbb{E}}_{\gamma}\left|\widehat{1_{2\cdot A}}(\gamma)\right|^2\right)^{1/2}\\ &= \|1_A\|_2\|1_{2\cdot A}\|_2\\ &= |A|\,.\end{aligned}$$

Using this and (1) we have

$$\begin{aligned}\sup_{\gamma\neq\mathbf{1}}\left|\widehat{f_A}(\gamma)\right|\alpha p^n &\geq \sup_{\gamma\neq\mathbf{1}}\left|\widehat{f_A}(\gamma)\right|\mathop{\mathbb{E}}_{\gamma}\left|\widehat{1_A}(\gamma)\right|\left|\widehat{1_{2\cdot A}}(\gamma)\right|\\ &\geq \mathop{\mathbb{E}}_{\gamma\neq\mathbf{1}}\left|\widehat{f_A}(\gamma)\right|\left|\widehat{1_A}(\gamma)\right|\left|\widehat{1_{2\cdot A}}(\gamma)\right|\\ &\geq \tfrac{1}{2}\alpha^3p^{2n}.\end{aligned}$$

In particular, there must exist some $\gamma\neq\mathbf{1}$ such that $|\widehat{f_A}(\gamma)|\geq\frac{1}{2}\alpha^2p^n$. (Compare this to the trivial upper bound $|\widehat{f_A}(\gamma)|\leq 2\alpha p^n$ from the triangle inequality.)

Let V' be the subspace which annihilates γ – that is, the set of all $x\in\mathbb{F}_p^n$ such that $\gamma\cdot x=0$ (recalling our identification of $\mathbb{F}_p^n$ with $\widehat{\mathbb{F}_p^n}$, this is equivalent to $\gamma(x)=1$ viewing γ as a character). This is a subspace of codimension 1. The key observation is that γ (viewed as a character) is now constant on cosets of V' – if the cosets of V' are $v_1+V',\ldots,v_p+V'$ and if $x\in v_i+V'$ then $\gamma(x)=\gamma(v_i)$.

We know that $|\widehat{f_A}(\gamma)|\geq\alpha^2p^n/2$. To see what this has to do with V', we write out the Fourier transform as follows. Let $V_1',\ldots,V_p'$ be the cosets of V'. Then

$$\begin{aligned}\widehat{f_A}(\gamma)&=\sum_{x\in A}(1_A(x)-\alpha 1_G(x))\overline{\gamma(x)}\\ &=\sum_{i=1}^{p}\left(\sum_{x\in V_i'}(1_A(x)-\alpha 1_G(x))\overline{\gamma(x)}\right)\\ &=\sum_{i=1}^{p}\overline{\gamma(v_i)}\left(|A\cap V_i'|-\alpha p^{n-1}\right).\end{aligned}$$

We want to show that there exists some i such that $|A\cap V_i'|-\alpha p^{n-1}\geq\frac{1}{4}\alpha^2p^{n-1}$. One immediate problem is that we only know about the absolute value of $\widehat{f_A}(\gamma)$. The second is that the sum above is a sum of complex values, so extracting information about individual summands from a bound on the sum is difficult. We will now show how to get around such difficulties.

Let $c\in\mathbb{C}$ be such that $\overline{c}\widehat{f_A}(\gamma)=|\widehat{f_A}(\gamma)|$ (so $|c|=1$), and consider

$$\langle f_A,c\gamma+1\rangle=\overline{c}\widehat{f_A}(\gamma)+\sum_x f_A(x)=\left|\widehat{f_A}(\gamma)\right|.$$

In particular, this inner product is a non-negative real number. The function $x\mapsto c\gamma(x)+1$ is constant on cosets of V' - say, takes the values $x_1,\ldots,x_p$. So if we split the inner product into a sum over V_i' for $1\leq i\leq p$ as above, then

$$\langle f_A,c\gamma+1\rangle=\sum_i x_i\left(|A\cap V_i'|-\alpha p^{n-1}\right).$$

Since the left-hand side is a non-negative real value, and is $\geq\frac{1}{2}\alpha^2p^n$, we have

$$\sum_i\mathrm{Re}(x_i)\left(|A\cap V_i'|-\alpha p^{n-1}\right)\geq\tfrac{1}{2}\alpha^2p^n.$$

By averaging (which is now possible since this is a sum of real numbers), there exists i such that

$$\mathrm{Re}(x_i)(|A \cap V_i'| - \alpha p^{n-1}) \geq \tfrac{1}{2}\alpha p^{n-1}.$$

Finally, we note that $\mathrm{Re}(x_i) \in [0, 2]$, and so we're done. (Note how vital it was that we introduced the $+1$, or else $\mathrm{Re}(x_i) \in [-1, 1]$, and we might have found a density decrement instead of an increment.) □

A lot of the above argument makes sense in any finite abelian group, such as $\mathbb{Z}/N\mathbb{Z}$. Where we made essential use of the fact that we're working in $\mathbb{F}_p^n$ was saying that there is a subspace V', which is large, on which $\gamma(x) = 1$. This is the utility of having plentiful subspaces around, which can exactly annihilate any character. In $\mathbb{Z}/N\mathbb{Z}$, this is no longer possible – for example, if $\gamma : x \mapsto e^{2\pi i x/N}$, then $\gamma(x) = 1$ if and only if $x = 0$. So we cannot hope to find a large subgroup on which γ vanishes exactly.

We will instead pass to the subset of those x where $\gamma(x) \approx 1$ – that is, where $|\gamma(x) - 1| \leq \epsilon$ for some small $\epsilon > 0$. With this choice, for suitable ϵ, something similar to the previous argument can be made to work for $\mathbb{Z}/N\mathbb{Z}$ – but the details become more complicated, since these sets are no longer closed under addition.

CHAPTER 3

Lectures Three and Four

4. Bohr sets

In this section we will define Bohr sets, which are a generalisation of subspaces that exist for any finite abelian group, and explore their properties. In this section G is an arbitrary finite abelian group, of order N.

Definition 2 (Bohr set)**.** Let $\Gamma \subset \widehat{G}$ and $\rho \in [0,2]$. The Bohr set with frequency set Γ and width ρ is the set

$$\mathrm{Bohr}(\Gamma;\rho) = \{x \in G : |1-\gamma(x)| \leq \rho \text{ for all } \gamma \in \Gamma\}.$$

If $\lambda > 0$ and $B = \mathrm{Bohr}(\Gamma;\rho)$ is a Bohr set then we will write B_λ for $\mathrm{Bohr}(\Gamma;\lambda\rho)$, which we call B dilated by λ. The size of Γ is called the rank of the Bohr set.

Important: The frequency set Γ and width ρ is not uniquely determined by the corresponding Bohr set! (For example, $\mathrm{Bohr}(\Gamma;2) = G$ for any Γ.) Formally, it would be most proper to always talk of triples $(\mathrm{Bohr}(\Gamma;\rho), \Gamma, \rho)$, but this notation is very cumbersome. Thus we adopt the convention that whenever we refer to a 'Bohr set' B, we are also implicitly fixing some Γ and ρ such that $B = \mathrm{Bohr}(\Gamma;\rho)$.

Before giving some examples, we note some basic properties.

(1) A Bohr set B is always a symmetric set (i.e. $B = -B$) which contains 0. Indeed, this is immediate from the fact that $\gamma(-x) = \overline{\gamma(x)}$ and $\gamma(0) = 1$ for any $\gamma \in \widehat{G}$.

(2) Bohr sets are decreasing in frequency sets, in that if $\Gamma \supseteq \Gamma'$ then $\mathrm{Bohr}(\Gamma;\rho) \subseteq \mathrm{Bohr}(\Gamma';\rho)$.

(3) Bohr sets are increasing in width, in that if $\rho \leq \rho'$ then $\mathrm{Bohr}(\Gamma;\rho) \subseteq \mathrm{Bohr}(\Gamma;\rho')$.

(4)
$$\mathrm{Bohr}(\Gamma;\rho_1) + \mathrm{Bohr}(\Gamma;\rho_2) \subseteq \mathrm{Bohr}(\Gamma;\rho_1+\rho_2).$$

This follows from the triangle inequality, since

$$|1-\gamma(x_1+x_2)| = |\gamma(-x_1) - \gamma(x_2)| \leq |1-\gamma(x_1)| + |1-\gamma(x_2)|\,.$$

In particular, $B + B_\lambda \subseteq B_{1+\lambda}$.

One should think of the Bohr sets with fixed frequency set Γ as a family of neighbourhoods of the origin – where we begin with $\mathrm{Bohr}(\Gamma;0)$ and expand outwards until eventually $\mathrm{Bohr}(\Gamma;2) = G$.

A Bohr set of rank d is the inverse image of a cube of dimension d: if we consider the map from $G \to \mathbb{C}^d$ where $x \mapsto (\gamma(x))_{\gamma\in\Gamma}$ then $\mathrm{Bohr}(\Gamma;\rho)$ is the inverse image of the cube of side-length 2ρ centred at 1. This inverse map is not a homomorphism or anything particularly well-behaved, but still this view of a Bohr set of rank d as the pullback of a d-dimensional cube provides useful intuition.

Examples. Before giving some concrete examples, it is convenient to note the following estimate. Recall that $e(x) = e^{2\pi i x}$. We note that if $\theta \notin \mathbb{Z}$ then

$$|1 - e(\theta)| = \left|e^{-\pi i\theta} - e^{\pi i\theta}\right| = 2\left|\sin(\pi\theta)\right|.$$

We now recall Jordan's inequality:

$$\tfrac{2}{\pi}\left|x\right| \leq \left|\sin(x)\right| \leq \left|x\right|,$$

valid for any $x \in (-\pi/2, \pi/2]$. In particular, if $\|\theta\|$ denotes the distance of θ from the nearest integer, then

$$4\|\theta\| \leq |1 - e(\theta)| \leq 2\pi\|\theta\|.$$

For our first example, recall that if $G = \mathbb{F}_p^n$ then the group of characters $\widehat{G}$ can be identified with $\mathbb{F}_p^n$ itself, where $\gamma \in \mathbb{F}_p^n$ is identified with the character $x \mapsto e(\gamma\cdot x/p)$. In particular, if $\rho < 4/p$, then $|1-\gamma(x)| \leq \rho$ implies $\|\gamma\cdot x/p\| < 1/p$. But $\gamma\cdot x \in \{0,\ldots,p-1\}$, and so the only way this is possible is if $\gamma \cdot x = 0$. That is, provided $\rho < 4/p$, we have shown that, for any $\Gamma \subset \mathbb{F}_p^n$,

$$\mathrm{Bohr}(\Gamma;\rho) = \{x \in \mathbb{F}_p^n : \gamma\cdot x = 0 \text{ for all } \gamma \in \Gamma\}.$$

That is, the Bohr set with frequency set Γ is precisely the subspace of $\mathbb{F}_p^n$ which is orthogonal to all $\gamma \in \Gamma$. This is very convenient, and goes a long way towards explaining why proofs over $\mathbb{F}_p^n$ are much more straightforward: provided the width is sufficiently small (less than some absolute constant depending only on p), Bohr sets in $\mathbb{F}_p^n$ are exactly subspaces (and vice versa). In particular they are closed under addition.

The advantage of Bohr sets in general is that they offer an analogue for 'subspaces', but they exist for any group, even those without subgroups. This is a good general heuristic picture to have in mind when thinking about Bohr sets: "A Bohr set of rank d plays the same role as a subspace of codimension $\leq d$."

Let's consider what Bohr sets look like in $\mathbb{Z}/N\mathbb{Z}$., when N is prime. Again, the group of characters can be identified with $\mathbb{Z}/N\mathbb{Z}$ itself, with $\gamma \in \{0,\ldots,N-1\}$ identified with the character $x \mapsto e(x\gamma/N)$. Consider first the case of rank 1. It is easy to see that $\mathrm{Bohr}(\Gamma;\rho)$ is just an arithmetic progression, centred at 0, of length $\approx \rho N$ – for example, when Γ consists of the character $\gamma : x \mapsto e(x/N)$, then $|1-\gamma(x)| \approx x/N$, and so $x \in \mathrm{Bohr}(\Gamma;\rho)$ if and only if $|x| \ll \rho N$. Changing to a different just dilates this interval, which is another arithmetic progression of the same length. Thus: "Bohr sets in $\mathbb{Z}/N\mathbb{Z}$ of rank 1 are exactly those symmetric arithmetic progressions containing 0."

Bohr sets of higher rank are a little more mysterious, and to understand their structure better we will need some tools from the geometry of numbers. We will explore this further in Chapter 4.

We now return to Bohr sets in general, over an arbitrary finite abelian group. The first basic question is: how large are Bohr sets? Heuristically, if $\gamma(x)$ were distributed equally over the unit circle, then $|1-\gamma(x)| \leq \rho$ would be true with 'probability' $\approx \rho$. Assuming this event is independent for each $\gamma \in \Gamma$, we might guess that the proportion of $x \in G$ that belong to a given Bohr set B of rank d is roughly $\approx \rho^d$, and so $|B| \approx \rho^d N$.

Note that this heuristic also agrees, up to a constant, with what we know about Bohr sets in $\mathbb{F}_p^n$: if $\rho < 4/p$ then $B = \mathrm{Bohr}(\Gamma;\rho)$ is the subspace of $\mathbb{F}_p^n$ which annihilates Γ, which has size $p^{-d'}N$, where $d' \leq |\Gamma|$ is the number of linearly independent elements in Γ. In particular, if Γ is linearly independent and $\rho \approx 4/p$, then $|B| = p^{-d}N \approx (\rho/4)^d N$.

Of course, this heuristic does not always work – for one thing, the distribution of $\gamma(x)$ will not be independent, especially if e.g. both γ and 2γ are elements of Γ (which can already be seen in the $\mathbb{F}_p^n$ subspace case, where d' may be much smaller than d). We can show, however, that this heuristic does work for providing a lower bound on the size of B.

The same idea also shows that dilating a Bohr set at worst reduces the size of the set by a factor exponential in d. This agrees with the heuristic that a Bohr set of d behaves like a cube in dimension d.

Lemma 6. *If B is a Bohr set of rank d and width $\rho \in (0,1]$ then*

$$|B| \geq (\rho/8)^d N.$$

Furthermore,

$$\left|B_{1/2}\right| \geq 8^{-d} \left|B\right|.$$

In particular, for any $0 < \delta < 1$, we have

$$\left|B_\delta\right| \geq (\delta/2)^{3d} \left|B\right|.$$

Proof. Let $B = \mathrm{Bohr}(\Gamma;\rho)$. We can cover the unit circle in $\mathbb{C}$ by at most $\lceil 2\pi/\rho\rceil$ many circles of radius $\rho/2$. In particular, G is covered by at most $\lceil 2\pi/\rho\rceil^d$ many sets of the shape

$$\{x \in G : \gamma(x) \in D_\gamma \text{ for all } \gamma \in \Gamma\},$$

where each D_γ is a circle of radius $\rho/2$ (possibly different circles for different γ). If X is any such set, then $X - X \subseteq B$ by the triangle inequality: suppose that $\gamma \in \Gamma$ and $x_1, x_2 \in X$, say $\gamma(x_1)$ and $\gamma(x_2)$ are both in the circle with centre a and radius $\rho/2$. Then

$$|1 - \gamma(x_1 - x_2)| = |\gamma(x_1) - \gamma(x_2)| \leq |a - \gamma(x_1)| + |a - \gamma(x_2)| \leq \rho.$$

In particular, $|X| \leq |B|$. It follows that

$$N \leq \lceil 2\pi/\rho\rceil^d \left|B\right|,$$

and the claim follows, since $\lceil x\rceil \leq x + 1 \leq (1 + 1/2\pi)x$ for any $x \geq 2\pi$, and $2\pi + 1 \leq 8$.

The second bound is proved similarly, except that now we cover just the part of the unit circle which is distance $\leq \rho$ from 1. This is covered by at most 8 circles of radius $\rho/4$, and hence B is covered by at most 8^d many sets of the shape

$$X' = \{x \in G : \gamma(x) \in D_\gamma \text{ for all } \gamma \in \Gamma\},$$

where each D_γ is a circle of radius $\rho/4$. As before, we have that each such X' satisfies $X' - X' \subseteq B_{1/2}$, and so $|X'| \leq \left|B_{1/2}\right|$, and thus $|B| \leq 8^d \left|B_{1/2}\right|$ as required.

To deduce the third bound, let $k \geq 1$ be such that $2^{-k} \leq \delta < 2^{-k+1}$. By k applications of the second bound,

$$\left|B_\delta\right| \geq \left|B_{1/2^k}\right| \geq 2^{-3kd} \left|B\right| \geq (\delta/2)^{3d} \left|B\right|$$

as required. □

Bohr sets are, in general, not even approximately group-like, and may grow exponentially under addition. Indeed, recall that $\mathrm{Bohr}(\Gamma;\rho)+\mathrm{Bohr}(\Gamma;\rho)\subseteq\mathrm{Bohr}(\Gamma;2\rho)$. If this containment is sharp, and we expect a Bohr set of rank d and radius ρ to have size $\approx \rho^d N$, then this suggests that $|B+B|\approx 2^d\,|B|$ – not so much a problem for $d=O(1)$, but as $d\to\infty$ this becomes very bad indeed!

Thus Bohr sets are, in general, not even approximately group-like. This quickly leads to disaster when naively trying to do Fourier analysis. We can salvage something, however. Note that if B is a Bohr set of rank d then, for any $\lambda>0$, the above heuristic suggests that $B+B_\lambda\approx B_{1+\lambda}\approx(1+\lambda)^d\,|B|$. In particular, if $\lambda\approx 1/d$, then this doubling constant becomes very small, on the order of $1+o(1)$, much more group-like!

The slogan here, then, is that a Bohr set B behaves like a group, and is approximately closed under addition, provided we only translate by elements in some narrow dilate $B_{O(1/d)}$. (As a sanity check, see what happens in $\mathbb{F}_p^n$ - as soon as the width drops below some absolute constant then the Bohr set doesn't change, and so any dilate of B is B again, and this is just saying that subspaces are closed under addition.)

Unfortunately, even this is not true in complete generality – basically because the heuristic that $|B|\approx\rho^d N$ is not definitely true, and it may be that $|B_{1+\lambda}|$ is much larger than we expect. Fortunately, this is not typical behaviour, and an ingenious argument of Bourgain shows that every Bohr set is 'close' to one that behaves how we'd expect. We first formally define what kind of behaviour we're after: a kind of continuity of size, in that small changes in the width should not change the size too much.

Definition 3 (Regularity[a])**.** A Bohr set B of rank d is regular if for all $0\leq\delta\leq 1/200d$ we have

$$|B_{1+\delta}|\leq(1+200d\delta)\,|B|$$

and

$$|B_{1-\delta}|\geq(1-200d\delta)\,|B|\,.$$

[a] The constant 200 here is fairly arbitrary – smaller constants also work, but the proofs become messier. The point is that 200 is a fixed, absolute, constant.

For example, if B is regular, then in particular, for any $0\leq\delta\leq\epsilon/200d$, we have

$$|B+B_\delta|\leq|B_{1+\delta}|\leq(1+200d\delta)\,|B|\leq(1+\epsilon)\,|B|\,.$$

Thus, as discussed above, regular Bohr sets have small sumset with their (narrow) dilates.

Not all Bohr sets are regular! Here's a simple example. Let $\Gamma\subseteq\mathbb{F}_2^n$ be some linearly independent set of size d, and consider the Bohr set in $\mathbb{F}_2^n$ with frequency set Γ and width $2-\frac{1}{1000d}$. Since the characters in $\widehat{G}$ only take the values ±1, if $|1-\gamma(x)|<2$ then $\gamma(x)=1$, and so B is the subspace of characters orthogonal to Γ, which has 2^{n-d}. On the other hand, if $\delta=1/200d$, then since $(1+\delta)(2-1/200d)\geq 2$ we see that $B_{1+\delta}=\mathbb{F}_2^n$, which has size 2^n, and so $|B_{1+\delta}|\geq 2^d\,|B|$. A slight change in the width has resulted in an exponential factor increase in the size. Similar examples can be given for any $\mathbb{F}_p^n$ and, with a little more work, for $\mathbb{Z}/N\mathbb{Z}$.

It's clear what's gone wrong here – we maliciously chose our initial width ρ to be very close to some significant threshold, and then dilating it by a factor of $1+\delta$ pushed us over this threshold, causing a massive jump in size. The key observation is that this malicious choice can be undone if we're allowed to tweak the initial width slightly.

Bourgain showed that this is always true – every Bohr set can be turned into a regular Bohr set by dilating the initial width. A slogan form of this result is that "bad choices for the width are avoidable".

Lemma 7 (Bourgain's Regularity Lemma)**.** *For any Bohr set B there exists $\lambda \in [\frac{1}{2}, 1]$ such that B_λ is regular.*

In the proof of Lemma 7, we will need the following charming elementary result. (This lemma is probably folklore, but I first learnt of it from an expository note on Bourgain's result by Ben Green [5].)

Lemma 8. *Let $\mathcal{I}$ be a collection of open intervals in $\mathbb{R}$ whose union contains a closed interval of length λ. There is a finite collection $I_1, \ldots, I_n \in \mathcal{I}$ of disjoint intervals with total length at least $\lambda/2$.*

Proof. By compactness, there is a finite subset of intervals from $\mathcal{I}$ that contains the same closed interval of length λ. Let $\mathcal{I}'$ be a minimal such set. Fix $x \in \mathbb{R}$, and suppose that there are at least two intervals in $\mathcal{I}'$ containing x. Let $I = (a_I, b_I)$ and $J = (a_J, b_J)$ be two such intervals, chosen such that $a_I < x$ is minimal and $b_J > x$ is maximal. In particular, if $(a,b) \in \mathcal{I}$ also contains x, then $a \geq a_I$ and $b \leq b_J$, and so $(a,b) \subseteq I \cup J$. By the minimality of $\mathcal{I}'$, we deduce that $(a,b) \notin \mathcal{I}'$, and so x is contained in at most two different intervals in $\mathcal{I}'$.

If we list $\mathcal{I}$ as $I_1, \ldots, I_n$, where $I_i = (a_i, b_i)$, ordered such that $a_1 \leq a_2 \leq \cdots \leq a_n$, then we must have

$$a_1 \leq a_2 \leq b_1 \leq a_3 \leq b_2 \leq a_4 \leq \cdots \leq b_{k-1} \leq b_k.$$

In particular the odd intervals $I_1 \cup I_3 \cup \cdots$ are all disjoint, and so are all the even intervals $I_2 \cup I_4 \cup \cdots$. By the pigeonhole principle at least one of them must have measure at least $\lambda/2$. □

We now prove Bourgain's regularity lemma. The basic idea is the following: regularity roughly says that perturbing the width by an (additive) factor of $O(1/d)$ does not change the size by more than $O(1)$. If we have repeated failures of regularity for every $\lambda \in [1/2, 1]$, then we can make $\approx d$ many steps (each of size $O(1/d)$) going from width $1/2$ to width 1, each time increasing the size of the Bohr set by a multiplicative factor. But this means that $|B| \geq C^d |B_{1/2}|$ which, for a suitably large constant $C > 8$, contradicts the fact that $|B| \leq 8^d |B_{1/2}|$ from Lemma 6. The previous covering lemma, and a careful choice of initial constants, allows us to carry out this procedure and get the desired contradiction.

Proof of Lemma 7. Let B be the Bohr set $\mathrm{Bohr}(\Gamma; \rho)$. To make things more visible, let $B(\delta) = B_\delta = \mathrm{Bohr}(\Gamma; \delta\rho)$.

Suppose that the lemma is false. This means that for every $\lambda \in [\frac{1}{2}, 1]$ there exists some $0 < \delta_\lambda \leq \frac{1}{200d}$ such that either

$$|B((1+\delta_\lambda)\lambda)| > (1 + 200\delta_\lambda d)\, |B(\lambda)|\, .$$

or

$$|B((1-\delta_\lambda)\lambda)| < (1-200\delta_\lambda d)\,|B(\lambda)|\,.$$

In either case, we have

$$|B((1+\delta_\lambda)\lambda)| > (1+100\delta_\lambda d)\,|B((1-\delta_\lambda)\lambda)|\,.$$

Consider the collection of intervals of the shape $I_\lambda = ((1-2\delta_\lambda)\lambda, (1+2\delta_\lambda)\lambda)$ for all $\lambda \in [\frac{1}{2}+\frac{1}{100d}, 1-\frac{1}{100d}]$. By Lemma 8, there is some finite set $\{\lambda_1 < \cdots < \lambda_k\}$ such that the corresponding I_{λ_i} are all disjoint and have total measure at least $1/4 - 1/100d \geq 1/5$, and so

$$\sum 4\delta_{\lambda_i}\lambda_i \geq 1/5,$$

and so

$$\sum \delta_{\lambda_i} \geq 1/20.$$

Since $(1-\delta_{\lambda_1})\lambda_1 \geq 1/2$ and $(1+\delta_{\lambda_k})\lambda_k \leq 1$ we have

$$\frac{|B(1/2)|}{|B|} \leq \frac{|B((1-\delta_{\lambda_1})\lambda_1)|}{|B((1+\delta_{\lambda_k})\lambda_k)|}.$$

We further note that, since the disjointness of the intervals above implies that $(1+\delta_{\lambda_i})\lambda_i \leq (1-\delta_{\lambda_{i+1}})\lambda_{i+1}$, we have

$$\frac{\left|B((1-\delta_{\lambda_{i+1}})\lambda_{i+1})\right|}{|B((1+\delta_{\lambda_i})\lambda_i)|} \geq 1.$$

Therefore, using our initial assumption,

$$\begin{aligned}\frac{|B(1/2)|}{|B|} &\leq \frac{|B((1-\delta_{\lambda_1})\lambda_1)|}{|B((1+\delta_{\lambda_k})\lambda_k)|}\\ &\leq \prod_{i=1}^{k} \frac{|B((1-\delta_{\lambda_i})\lambda_i)|}{|B((1+\delta_{\lambda_i})\lambda_i)|}\\ &< \prod_{i=1}^{k}(1+100\delta_{\lambda_i}d)^{-1}.\end{aligned}$$

Using the inequality $1+x \geq e^{x/2}$, valid for all $0 \leq x \leq 1$, this implies

$$\frac{|B(1/2)|}{|B|} \leq \exp(-\tfrac{50}{20}d)) < 8^{-d},$$

say, since $5/2 \geq \log 8$. By Corollary 6, however, the left hand side is at least 8^{-d} and we have a contradiction. □

The following lemmas indicate how regularity of Bohr sets will be exploited. It allows us to remove convolutions by a narrow dilate of B (with a small error).

Lemma 9. *If B is a regular Bohr set of rank d and $B' \subseteq B_\delta$, with $0 < \delta \leq 1/200d$, then for any function f supported on B satisfying $|f(x)| \leq M$ for all $x \in B$,*

$$\langle f, 1_B * 1_{B'}\rangle = \langle f, 1_B\rangle\,|B'| + O(\delta d M\,|B|\,|B'|).$$

In particular, if $A \subseteq B$, then

$$\langle 1_A, 1_B * 1_{B'}\rangle = |A|\,|B'| + O(\delta d\,|B|\,|B'|).$$

Proof. We have, since f is supported on B,

$$\langle f, 1_B * 1_{B'}\rangle - \langle f, 1_B\rangle |B'| = \sum_{x\in B} f(x)\left(1_B * 1_{B'}(x) - |B'|\right).$$

By the triangle inequality, this is at most

$$\begin{aligned} M\sum_{x\in B}\lvert 1_B * 1_{B'}(x) - |B'|\rvert &= M\sum_{x\in B}\left\lvert \sum_{y\in B'}(1_B(x-y)-1)\right\rvert \\ &\leq M\sum_{y\in B'}\sum_{x\in B}\lvert 1_B(x-y)-1\rvert \\ &= M\sum_{y\in B'}\lvert B\backslash(B+y)\rvert. \end{aligned}$$

We now note that $B_{1-\delta}\subseteq B+y$ – indeed, if $z\in B_{1-\delta}$ and $y\in B_\delta$ then $z-y\in B_{1-\delta}+B_\delta\subseteq B$. Therefore, by the definition of regularity,

$$\lvert B\backslash(B+y)\rvert \leq \lvert B\backslash B_{1-\delta}\rvert \ll \delta d\,|B|,$$

and the proof is complete. □

5. Bourgain's bound for Roth's theorem

We will now prove Bourgain's bound for sets without three-term arithmetic progressions. The overall strategy is to mimic the proof we did in $\mathbb{F}_p^n$, but with Bohr sets playing the role of subspaces. The main complication is that since Bohr sets are not closed under addition by themselves, but are approximately closed under addition by a narrow dilate (at least, if the Bohr sets are regular), we will have to work with several widths of the same Bohr set simultaneously.

Our goal is the following result.

Theorem 5 (Bourgain 1999)**.** *If $A\subset\{1,\ldots,N\}$ has no non-trivial three-term arithmetic progressions then*

$$|A| \ll \left(\frac{\log\log N}{\log N}\right)^{1/2} N.$$

In particular, $|A|/N\to 0$ as $N\to\infty$.

An immediate problem if we try to prove this theorem is that $\{1,\ldots,N\}$ is not a group! Everything we've developed in this chapter has been for finite abelian groups. So we will in fact prove the following.

Theorem 6 (Bourgain 1999)**.** *Let G be a finite abelian group of odd order N. If $A\subseteq G$ has no non-trivial three-term arithmetic progressions then*

$$|A| \ll \left(\frac{\log\log N}{\log N}\right)^{1/2} N.$$

(Note that this also includes the case when $G=\mathbb{F}_p^n$ with $p\geq 3$ an odd prime, but of course in this case we have already proved the much better bound $|A|\ll N/\log N$.)

Even though $\{1,\ldots,N\}$ is not a group, there is a neat trick that allows us to deduce Theorem 5 from Theorem 6.

Proof of Theorem 5 assuming Theorem 6. Suppose $A \subseteq \{1, \dots, N\}$ contains no non-trivial 3APs. Let $M = 2N - 1$. Suppose that A had a non-trivial 3AP modulo M. This means that there are distinct $x, y, z \in A$ such that $x + y \equiv 2z \pmod{M}$. But since $1 \leq x, y, z \in N$, we have

$$-M < 2 - 2N \leq x + y - 2z \leq 2N - 2 < M.$$

Therefore $x + y - 2z \equiv 0 \pmod{M}$ implies $x + y - 2z = 0$, and we have found a genuine non-trivial 3AP in A, which is a contradiction. Therefore A, viewed as a subset of $\mathbb{Z}/M\mathbb{Z}$, also has no non-trivial 3APs, and so Theorem 6 applies with $G = \mathbb{Z}/M\mathbb{Z}$. Therefore

$$|A| \ll \left(\frac{\log\log M}{\log M}\right)^{1/2} M \ll \left(\frac{\log\log N}{\log N}\right)^{1/2} N.$$

□

As for the proof of Meshulam's theorem, we will first state the density increment lemma we will use, and then show how Theorem 6 follows from it.

Lemma 10. *Let B be a regular Bohr set of rank d and width ρ. Let $A \subseteq B$ be a subset of density $\alpha = |A| / |B|$. Suppose that A has no non-trivial three-term arithmetic progressions. Then there is a constant $c > 0$ such that either*

1. $|A| \ll (d/\alpha)^{O(d)} |B|^{1/2}$, *or*
2. *there is a regular Bohr set $B' \subseteq B$ of rank $\leq d+1$ and width $\gg \rho(\alpha/d)^{O(1)}$ and x such that*
$$\frac{|(A - x) \cap B'|}{|B'|} \geq (1 + c\alpha)\alpha.$$

We will now prove Theorem 6 by repeated applications of Lemma 10.

Proof. Let $A \subseteq G$ be a fixed set of density $\alpha > 0$ without non-trivial 3APs. We can assume, without loss of generality, that $\alpha \geq 1/\log N$, or else we are done immediately.

Let $k \geq 0$ be maximal such that the following holds. There is a sequence of sets $A_0, \dots, A_k$ and associated Bohr sets $B_0, \dots, B_k$, with ranks $d_0, \dots, d_k$ and widths $\rho_0 \geq \dots \geq \rho_k$, such that

1. $A_0 = A$ and $B_0 = G$, with $d_0 = 1$ and $\rho_0 = 1$ (taking the frequency set to be just the trivial character, for example),
2. $A_i \subseteq B_i$,
3. A_i has no non-trivial 3APs,
4. if $\alpha_i = |A_i| / |B_i|$ then
$$\alpha_{i+1} \geq (1 + c\alpha_i)\alpha_i,$$
where $c > 0$ is the constant from Lemma 10,
5. $d_i \leq i + 1$, and
6. $\rho_{i+1} \gg (\alpha/d_i)^{O(1)}\rho_i$.

Just as in the proof of Theorem 4, part (4) implies that $k \ll \alpha^{-1}$.

We now apply Lemma 10 to $A_k \subseteq B_k$. By maximality of k, the second condition of Lemma 10 can't hold, and so (since $d_k \leq k + 1 \ll \alpha^{-1}$)

$$\frac{1}{\log N} \leq \alpha \leq \alpha_k \ll (d_k/\alpha)^{O(d_k)} |B_k|^{-1/2} \ll (1/\alpha)^{O(\alpha^{-1})} |B_k|^{-1/2}.$$

We now compare this to our lower bound for $|B_k|$. Since the rank of each Bohr set is $\leq k+1 \ll \alpha^{-1}$, we have for $0 \leq i < k$, the width relationship

$$\rho_{i+1} \gg \alpha^{O(1)}\rho_i,$$

and so $\rho_k \gg \alpha^{O(d_k)} \gg \alpha^{O(\alpha^{-1})}$. By our size lower bound for Bohr sets, Lemma 6, we have

$$|B_k| \geq (\rho_k/8)^{d_k} N \gg \alpha^{O(\alpha^{-2})}N.$$

Therefore,

$$\frac{1}{\log N} \ll \alpha^{-O(\alpha^{-1})}\left|B_k\right|^{-1/2} \ll \alpha^{-O(\alpha^{-2})}N^{-1/2}.$$

Rearranging and taking logarithms, this implies

$$\alpha^{-2}\log(1/\alpha) \gg \log N.$$

Since we are assuming that $\alpha \geq 1/\log N$, we have $\log(1/\alpha) \ll \log\log N$, and so

$$\alpha^{-2} \gg \frac{\log N}{\log\log N},$$

and so $\alpha \ll (\log\log N/\log N)^{1/2}$ as required. □

Before we prove the density increment lemma Lemma 10, we need to prove two supporting technical lemmas. These are to compensate for the fact that Bohr sets are not closed under addition, and we need to work with narrower Bohr sets instead and use regularity.

The first of our two supporting lemmas will be used to replace the fact that, in $\mathbb{F}_p^n$, we could exactly work out the number of 3APs when one of the copies of A was replaced by G: namely that $\langle 1_G * 1_A, 1_{2\cdot A}\rangle = \alpha^2 \left|G\right|^2$. This is no longer possible if we replace G by some Bohr set. We will show that, using regularity, we can recover a suitable lower bound for this count, if instead of replacing A we replace $2\cdot A$ by $2\cdot B_\delta$, provided δ is sufficiently small – or at least, either this is possible, or else we have a density increment anyway.

Note that if A is a random subset of B of density α, then we expect for a Bohr set B', we have $\langle 1_A * 1_A, 1_{B'}\rangle \approx \alpha^2\langle 1_B * 1_B, 1_{B'}\rangle$. Provided B is regular and B' is contained in some suitable dilate of B, we also have $\langle 1_B * 1_B, 1_{B'}\rangle \approx |B|\,|B'|$ (since 1_B is approximately invariant under translations by B'). The following lemma says that (with B' replaced by $2\cdot B_\delta$) either this approximation is true as a lower bound, or else A has a strong density increment.

Lemma 11. *Let B be a regular Bohr set of rank d and width ρ. Suppose that $\delta \leq c_0\alpha/d$ for some sufficiently small constant $c_0 > 0$ such that B_δ is also regular. Let $A \subseteq B$ with density $\alpha = |A|/|B|$. Either*

1. *(many 'progressions') $\langle 1_A * 1_A, 1_{2\cdot B_\delta}\rangle \geq \frac{1}{2}\alpha^2\,|B|\,|B_\delta|$ or*
2. *(density increment) there is a regular Bohr set B' of rank $\leq d$ and width $\gg \delta^2\rho$ and an x such that*

$$\frac{|(A-x)\cap B'|}{|B'|} \geq (1+1/256)\alpha.$$

Proof. If the first condition fails then

$$\tfrac{1}{2}\alpha^2\,|B|\,|B_\delta| > \langle 1_A * 1_A, 1_{2\cdot B_\delta}\rangle = \langle 1_A, 1_{2\cdot B_\delta} \circ 1_A\rangle.$$

This means that there can't be too many elements of A where $1_{2\cdot B_\delta}\circ 1_A$ is large. More precisely, decompose $A = A_{\mathrm{large}}\sqcup A_{\mathrm{small}}$, where

$$A_{\mathrm{large}} = \{x\in A : 1_{2\cdot B_\delta}\circ 1_A(x) > \tfrac{3}{4}\alpha\,|B_\delta|\}.$$

We have

$$\tfrac{1}{2}\alpha^2\,|B|\,|B_\delta| > \langle 1_A, 1_{2\cdot B_\delta}\circ 1_A\rangle \geq \tfrac{3}{4}\alpha\,|B_\delta|\,|A_{\mathrm{large}}|\,,$$

and so $|A_{\mathrm{large}}| < \frac{2}{3}|A|$, and hence $|A_{\mathrm{small}}|\geq \frac{1}{3}|A|$.

So we know that A_{small} is large, so there are many elements in A where $1_{2\cdot B_\delta}\circ 1_A$ is small. We now show how to upgrade this to find many elements in B where this convolution is small. Let $c\in[1/2,1]$ be such that $B_{c\delta^2}$ is regular. The key is to note that, by regularity of B_δ, for any $z\in B_{c\delta^2}$,

$$|(2\cdot B_\delta - 2z)\backslash 2\cdot B_\delta| = |(B_\delta - z)\backslash B_\delta| \leq \left|B_{(1+c\delta)\delta}\backslash B_\delta\right| \ll \delta d\,|B_\delta|\,,$$

and hence for any $y\in A_{\mathrm{small}}$ we have

$$\begin{aligned}1_{2\cdot B_\delta}\circ 1_A(y+2z) &= |(2\cdot B_\delta - 2z)\cap(A+y)|\\ &\leq |2\cdot B_\delta\cap(A+y)| + O(\delta d\,|B_\delta|)\\ &= 1_{2\cdot B_\delta}\circ 1_A(y) + O(\delta d\,|B_\delta|).\end{aligned}$$

In particular, for any $x\in A_{\mathrm{small}} + 2\cdot B_{c\delta^2}$, since $\delta\leq c_0\alpha/d$, provided c_0 is a small enough, we have

$$1_{2\cdot B_\delta}\circ 1_A(x) < \tfrac{7}{8}\alpha\,|B_\delta|\,.$$

We therefore let $B_{\mathrm{small}} = B\cap(A_{\mathrm{small}} + 2\cdot B_{c\delta^2})$. This is a subset of B where $1_{2\cdot B_\delta}\circ 1_A$ is small (as is A_{small}, but B_{small} is probably much larger).

How large is B_{small}? We don't know, but we will show that whether B_{small} is large or small, we can obtain a density increment.

Case 1: Suppose that $|B_{\mathrm{small}}| < \frac{1}{16}|B|$. In this case we consider the convolution $\langle 1_{A_{\mathrm{small}}} * 1_{2\cdot B_{c\delta^2}}, 1_B\rangle$. By regularity, and noting that $2\cdot B_{c\delta^2}\subseteq B_{c\delta^2}+B_{c\delta^2}\subseteq B_{2c\delta^2}$,

$$\langle 1_{A_{\mathrm{small}}} * 1_{2\cdot B_{c\delta^2}}, 1_B\rangle = \langle 1_{A_{\mathrm{small}}}, 1_B * 1_{2\cdot B_{c\delta^2}}\rangle = |A_{\mathrm{small}}|\,|B_{c\delta^2}| + O(\delta^2 d\,|B|\,|B_{c\delta^2}|).$$

(Note that the adjoint property would suggest a $\circ 1_{2\cdot B_{c\delta^2}}$ here in the second expression, but since Bohr sets are symmetric, it is the same whether we write $\circ$ or $*$ here! This kind of substitution, between $\circ$ and $*$, which are equivalent for symmetric sets, will doubtless happen again.)

Provided δ is small enough, this is at least (recall that $|A_{\mathrm{small}}|\geq\frac{1}{3}|A|$)

$$|A_{\mathrm{small}}|\,|B_{c\delta^2}| - \tfrac{1}{8}|A|\,|B_{c\delta^2}| \geq \tfrac{1}{8}|A|\,|B_{c\delta^2}|\,.$$

Since $1_{A_{\mathrm{small}}} * 1_{2\cdot B_{c\delta^2}}$ is supported, inside B, on B_{small}, we have

$$\begin{aligned}\langle 1_{A_{\mathrm{small}}} * 1_{2\cdot B_{c\delta^2}}, 1_B\rangle &\leq |B_{\mathrm{small}}|\max_x(1_{A_{\mathrm{small}}} * 1_{2\cdot B_{c\delta^2}}(x))\\ &\leq \tfrac{1}{16}|B|\max_x(1_A * 1_{2\cdot B_{c\delta^2}}(x)).\end{aligned}$$

Comparing the upper and lower bounds, we deduce that

$$\max_x |(A-x)\cap 2\cdot B_{c\delta^2}| = \max_x 1_A * 1_{2\cdot B_{c\delta^2}}(x) \geq 2\alpha\,|B_{c\delta^2}|\,,$$

and we have a density increment (even better than we needed), with $B' = 2\cdot B_{c\delta^2}$. Here we are using the observation that if $B = \mathrm{Bohr}(\Gamma;\rho)$ is a Bohr set then $2\cdot B$ is also a Bohr set of the same rank and width:

$$2\cdot\mathrm{Bohr}(\Gamma;\rho) = \mathrm{Bohr}(2^{-1}\Gamma;\rho),$$

where

$$2^{-1}\Gamma = \{x \mapsto \gamma(2^{-1}x) : \gamma \in \Gamma\}.$$

Here we use $2^{-1}x$ to denote the inverse homomorphism to $x \mapsto 2x$, which exists since G is a finite group of odd order, so $x \mapsto 2x$ is an injective, and hence bijective, homomorphism. Furthermore, if B is regular then $2 \cdot B$ will also be regular, since $\lvert(2 \cdot B)_{1+\delta}\rvert = \lvert 2 \cdot B_{1+\delta}\rvert = \lvert B_{1+\delta}\rvert$.

Case 2: Suppose that $\lvert B_{\mathrm{small}}\rvert \geq \frac{1}{16}\lvert B\rvert$. In this case we consider the inner product $\langle 1_{2\cdot B_\delta} \circ 1_A, 1_B\rangle$. As above, by regularity, provided δ is sufficiently small, we have

$$\langle 1_{2\cdot B_\delta} \circ 1_A, 1_B\rangle \geq (1 - \tfrac{1}{256})\lvert A\rvert \lvert B_\delta\rvert .$$

For an upper bound, we recall that if $x \in B_{\mathrm{small}}$ then $1_{2\cdot B_\delta} \circ 1_A(x) \leq \frac{7}{8}\alpha \lvert B_\delta\rvert$. Also, for any $x \in B$, either we have a density increment (with $B' = 2 \cdot B_\delta$), or

$$1_{2\cdot B_\delta} \circ 1_A(x) = \lvert (A + x) \cap 2 \cdot B_\delta\rvert \leq (1 + 1/256)\alpha \lvert B_\delta\rvert .$$

Combining these upper bounds, we deduce that

$$\langle 1_{2\cdot B_\delta} \circ 1_A, 1_B\rangle \leq \tfrac{7}{8}\alpha \lvert B_\delta\rvert \lvert B_{\mathrm{small}}\rvert + (1 + 1/256)\alpha \lvert B_\delta\rvert (\lvert B\rvert - \lvert B_{\mathrm{small}}\rvert).$$

Comparing our lower and upper bounds and simplifying yields

$$(\tfrac{1}{8} + \tfrac{1}{256})\lvert B_{\mathrm{small}}\rvert \leq \tfrac{1}{128}\lvert B\rvert ,$$

which contradicts our lower bound on $\lvert B_{\mathrm{small}}\rvert$, and so we must have the required density increment. □

The previous lemma shows our need to work on two different scales at once, and to count 3APs where two elements come from B but the middle element comes from a narrowed copy B_δ. This suggests that when working with $A \subseteq B$ we need to count 3APs where two elements come from A and the third comes from $A \cap B_\delta$. There is a problem with this though – we don't know how large $A \cap B_\delta$ is. Indeed, it might even be empty! B_δ is (possibly) much smaller than B, so might entirely miss A. To avoid this, we show that there exists some translate of A which is reasonably large in both a narrowed copy of B and also in a doubly narrowed copy of B – or, as above, we have a density increment that we're happy with.

Lemma 12. *Let B be a regular Bohr set of rank d and suppose $A \subseteq B$ has density $\alpha = \lvert A\rvert / \lvert B\rvert$. Suppose that $B', B'' \subseteq B_\delta$ where $\delta = c_0\alpha\epsilon/d$ for some sufficiently small absolute constant $c_0 > 0$. Then either*

1. *there is an $x \in B$ such that $\lvert (A - x) \cap B'\rvert \geq (1-2\epsilon)\alpha \lvert B'\rvert$ and $\lvert (A - x) \cap B''\rvert \geq (1 - 2\epsilon)\alpha \lvert B''\rvert$, or*
2. *there is an x such that*

$$\max\left(\frac{\lvert (A - x) \cap B'\rvert}{\lvert B'\rvert}, \frac{\lvert (A - x) \cap B''\rvert}{\lvert B''\rvert}\right) \geq (1 + \epsilon)\alpha.$$

Proof. By regularity (in particular the second conclusion of Lemma 20),

$$\begin{aligned}\langle 1_A * 1_{B'}, 1_B\rangle &= \langle 1_A, 1_B * 1_{B'}\rangle \\ &= \lvert A\rvert \lvert B'\rvert + O(\delta d \lvert B\rvert \lvert B'\rvert) \\ &= \alpha \lvert B\rvert \lvert B'\rvert + O(\delta d \lvert B\rvert \lvert B'\rvert).\end{aligned}$$

In particular, provided $\delta \leq c\alpha/d$ for some small enough absolute constant $c > 0$, we have

$$\langle 1_A * 1_{B'}, 1_B\rangle \geq (1 - \epsilon/2)\alpha \lvert B\rvert \lvert B'\rvert$$

and similarly

$$\langle 1_A * 1_{B''}, 1_B\rangle \geq (1-\epsilon/2)\alpha\,|B|\,|B''|\,.$$

In particular, if $\mu_{B'} = \frac{1}{|B'|}1_{B'}$ and $\mu_{B''} = \frac{1}{|B''|}1_{B''}$ then

$$\langle 1_A * \mu_{B'} + 1_A * \mu_{B''}, 1_B\rangle \geq (2-\epsilon)\alpha\,|B|\,.$$

By the pigeonhole principle, there exists some $x \in B$ such that

$$1_A * \mu_{B'}(x) + 1_A * \mu_{B''}(x) \geq (2-\epsilon)\alpha.$$

If $1_A*\mu_{B'}(x) \geq (1+\epsilon)\alpha$ then we are in the second case, and similarly for $1_A*\mu_{B''}(x)$. Thus either the second case holds, or else both

$$1_A * \mu_{B'}(x) \geq (1-2\epsilon)\alpha$$

and

$$1_A * \mu_{B''}(x) \geq (1-2\epsilon)\alpha$$

as required. □

We are now ready to prove our density increment result, Lemma 10. The overall structure of the proof is very similar to the simpler case in $\mathbb{F}_p^n$, Lemma 5, but there are complications due to having to work with Bohr sets of different widths.

Proof of Lemma 10. Let $A \subseteq B$ with density $\alpha = |A|\,/\,|B|$, where B is a regular Bohr set of rank d and width ρ. We need to work with different layers of Bohr sets in this proof, so it's convenient to define them now: let $B^{(1)} = B_{\delta_1}$ and $B^{(2)} = (B^{(1)})_{\delta_2} = B_{\delta_1\delta_2}$, where $\delta_i = c_i\alpha^2/d$, with c_1, c_2 some absolute constants chosen to be sufficiently small and such that $B^{(1)}$ and $B^{(2)}$ are themselves regular.

We begin by applying Lemma 12 with $B^{(1)}, B^{(2)}$ playing the roles of B', B'', and $\epsilon = c\alpha$, where $c > 0$ is some small constant we'll choose later. If the second case holds, then we have a density increment as needed. Otherwise, there is some x such that if we let $A_1 = (A-x)\cap B^{(1)}$, with density $\alpha_1 = |A_1|\,/\,\left|B^{(1)}\right|$, and similarly $A_2 = (A-x)\cap B^{(2)}$, with density $\alpha_2 = |A_2|\,/\,\left|B^{(2)}\right|$, then $\min(\alpha_1,\alpha_2) \geq (1-2\epsilon)\alpha$. (In particular, provided $\epsilon \leq 1/4$, we have $\alpha_1 \geq \alpha/2$.)

Crucially, because A itself has no non-trivial 3APs, and 3APs are translation invariant, there are still no non-trivial solutions to $x+y=2z$ where $x,y \in A_1$ and $z \in A_2$. This means that

$$\langle 1_{A_1} * 1_{A_1}, 1_{2\cdot A_2}\rangle = |A_2|\,.$$

On the other hand, Lemma 11 implies that either we have a suitable density increment, and we are done, or else

$$\langle 1_{A_1} * 1_{A_1}, 1_{2\cdot B^{(2)}}\rangle \geq \tfrac{1}{2}\alpha_1^2|B^{(1)}||B^{(2)}|.$$

If $\alpha_1 < 2|B^{(1)}|^{-1/2}$, then we are in the first case: by repeated applications of the second part of Lemma 6 we have that $\left|B^{(1)}\right| \geq (\delta_1)^{O(d)}\,|B|$, and hence

$$\alpha \leq 2\alpha_1 \ll |B^{(1)}|^{-1/2} \ll (d/\alpha)^{O(d)}\,|B|^{-1/2}$$

as required. Otherwise, if $f = 1_{2\cdot A_2} - \alpha_2 1_{2\cdot B^{(2)}}$, then

$$\langle 1_{A_1} * 1_{A_1}, f\rangle \leq |A_2| - \tfrac{1}{2}\alpha_1^2|B^{(1)}|\,|A_2| \leq -\tfrac{1}{4}\alpha_1^2|B^{(1)}|\,|A_2|\,.$$

By Parseval's identity and the triangle inequality (just as in the proof of Lemma 5) we deduce that

$$\mathop{\mathbb{E}}_{\gamma}\left|\widehat{f}(\gamma)\right|\left|\widehat{1_{A_1}}(\gamma)\right|^2 \gg \alpha^2 |B^{(1)}|\,|A_2|\,.$$

Again, just as in the proof of Lemma 5, since by Parseval's identity we have $\mathbb{E}_\gamma|\widehat{1_{A_1}}(\gamma)|^2 = |A_1|$, we deduce that there exists some character λ such that

$$\left|\widehat{f}(\lambda)\right| \gg \alpha_1\,|A_2|\,.$$

We now simplify matters by noting that if $f_A = 1_{A_2} - \alpha_2 1_{B^{(2)}}$, then for any x, we have $f(2x) = f_A(x)$, and so

$$\widehat{f_A}(2\lambda) = \sum_x f(2x)\overline{\lambda(2x)} = \sum_y f(y)\overline{\lambda(y)} = \widehat{f}(\lambda).$$

In particular, there is some γ such that $|\widehat{f_A}(\gamma)| \gg \alpha_1\,|A_2|$.

We let B' be the Bohr set formed by adding γ to the frequency set of $B^{(2)}$ and then multiplying the width by a factor of $c_3\alpha^2/d$, where $c_3 > 0$ is another constant chosen in particular so that B' is regular. We will first use regularity to replace $f_A = 1_{A_2} - \alpha_2 1_{B^{(2)}}$ by $f'_A = 1_{A_2} - \alpha_2 1_{B^{(2)}+B'}$. We have that

$$\left|\widehat{f_A}(\gamma) - \widehat{f'_A}(\gamma)\right| \leq \alpha_2\left|(B^{(2)}+B')\backslash B^{(2)}\right| \ll c_3\alpha_2\alpha^2\left|B^{(2)}\right|,$$

and in particular, assuming c_3 is sufficiently small enough, we still have

$$\left|\widehat{f'_A}(\gamma)\right| \gg \alpha\,|A_2|\,.$$

As in the proof of Lemma 5, let $\theta \in \mathbb{C}$ be such that $\overline{\theta}\widehat{f'_A}(\gamma) = |\widehat{f'_A}(\gamma)|$, so that

$$\langle f'_A, \theta\gamma(y) + 1\rangle = \left|\widehat{f'_A}(\gamma)\right| + \sum_x f'_A(x)$$

and hence, since by regularity

$$\sum_x f'_A(x) = |A_2| - \alpha_2|B^{(2)} + B'| = -\alpha_2|(B^{(2)}+B')\backslash B^{(2)}| \ll c_3\alpha_2\alpha^2|B^{(2)}|,$$

provided c_3 is small enough, we have

$$\langle f'_A, \theta\gamma + 1\rangle \gg \alpha_1\,|A_2| \gg \alpha\,|A_2|\,.$$

In the proof of Lemma 5 we divided the sum into cosets $v + V'$. In our present case, there is no such neat decomposition into cosets, so instead we average over all translates $x + B'$ as x ranges over $B^{(2)}$.

Thus, by regularity of $B^{(2)}$, (and since $|f'_A(x)| \ll 1$ for all x)

$$\begin{aligned}\sum_{x\in B^{(2)}}\left(\sum_{y\in B'+x} f'_A(y)\overline{(\theta\gamma(y)+1)}\right) &= \langle 1_{B^{(2)}} * 1_{B'}, f'_A(\theta\gamma+1)\rangle \\ &= |B'|\,\langle 1_{B^{(2)}}f'_A, \theta\gamma+1\rangle + O(c_3\alpha^2|B^{(2)}|\,|B'|).\end{aligned}$$

We relate the value of this inner product to that above by regularity yet again (and using that $|f'_A(\theta\gamma+1)| \ll 1$ and that f'_A is supported on $B^{(2)} + B'$):

$$\langle f'_A, \theta\gamma+1\rangle - \langle 1_{B^{(2)}}f'_A, \theta\gamma+1\rangle \ll |(B^{(2)}+B')\backslash B^{(2)}| \ll c + 3\alpha^2|B^{(2)}|,$$

and so, provided we choose c_3 small enough, we have

$$\langle 1_{B^{(2)}} f'_A, \theta\gamma + 1\rangle \gg \alpha\,|A_2|$$

and hence

$$\sum_{x\in B^{(2)}} \left(\sum_{y\in B'+x} f'_A(y)\overline{(\theta\gamma(y)+1)} \right) \gg \alpha\,|A_2|\,|B'|\,.$$

Finally, we note that while $\gamma(y)$ is not constant on the translates $B'+x$, it is approximately constant: indeed, if $y = t + x$ where $t \in B'$, then

$$|\gamma(y) - \gamma(x)| = |1 - \gamma(t)| \ll c_3\alpha/d,$$

since γ was included in the frequency set of B'. Therefore,

$$\sum_{x\in B^{(2)}} \overline{(\theta\gamma(x)+1)} \left(\sum_{y\in B'+x} f'_A(y) \right) \geq c_4\alpha_1\,|A_2|\,|B'| - O(c_3\tfrac{\alpha}{d}\left|B^{(1)}\right|\left|B^{(2)}\right|).$$

Once again, provided we have chosen $c_3 > 0$ small enough, this right-hand side is at least $\frac{1}{2}c_4\alpha_1\,|A_2|\,|B'|$. Taking the real parts and averaging over all $x \in B^{(2)}$, as in the proof of Lemma 5, we deduce that there exists some $x \in B^{(2)}$ such that

$$|A_2 \cap (B'+x)| - \alpha_2\,|B'| = \sum_{y\in B'+x} f'_A(y) \gg \alpha\alpha_2\,|B'|\,.$$

In particular, there is an absolute constant $c > 0$ such that

$$\frac{|(A_2 - x)\cap B'|}{|B'|} \geq (1 + c\alpha)\alpha_2 \geq (1+c\alpha)(1-2\epsilon)\alpha.$$

If we choose $\epsilon = c\alpha/8$, then the right-hand side is $\geq (1 + \frac{c}{4}\alpha)\alpha$, and we are done, since A_2 itself was a subset of a translate of A. □

CHAPTER 4

Lecture Five

In this chapter we will mostly return to the simpler model setting of $\mathbb{F}_p^n$ and explore how one might strengthen the density increment method to go beyond Meshulam's density bound of $O(1/\log N)$. (Recall our convention that N is the size of the ambient group, which here is p^n, so $\log N \asymp n$).

6. Density increment strengths

It is convenient to introduce the following definition.

Definition 4. Let $A \subseteq \mathbb{F}_p^n$ with density α. We say A has a density increment of strength $[\delta, d]$ if there is some subspace V of codimension $O(d)$ and some x such that

$$\frac{|(A-x)\cap V|}{|V|} \geq (1+c\delta)\alpha,$$

where c is some constant.

In this language, therefore, Meshulam's proof can be phrased as follows.

Lemma 13 (Meshulam's increment). *If $A \subseteq \mathbb{F}_p^n$ is a set of density α with no non-trivial three-term arithmetic progressions then either $\alpha \ll N^{-1/2}$ or A has a density increment of strength $[\alpha, 1]$.*

Let's now recap how we applied this increment to obtain Meshulam's bound: after $O(1/\alpha)$ many steps we must halt since the density is > 1, and each step loses 1 in the codimension, so the final codimension loss is $O(\alpha^{-1})$. We halt with a vector space of size

$$N' \gg p^{-O(\alpha^{-1})}N$$

such that the $\alpha \ll (N')^{-1/2}$. We can assume that $\alpha \gg N^{-1/4}$ (or we are already done), and hence $N' \ll N^{1/4}$, which means $\alpha^{-1} \gg \log N$.

In general, suppose we have a result like (assuming A has no non-trivial three-term arithmetic progressions) either $\alpha \ll N^{-1/2}$ or A has a density increment of strength $[\delta, d]$. This density doubles after $\ll \delta^{-1}$ many steps, and we can double at most $\ll \log(1/\alpha)$ times before we are > 1, so (recalling that $\lesssim$ hides log factors) this increment can apply $\lesssim \delta^{-1}$ many times. Each time we have lost d in the codimension, so the final codimension loss is $\lesssim \delta^{-1}d$. We halt with a vector space of size

$$N' \gg p^{-\tilde{O}(\delta^{-1}d)}N$$

such that the $\alpha \ll (N')^{-1/2}$. We can assume that $\alpha \gg N^{-1/4}$ (or we are already done), and hence $N' \ll N^{1/4}$, which means $\delta^{-1}d \gtrsim \log N$.

To summarise: a density increment lemma of strength $[\delta, d]$ iterated implies that $\delta^{-1}d \gtrsim \log N$. Let's see a couple of examples.

(1) Meshulam's $[\alpha, 1]$ implies $\alpha^{-1} \gtrsim \log N$, or $\alpha \lesssim 1/\log N$.
(2) A density increment of strength $[1, \alpha^{-1}]$ would also yield $\alpha^{-1} \gtrsim \log N$.
(3) A very strong density increment of $[1, 1]$ would imply $1 \gtrsim \log N$. In fact since the $\gtrsim$ is only hiding a single $\log(1/\alpha)$ factor this would imply the very strong density bound of $\alpha \leq N^{-c}$ for some constant $c > 0$. (This is actually true for $\mathbb{F}_p^n$, as was shown by Ellenberg and Gijswijt using Croot-Lev-Pach's polynomial method, but so far this seems out of reach of the density increment method.)

Let's return to the general setting of Bohr sets (in particular to the cyclic group $\mathbb{Z}/N\mathbb{Z}$) to see how things play out there.

Definition 5. Let B be a regular Bohr set of rank r and width ρ, and $A \subseteq B$ with density α. We say A has a density increment of strength $[\delta, d]$ if there is some regular Bohr set $B' \subseteq B$ of rank $r + O(d)$ and width $\gg \alpha^{O(1)}\rho$, and some x such that

$$\frac{|(A - x) \cap V|}{|V|} \geq (1 + c\delta)\alpha,$$

where c is some constant.

This is the same as the special case above, except that now we have a width parameter to also keep track of. The important thing about this is that it affects the size of the Bohr set; recall that if $B \subseteq G$ is a Bohr set of rank r and width ρ then

$$|B| \geq \rho^{O(r)} N.$$

Again, we restate Bourgain's density increment (informally) in these terms.

Lemma 14 (Bourgain's increment)**.** *If B is a regular Bohr set and $A \subseteq B$ is a set of density α with no non-trivial three-term arithmetic progressions then either $\alpha \ll |B|^{-1/2}$ or A has a density increment of strength $[\alpha, 1]$.*

In general, suppose we have such a lemma with either $\alpha \ll |B|^{-1/2}$ or A has a density increment of strength $[\delta, d]$. As above (starting with $A \subseteq G$ with density α) we can iterate this $\lesssim \delta^{-1}$ many times, and then halt with a Bohr set of rank $\lesssim \delta^{-1} d$. The width of this Bohr set is like $\exp(-\tilde{O}(\delta^{-1}))$. We halt with a Bohr set of size

$$N' \gg \exp(-\tilde{O}(\delta^{-2} d))N$$

such that the $\alpha \ll (N')^{-1/2}$. We can assume that $\alpha \gg N^{-1/4}$ (or we are already done), and hence $N' \ll N^{1/4}$, which means $\delta^{-2} d \gtrsim \log N$.

Notice the key difference with the $\mathbb{F}_p^n$ case! There we ended up with $\delta^{-1} d \gtrsim \log N$, while for Bohr sets we have $\delta^{-2} d \gtrsim \log N$, because the width is also decaying, and we have lost a multiplicative factor in the width a further δ^{-1} many times. Let's see a couple of examples.

(1) Bourgain's $[\alpha, 1]$ implies $\alpha^{-2} \gtrsim \log N$, or $\alpha \lesssim 1/(\log N)^{1/2}$.
(2) A density increment of strength $[1, \alpha^{-1}]$ would yield $\alpha^{-1} \gtrsim \log N$, or $\alpha \lesssim 1/\log N$. Note that now these two cases are actually different, and this is better than the above.
(3) A very strong density increment of $[1, 1]$ would still imply $1 \gtrsim \log N$. In fact since now the $\gtrsim$ is only hiding a $\log(1/\alpha)^2$ factor this would imply the very strong density bound of $\alpha \leq \exp(-O(\sqrt{\log N}))$, which matches Behrend's lower bound.

(4) A density increment of strength $[1, (\log(1/\alpha))^{O(1)}]$ would already imply $\alpha \leq \exp(-(\log N)^c)$ for some constant $c > 0$. This is exactly the increment obtained in the recent breakthrough of Kelley and Meka, which we will discuss later in the course.

I will now return to the model setting of $\mathbb{F}_p^n$, but hopefully this discussion explains why a density increment of strength $[1, \alpha^{-1}]$ is better than $[\alpha, 1]$, even if they give identical results in $\mathbb{F}_p^n$.

7. An improved density increment

In this section I will sketch how to go about getting an increment of strength $[1, \alpha^{-1}]$ (under the usual assumptions of no three-term arithmetic progressions).

We begin as before: under the assumption that $A \subseteq \mathbb{F}_p^n$ has no non-trivial three-term arithmetic progressions and $\alpha \gg N^{-1/2}$ we deduce that, with $f_A = 1_A - \alpha$,

$$\mathop{\mathbb{E}}_{\gamma} |\widehat{f_A}(\gamma)| |\widehat{1_A}(\gamma)|^2 \gg \alpha \left|A\right|^2.$$

In Meshulam's proof we then noted that by Parseval's identity $\mathbb{E}_\gamma |\widehat{1_A}(\gamma)|^2 = |A|$, whence there must exist some γ such that $|\widehat{f_A}(\gamma)| \gg \alpha \left|A\right|$. Again recall that the Fourier transform of f_A is in fact equal to that of $\widehat{1_A}$ everywhere except the trivial character, where it is equal to 0.

This part of Meshulam's argument can therefore be summarised as: if $A \subseteq \mathbb{F}_p^n$ has no non-trivial three-term arithmetic progressions and $\alpha \gg N^{-1/2}$ then there is a non-trivial character γ such that the Fourier coefficient is large, $|\widehat{1_A}(\gamma)| \gg \alpha \left|A\right|$. We will now consider not just a single such character, but a whole collection, and so introduce the following definition.

Definition 6 (Large spectrum)**.** Let $A \subseteq G$ and $\eta \in [0, 1]$. We define the η-large spectrum of A as

$$\Delta_\eta(A) = \{\gamma \not\equiv 1 : |\widehat{1_A}(\gamma)| \geq \eta \left|A\right|\}.$$

For example note that $\Delta_0(A) = \widehat{G} \backslash \{1\}$ and $\Delta_1(A)$ is the collection of γ such that $\gamma(a) = 1$ for all $a \in A$. The following simple fact is very important to bear in mind.

Lemma 15. *If $A \subseteq G$ with density α and $\eta \in [0, 1]$ then*

$$|\Delta_\eta(A)| \leq \eta^{-2}\alpha^{-1}.$$

Proof.

$$\eta^2 \left|A\right|^2 |\Delta_\eta(A)| \leq \sum_\gamma |\widehat{1_A}(\gamma)|^2 = \left|A\right| N$$

by Parseval's theorem (or just expand out and use orthogonality). Now rearrange. □

Meshulam's argument can be phrased as "$\Delta_{c\alpha}(A)$ is not empty" (for some constant $c > 0$). Can we do better? Yes! Recall that we actually know that

$$\mathop{\mathbb{E}}_{\gamma} |\widehat{f_A}(\gamma)| |\widehat{1_A}(\gamma)|^2 \gg \alpha \left|A\right|^2.$$

This is a lot stronger than just "$\Delta_\alpha(A)$ is not empty". Let's explore why. First of all we'll undo our normalisation and replace $\widehat{f_A}$ by $\widehat{1_A}$ to get

$$\sum_{\gamma\not\equiv 1}\lvert\widehat{1_A}(\gamma)\rvert^3 \gg \lvert A\rvert^3 .$$

Now the contribution to the left-hand side from $\gamma\not\in\Delta_{c\alpha}(A)$ is

$$\leq c\alpha\,\lvert A\rvert\sum_{\gamma}\lvert\widehat{1_A}(\gamma)\rvert^2 = c\,\lvert A\rvert^3 ,$$

and in particular if we choose the constant $c>0$ small enough, then we deduce that

$$\sum_{\gamma\in\Delta_{c\alpha}(A)}\lvert\widehat{1_A}(\gamma)\rvert^3 \gg \lvert A\rvert^3 .$$

Again, note this in particular implies that the left-hand side is not zero, so $\Delta_{c\alpha}(A)$ is not empty, which is where Meshulam (and indeed Roth and Bourgain) halt.

But giving a good lower bound for this sum is saying a lot more than just saying it's not zero. We apply dyadic pigeonholing to turn this into something nicer; note that since for $\gamma\in\Delta_{c\alpha}(A)$ we have

$$\alpha\,\lvert A\rvert \ll \lvert\widehat{1_A}(\gamma)\rvert \leq \lvert A\rvert$$

there are only $\log(1/\alpha)\lesssim 1$ many dyadic scales that $\lvert\widehat{1_A}(\gamma)\rvert/\lvert A\rvert$ can live on. Choose a single dyadic scale on which the sum is $\gtrsim \lvert A\rvert^3$ – this means we have chosen some $\alpha\ll\eta\ll 1$ such that if

$$\Delta = \{\gamma\not\equiv 1 : \lvert\widehat{1_A}(\gamma)\rvert\in[\eta,2\eta]\,\lvert A\rvert\}$$

then

$$\sum_{\gamma\in\Delta}\lvert\widehat{1_A}(\gamma)\rvert^3 \gtrsim \lvert A\rvert^3 .$$

But since $\lvert\widehat{1_A}(\gamma)\rvert^3 \asymp \eta^3\,\lvert A\rvert^3$ on Δ this is equivalent to saying $\lvert\Delta\rvert\gtrsim\eta^{-3}$. Now we observe that $\Delta\subseteq\Delta_\eta(A)$. We have proved the following.

Lemma 16. *If $A\subseteq G$ has density α and no non-trivial three-term arithmetic progressions then either $\alpha\ll N^{-1/2}$ or there is some $\alpha\ll\eta\leq 1$ such that*

$$\lvert\Delta_\eta(A)\rvert\gtrsim\eta^{-3}.$$

To see why this is stronger, think about what happens at the two extremes, where $\eta\approx 1$ or $\eta\approx\alpha$. In the former our lower bound is not great, only $\gtrsim 1$, but in particular $\Delta_\eta(A)$ is not empty – so we have found a non-trivial charater γ such that $\lvert\widehat{1_A}(\gamma)\rvert\gg\lvert A\rvert$. This is much better than $\gg\alpha\,\lvert A\rvert$, and in particular the averaging argument we saw before immediately gives a density increment of strength $[1,1]$, fantastic!

So we are very happy if $\eta\approx 1$. What if $\eta\approx\alpha$? Now our Fourier coefficients are much smaller, possibly only of size $\asymp\alpha\,\lvert A\rvert$, the same size as Meshulam found. But now we know much more – there exists not just one of them, but lots, α^{-3} as many.

But we need to find a way to make use of this fact – we want to find a density increment that makes use of not just one large Fourier coefficient, but a whole bunch at once.

Lemma 17. *Let $A \subseteq G$ with density α and $\eta \in [0,1]$. If $\Delta \subseteq \Delta_\eta(A)$ and $V = \Delta^\perp$ is the subspace which annihilates Δ then there exists some x such that*

$$\frac{|(A-x)\cap V|}{|V|} \geq (1+\eta^2\left|\Delta\right|)\alpha.$$

In particular, A has a density increment of strength $[\eta^2\left|\Delta\right|, \dim(\Delta)]$.

Here $\dim(\Delta)$ is the dimension of the space spanned by Δ (in $\mathbb{F}_p^n$). There is a more robust notion that makes sense in any group, which we will examine below.

Proof. Since $\gamma(x) = 1$ for all $\gamma \in \Delta$ and $x \in V$ we know that $\widehat{1_V}(\gamma) = |V|$ for $\gamma \in \Delta$. This means that

$$\eta^2\left|A\right|^2\left|\Delta\right|\left|V\right|^2 \leq \sum_{\gamma\in\Delta}|\widehat{1_A}(\gamma)|^2|\widehat{1_V}(\gamma)|^2.$$

Note that Δ does not contain the trivial character, and so we can add this contribution on, and deduce that

$$(1+\eta^2\left|\Delta\right|)\left|A\right|^2\left|V\right|^2 \leq \sum_{\gamma}|\widehat{1_A}(\gamma)|^2|\widehat{1_V}(\gamma)|^2.$$

Apply Parseval's identity to go back to physical space (recalling that $\widehat{1_A\circ 1_A} = |\widehat{1_A}|^2$) this is saying that

$$(1+\eta^2\left|\Delta\right|)\alpha\left|A\right|\left|V\right|^2 \leq \langle 1_A\circ 1_A, 1_V\circ 1_V\rangle = \langle 1_A\circ 1_V, 1_A\circ 1_V\rangle.$$

But the right-hand side is trivially at most

$$\|1_A\circ 1_V\|_\infty\|1_A\circ 1_V\|_1 = \|1_A\circ 1_V\|_\infty\left|A\right|\left|V\right|.$$

Therefore

$$\|1_A\circ 1_V\|_\infty \geq (1+\eta^2\left|\Delta\right|)\alpha\left|V\right|.$$

But the left-hand side is exactly $1_A\circ 1_V(x) = |(A-x)\cap V|$ for some x, and we are done. □

If we use this lemma with the information that 16 provides (ignoring log factors) then we deduce that A has a density increment of strength $[\eta^{-1}, \eta^{-3}]$. Here we have used the trivial fact that $\dim(\Delta) \leq |\Delta|$. Note that the size of the density increment is actually very good – we have $\alpha \mapsto (1+\eta^{-1})\alpha$. The downside is the codimension cost. In particular if $\eta \approx \alpha$ (as is indeed possible) then this is like $[\alpha^{-1}, \alpha^{-3}]$ which quantitatively (ignoring log factors) is no better than $[1, \alpha^{-3}]$. This iterated would only lead to a bound for the density of sets without three-term progressions like $\ll 1/(\log N)^{1/3}$.

To do better we will first see whether we can improve on the trivial fact that $\dim(\Delta) \leq |\Delta|$. For general sets, of course, this cannot be improved – we can take the entirety of Δ to be linearly independent elements in $\mathbb{F}_p^n$. A crucial fact is that, when Δ is a spectrum, there is an improvement available.

Lemma 18 (Chang's lemma)**.** *If $A \subseteq G$ with density α and $\eta \in [0,1]$ then*

$$\dim(\Delta_\eta(A)) \ll \eta^{-2}\log(1/\alpha).$$

In particular $\dim(\Delta) \lesssim \eta^{-2}$. This is much better than the cruder $\dim(\Delta) \leq |\Delta| \lesssim \eta^{-3}$ that we used above, and in particular leads to a density increment of the strength $[\eta^{-1}, \eta^{-2}]$, which in the worst case looks like $[1, \alpha^{-2}]$.

Proof. Let $\Gamma \subseteq \Delta$ be a maximal linearly independent subset. It suffices to bound the size of Γ. This falls into two stages: we show that any subset of Δ has many solutions to linear equations, and compare that to the fact that Γ can have (thanks to linear independence) only trivial solutions.

We first define, for any $m \geq 1$, the m-fold additive energy of a set

$$E_{2m}(X) = \{(x_1, \ldots, x_{2m}) \in X^{2m} : x_1 + \cdots + x_m = x_{m+1} + \cdots + x_{2m}\}.$$

(Note that this makes sense for any finite set in an abelian group, in particular for any set of characters.) We have the trivial bounds

$$\left|X\right|^m \leq E_{2m}(X) \leq \left|X\right|^{2m-1},$$

the lower bound coming from the trivial solutions where $x_{m+i} = x_i$ for $1 \leq i \leq m$.

We claim that, for any $m \geq 1$ and $\Delta' \subseteq \Delta$,

$$E_{2m}(\Delta') \geq \eta^{2m}\alpha \left|\Delta'\right|^{2m}.$$

For example, with $m = 1$ this is saying that

$$E_2(\Delta) \geq \eta^2\alpha \left|\Delta\right|^2.$$

The left-hand side is just counting the number of $\gamma_1 = \gamma_2$ with $\gamma_i \in \Delta$, which is of course $|\Delta|$, and therefore this is reproving that $|\Delta| \leq \eta^{-2}\alpha^{-1}$. Chang's lemma is taking advantage of the fact that this holds even as $m \to \infty$.

Before proving this general bound we use it to deduce Chang's lemma. The key observation is that there are no non-trivial linear dependencies between elements of Γ, and hence

$$E_{2m}(\Gamma) \leq m! \left|\Gamma\right|^m \leq m^m \left|\Gamma\right|^m.$$

Comparing this to the lower bound above, for any $m \geq 1$,

$$|\Gamma| \leq \eta^{-2} m \alpha^{-1/m}.$$

In particular choosing $m = \lceil \log(1/\alpha) \rceil$ this is $\ll \eta^{-2}\log(1/\alpha)$ as required.

It remains to prove the energy lower bound. The starting point is the fact that

$$\eta \left|A\right| \left|\Delta'\right| \leq \sum_{\gamma \in \Delta'} |\widehat{1_A}(\gamma)| = \sum_{\gamma \in \Delta'} \left| \sum_{a \in A} \gamma(a) \right|.$$

We would like to change the order of summation here, but the absolute value prevents us. Undeterred, we just add in a sign and do it anyway – let c_γ be such that $c_\gamma \widehat{1_A}(\gamma) = |\widehat{1_A}(\gamma)|$. Therefore

$$\eta \left|A\right| \left|\Delta'\right| = \sum_{\gamma \in \Delta'} c_\gamma \widehat{1_A}(\gamma) = \sum_{a \in A} \sum_{\gamma \in \Delta'} c_\gamma \gamma(a).$$

We now apply Hölder's inequality to bound the right-hand side above by

$$\left|A\right|^{1-1/2m} \left(\sum_{a \in A} \left| \sum_{\gamma \in \Delta'} c_\gamma \gamma(a) \right|^{2m} \right)^{1/2m}.$$

Taking $2m$-powers and rearranging we have

$$\eta^{2m} \left|A\right| \left|\Delta'\right|^{2m} \leq \sum_{a \in A} \left| \sum_{\gamma \in \Delta'} c_\gamma \gamma(a) \right|^{2m}.$$

We now exploit the fact that the right-hand side is a sum of non-negative terms to remove the restriction that $a \in A$:

$$\eta^{2m} |A| |\Delta'|^{2m} \leq \sum_{x \in G} \left| \sum_{\gamma \in \Delta'} c_\gamma \gamma(x) \right|^{2m} .$$

Now we expand out the power so that the right-hand side is equal to

$$\sum_{\gamma_1, \ldots, \gamma_{2m}} c_{\gamma_1} \cdots \overline{c_{\gamma_{2m}}} \sum_{x \in G} \gamma_1(x) \cdots \overline{\gamma_{2m}(x)}.$$

By orthogonality the inner sum is 0 if $\gamma_1 + \cdots - \gamma_{2m} = 0$ and N otherwise. Therefore

$$\eta^{2m} |A| |\Delta'|^{2m} \leq \sum_{\gamma_1, \ldots, \gamma_{2m} \in \Delta'} c_{\gamma_1} \cdots \overline{c_{\gamma_{2m}}} 1_{\gamma_1 + \cdots + \gamma_m = \gamma_{m+1} + \cdots + \gamma_{2m}}.$$

The right-hand side is almost the additive energy $E_{2m}(\Delta')$, except for the annoying sign terms. But we can just throw them away by the triangle inequality – their inclusion could only be introducing cancellation into the count, so the right-hand side is at most $E_{2m}(\Delta')$, and the proof is complete. □

It is an instructive exercise to explore where exactly this proof used that G was a group. More robust notions are available where G can be replaced by a Bohr set.

Using Chang's lemma we have shown that if a set A lacks three-term progressions then ($\alpha \ll N^{-1/2}$ or) A has a density increment of strength $[1, \alpha^{-2}]$. Our goal is to obtain a strength of $[1, \alpha^{-1}]$ – this leads to nothing new in the $\mathbb{F}_p^n$ model setting, but (after a translation into Bohr sets) would give a $\ll 1/(\log N)$ bound in the integers.

Can we improve Chang's lemma? An example of Ben Green shows that this is impossible – there do exist examples of sets A and η such that $|\Delta_\eta(A)| \approx \eta^{-2} \log(1/\alpha)$ and $\Delta_\eta(A)$ is linearly independent.

We will sketch Green's example (a so-called 'niveau set') in the simplest setting of $\mathbb{F}_2^n$ (when it is in fact a Hamming ball). The pleasing feature of this group is that elements can be identified with subsets of $[n]$. Identifying the dual group also with subsets of $[n]$, we have that (viewing γ, x as subsets of $[n]$)

$$\gamma(x) = (-1)^{|\gamma \cap x|}.$$

In particular there are n linearly independent characters, corresponding to the basis vectors in $\mathbb{F}_2^n$ or equivalently the singleton sets $\{i\}$, for each of which the character evaluates to

$$\{i\}(x) = (-1)^{1_{i \in x}}.$$

Note that

$$\binom{n-1}{k} - \binom{n-1}{k-1} = \frac{n-2k}{n} \binom{n}{k}.$$

In particular if we let A be the collection of all sets of size $\leq n/2 - \sqrt{n}$ then

$$\widehat{1_A}(i \in A) = \sum_{x \in A} 1_{i \in x} - \sum_{x \in A} 1_{i \notin x} = \sum_{k \leq n/2 - \sqrt{n}} \frac{n-2k}{n} \binom{n}{k} \geq \frac{1}{\sqrt{n}} |A| .$$

We have $|A| \gg 2^n$ (e.g. via standard concentration of measure, since we're estimating the probability that a random collection of n 0/1 has at most $n/2 - \sqrt{n}$ many 1s, which is at most one standard deviation less than the expected count) and so this produces a set of density $\alpha \gg 1$ and $\eta \approx 1/\sqrt{n}$ such that there are $n \asymp \eta^{-2}$ many independent elements in $\Delta_\eta(A)$. We leave as an exercise extending this construction to handle smaller densities α while achieving $\asymp \eta^{-2}\log(1/\alpha)$ many independent elements. (Hint: Take the previous set and add an orthogonal subspace.)

We cannot hope therefore for a simple win to improving density increment via improving Chang's lemma. An obvious weakness remaining is that our density increment was actually of strength $[\eta^{-1}, \eta^{-2}]$, and the first parameter is much stronger than we need. Can we use this to improve the second parameter? Examining Lemma 17 we see that it would suffice to product some $\Delta' \subseteq \Delta$ with $|\Delta'| \approx \eta^{-2}$ with smaller dimension. That is, can we improve Chang's lemma if we're prepared to pass to a smaller subset? The answer, fortunately, is yes.

Lemma 19 (Improved Chang's lemma). *If $A \subseteq G$ with density α and $\eta \in [0,1]$ then there there exists $\Delta \subseteq \Delta_\eta(A)$ such that*

$$\dim(\Delta) \ll \eta^{-1}\log(1/\alpha)$$

and

$$|\Delta| \ll \eta\,|\Delta_\eta(A)|\,.$$

(Note that this does not contradict Green's example.)

Proof. This is a sketch proof, sweeping away some unpleasant technical calculations.

Let $d \approx C\eta^{-1}\log(1/\alpha)$, for some constant $C > 0$. The idea is to choose d elements of $\Delta_\eta(A)$ uniformly at random, say Γ, and then look at the subset of $\Delta_\eta(A)$ which they span.

The plan is that if the conclusion fails then this Γ should be close to being independent, and therefore should have small energy – but this will contradict the fact that as a random subset of the spectrum it should have large energy.

In particular, suppose that the conclusion fails, so every subset of $\Delta_\eta(A)$ of size at least $\eta\,|\Delta_\eta(A)|$ has dimension $> d$. What this means is that, when we were selecting our d random elements, at each stage there was a probability at least η that the next element was not in the span of the previous elements.

From this it follows that, for any k, the probability that Γ has dimension $d-k$ is at most the probability that k events with probability $\leq \eta$ occur in d independent trials, which is

$$\binom{d}{k}\eta^k \leq (d\eta)^k/k!.$$

Instead of considering the energy E_{2m} as in the proof of Chang's lemma, we will consider the 'restricted energy' $E^\sharp_{2m}(\Gamma)$, defined to be the count of $\gamma_1, \ldots, \gamma_{2m} \in \Gamma$ such that $\gamma_1 + \cdots + \gamma_m = \gamma_{m+1} + \cdots + \gamma_{2m}$ and $\gamma_i \neq \gamma_j$ for all $i \neq j$.

We first claim that if $\dim(\Gamma) = d-k$ then

$$E^\sharp_{2m}(\Gamma) \leq (2m)!4^k$$

Indeed, write $\Gamma = \Gamma_0 \sqcup \Gamma_1$ with Γ_0 linearly independent and $|\Gamma_0| = d-k$. We write a tuple counted by $E^\sharp_{2m}$ as

$$\pm\gamma_1 + \cdots \pm \gamma_r = \pm\lambda_1 \pm \cdots \pm \lambda_{2m-r}$$

There are at most 4^k possible choices a tuple of λ_i (since they must be distinct). Once these are fixed the elements appearing on the left-hand side is also fixed by linear independence of Γ_0, and hence we only need to take the ordering into account, which is a factor of $(2m)!$.

Summing over all possible values of k we deduce that

$$\mathbb{E}\, E^{\sharp}_{2m}(\Gamma) \leq \sum_k \frac{(d\eta)^k}{k!}(2m)!4^k \leq m^{2m}e^{4d\eta} \leq m^{2m}\alpha^{-O(1)},$$

say. On the other hand, by linearity of expectation,

$$\mathbb{E}\, E^{\sharp}_{2m}(\Gamma) \approx \left(\frac{d}{|\Delta|}\right)^{2m} E^{\sharp}_{2m}(\Delta).$$

(Note it is vital here that we were considering $E^{\sharp}_{2m}$ rather than E_{2m} here, so that the probability that a given $2m$-tuple lies in Γ is $(d/\left|\Delta\right|)^{2m}$.)

It follows that

$$E^{\sharp}_{2m}(\Delta) \ll m^{2m}\alpha^{-O(1)}d^{-2m}\left|\Delta\right|^{2m}.$$

We now make the major simplifying assumption that $E^{\sharp}_{2m}(\Delta) \approx E_{2m}(\Delta)$. This is definitely not true in general! For a full proof one must decompose the full energy E_{2m} into a weighted sum of smaller restricted energies $E^{\sharp}_{2t}$ for $t \leq m$ and perform this argument for each summand.

Assuming this simplifying assumption though, we deduce that

$$\alpha^{-O(1)}m^{2m}d^{-2m}\left|\Delta\right|^{2m} \gg E_{2m}(\Delta) \geq \eta^{2m}\alpha\left|\Delta\right|^{2m}.$$

Taking mth roots and rearranging we deduce that

$$\alpha^{-O(1/m)}m\eta^{-1} \gg d.$$

Taking m some multiple of $\log(1/\alpha)$ and recalling d is a large multiple of $\eta^{-1}\log(1/\alpha)$ results in a contradiction, and we are done. □

Employing this improved Chang's lemma we obtain a density increment of strength $[1,\alpha^{-1}]$ and thence a density bound of $\ll 1/\log N$, as required. In particular a robust version of this proof over Bohr sets delivers the same result for the integers.

CHAPTER 5

Lecture Six

8. Bateman-Katz improvement

In 2012 Bateman and Katz delivered the first improvement on Meshulam's bound.

Theorem 7 (Bateman-Katz)**.** *If $A \subseteq \mathbb{F}_3^n$ contains no non-trivial three-term arithmetic progressions then*

$$|A| \ll \frac{3^n}{n^{1+c}},$$

where $c > 0$ is some small constant.

The first part of their proof is standard, and follows along the same lines as above: if A has no non-trivial three-term arithmetic progressions then (either $\alpha \ll N^{-1/2}$ or) there is some $1 \geq \eta \gg \alpha$ such that

$$|\Delta_\eta(A)| \gtrsim \eta^{-3}.$$

To obtain a bound like $\ll 1/(\log N)^{1+c}$ we have seen that it suffices to obtain from this a density increment of strength $[1, \alpha^{-1+c}]$ for some constant $c > 0$. The previous argument utilising improved Chang's lemma obtains a density increment of strength $[1, \eta^{-1}]$, so we are done already if this large spectrum occurs at $\eta \gg \alpha^{1-c}$.

I'll now sketch what we do in the hardest case, when $\eta \approx \alpha$. This means we have

$$|\Delta_\alpha(A)| \approx \alpha^{-3}$$

(up to logarithmic factors). The first observation is that we have already obtained a density increment of strength $[1, \alpha]$, and only need to improve upon this slightly – in particular, if we can improve the statement of Chang's lemma (or improved Chang's lemma) even a little then we are done. The only thing these proofs required was the energy estimate

$$E_{2m}(\Delta) \geq \eta^{2m}\alpha \left|\Delta\right|^{2m}.$$

In particular, and this is the first main idea, we can actually assume that this lower bound is approximately also an upper bound – for if some energy was larger then we could leverage this into a better version of Chang's lemma and be done.

Therefore we may assume that $E_{2m}(\Delta) \approx \eta^{2m}\alpha \left|\Delta\right|^{2m}$ for all $m \geq 2$. For example (recalling $\eta \approx \alpha$ and $abs\Delta \approx \alpha^{-3}$)

$$E_4(\Delta) \approx \eta^4\alpha \left|\Delta\right|^4 \approx \alpha^2 \left|\Delta\right|^3 \approx \left|\Delta\right|^{2+1/3}.$$

On the face of it, this is not particularly useful information – we know that $E_4(\Delta) \in [\left|\Delta\right|^2, \left|\Delta\right|^3]$ and this falls in the middle of this range, far away from any extremes where might hope to gain some kind of information about what Δ looks like. Using E_4 alone this is true.

But Bateman and Katz observed that the knowledge of $E_8(\Delta)$ combined with $E_4(\Delta)$ does allow us to say something. We know that

$$E_8(\Delta) \approx \alpha^6 \left|\Delta\right|^7 .$$

In particular, if we normalise $e_4(\Delta) = E_4(\Delta)/\left|\Delta\right|^3$ and $e_8(\Delta) = E_8(\Delta)/\left|\Delta\right|^7$ then

$$e_4^3 \approx e_8.$$

Why is this significant? Because for any set Δ we always have

$$e_4^3 \leq e_8.$$

Here is a simple proof using two applications of the Cauchy-Schwarz inequality:

$$\begin{aligned} E_4(\Delta)^4 &= \langle 1_\Delta \circ 1_\Delta, 1_\Delta \circ 1_\Delta\rangle^4 \\ &= \langle 1_\Delta, 1_\Delta * 1_\Delta \circ 1_\Delta\rangle^4 \\ &\leq \left|\Delta\right|^2 \langle 1_\Delta * 1_\Delta \circ 1_\Delta, 1_\Delta * 1_\Delta \circ 1_\Delta\rangle^2 \\ &= \left|\Delta\right|^2 \langle 1_\Delta \circ 1_\Delta, 1_\Delta * 1_\Delta \circ 1_\Delta \circ 1_\Delta\rangle^2 \\ &\leq \left|\Delta\right|^2 E_4(\Delta)E_8(\Delta). \end{aligned}$$

Now we do know something interesting about our spectrum Δ – it is a set where this inequality is nearly sharp. In general, whenever you see an inequality nearly sharp, you might expect some kind of inverse result.

That there is something interesting you can say about sets with $e_4^3 \approx e_8$ is the main idea behind the proof of Bateman and Katz.

9. Additively non-smoothing sets

Let Δ be a set such that $E_4(\Delta) = \tau\left|\Delta\right|^3$. As we have seen, two applications of the Cauchy-Schwarz inequality imply that $E_8(\Delta) \geq \tau^3\left|\Delta\right|^7$. Bateman and Katz call sets which almost achieve this lower bound 'non-smoothing sets': the idea is that these are sets where looking at 4-fold sums doesn't give us any more interesting information than looking at the 2-fold sums did. In other words, these sets do not 'smooth out' under repeated addition.

The classic example of a *smoothing* set is a random set – if Δ is a random subset of a subspace then, with high probability, $\Delta + \Delta$ fills out the entire subspace, and in particular is a much more structured set than Δ. The idea is that knowing our set is *non-smoothing* should therefore be telling us that Δ is non-random in some way (which in turn we eventually hope to exploit as a density increment somehow).

To see what kind of information we should hope for we will consider two examples of non-smoothing sets. First a silly example: if Δ is a subspace, or in general a very structured set, then it must be non-smoothing, since $1 \approx e_4 \approx e_4^3 \approx e_8$. As we have seen, however, when Δ is the spectrum we're interested in we know that $e_4 \approx \left|\Delta\right|^{-2/3}$, so we need examples that allow for any values of e_4.

For the first example, consider

$$\Delta_1 = H \oplus D \quad\text{and}\quad \Delta_2 = \bigsqcup_{i=1}^{L} H_i,$$

where H and H_i are subgroups (and the H_i are all the same size, say $\left|H_i\right| \approx K$), and D is 'dissociated' in some appropriate sense. For Δ_1, we expect that $\Delta_1 + \Delta_1 =$

$H \oplus D \oplus D$. On H, we have $1_{\Delta_1} * 1_{\Delta_1} \approx |\Delta_1|$, and on the rest of $\Delta_1 + \Delta_1$, we have $1_{\Delta_1} * 1_{\Delta_1} \approx |H|$. Therefore

$$E_4(\Delta_1) \approx |H| \, |\Delta_1|^2 + |\Delta_1 + \Delta_1| \, |H|^2 \approx \frac{1}{|D|} |\Delta_1|^3 .$$

In particular, if $|D| \approx \tau^{-1}$ and $|H| \approx \tau |\Delta|$, then Δ_1 has $E_4(\Delta_1) \approx \tau |\Delta_1|^3$. Moreover, a similar calculation shows that

$$E_8(\Delta_1) \approx |H| \, (|H|^3 \, |D|^2)^2 + |\Delta_1 + \Delta_1 + \Delta_1 + \Delta_1| \, |H|^6 \approx \tau^3 |\Delta_1|^7 .$$

In particular, Δ_1 is additively non-smoothing. For Δ_2, on the other hand, assuming the H_i are 'spread out' enough that they do not additively interact much with each other,

$$\Delta_2 + \Delta_2 = \bigsqcup_{i=1}^{L} H_i \cup \bigsqcup_{1 \leq i \neq j \leq L} (H_i + H_j).$$

On the first part, which has size $|\Delta_2|$, we have $1_{\Delta_2} * 1_{\Delta_2} \approx K$. On the second part, which has size $L^2K^2 \approx |\Delta_2|^2$, we have $1_{\Delta_2} * 1_{\Delta_2} \approx 1$. Therefore

$$E_4(\Delta_2) \approx K^2 |\Delta_2| + |\Delta_2|^2 .$$

In particular, if $L \approx \tau^{-1/2}$, then $E_4(\Delta_2) \approx \tau |\Delta|^3$. Similarly,

$$E_8(\Delta_2) \approx K^6 |\Delta_2| + |\Delta_2|^4 \approx \tau^3 |\Delta_2|^7 ,$$

and hence Δ_2 is also additively non-smoothing.

Note that Δ_1 and Δ_2, although highly structured sets, have qualitatively different kinds of structure. The former is the union of $\approx \tau^{-1}$ many cosets, each of which is a translate of the same subgroup, while the latter is the union of $\approx \tau^{-1/2}$ many cosets, each of which comes from a different subgroup, which do not interact much.

The philosophy behind the structural results for additive non-smoothing sets is that these two kinds of structure (and natural interpolations between the two) are the only ways that a set can be additively non-smoothing.

To motivate the form the structural theorem takes, note that in either construction there is a set $X \subset \Delta$ and a subgroup $H \subset \Delta$ such that $|X| \, |H| \approx \tau |\Delta|^2$ and $|X + H| \ll |X|$. Indeed, for Δ_1 we take $X = \Delta_1$ and H to be the H in its construction, and for Δ_2 we take $X = H = H_i$ for some arbitrary $1 \leq i \leq L$.

Theorem 8. *If $e_4(\Delta) = \tau$ and $e_8(\Delta) \ll \tau^3$ then there are $X, H \subseteq \Delta$ such that*

(1)
$$|H| \, |X| \asymp \tau |\Delta|^2 ,$$

(2)
$$|H + H| \ll |H| ,$$

and

(3)
$$|X + H| \ll |X| .$$

Of course this is only an informal statement (in fact in the actual statement one proves a result about the additive energies, which can then be convered into this form using standard tools from additive combinatorics).

The way to read this conclusion is that we find some sets H, X in Δ such that H is very structured (behaves like a subspace) and X is very structured under translates from H – but X itself may not be structured!

10. Back to density increments

Let us apply this to our spectral information to see what it says. We have $\tau \approx \alpha^2$ now, so applying this structural result produces $X, H \subseteq \Delta = \Delta_\alpha(A)$ so that H is very structured and $\lvert H\rvert \lvert X\rvert \gg \alpha^2 \lvert\Delta\rvert^2 \approx \alpha^{-4}$.

We now recall that our density increment procedure produces, for any set $\Delta' \subseteq \Delta$, an increment of strength

$$[\alpha^2 \lvert\Delta'\rvert, \dim(\Delta')].$$

Our initial goal was to produce an increment of strength $[1, \alpha^{1-c}]$, but in the model setting $\mathbb{F}_p^n$ we would also be happy with an increment of the strength $[\alpha^{1-c}, 1]$. Since H is a very structured set of size $\alpha^{-O(1)}$ we can assume that it has dimension $O(1)$ (up to logarithmic factors). In particular, if $\lvert H\rvert \gg \alpha^{-1-c}$ then we are done.

The hardest case remaining therefore is when $\lvert H\rvert \approx \alpha^{-1}$, whence $\lvert X\rvert \approx \alpha^{-3}$, which is the size of the full spectrum, so we may as well assume that $X \approx \Delta$.

To summarise our discussion so far: either we have an increment good enough to get our result, or we have found some very structured set $H \subseteq \Delta$ of size $\approx \alpha^{-1}$ such that $H + \Delta \approx \Delta$.

If we pretend that H is a subspace, this basically means that Δ is the union of $\approx \alpha^{-2}$ disjoint cosets of H. Finishing off the proof now is a little delicate, since one must deduce a strong increment from this information. The way that Bateman and Katz proceed is via 'quotienting out' the entire space $\mathbb{F}_p^n$ by the subspace H, and then a localised Parseval-type argument finds a single large Fourier coefficient, and hence an increment of strength $[\alpha^{1-c}, 1]$.

This suffices to achieve a bound like $\alpha \ll 1/(\log N)^{1+c}$ in $\mathbb{F}_p^n$. As we have seen, however, even assuming that the above is robust enough to be translated to Bohr sets, an increment of strength $[\alpha^{1-c}, 1]$ would only obtain a bound of $\alpha \ll 1/(\log N)^{1/2+c}$.

To achieve a bound like $\alpha \ll 1/(\log N)^{1+c}$ we need to convert this final situation into an increment of strength $[1, \alpha^{-1+c}]$. With Olof Sisask we found a method that does this, which we call 'spectral boosting'. The idea of spectral boosting is to that a subset of a spectrum with an unusually large amount of additive structure is forced to have some translate lie in a spectrum of a higher level. That is, the 'spectral level' of a set is automatically 'boosted' by its inherent additive structure.

In our set-up, we have some $H \subseteq \Delta = \Delta_\alpha(A)$ of size $\approx \alpha^{-1}$ and constant dimension such that $\Delta + H \approx \Delta$. Spectral boosting allows us to essentially assume that (some translate of) H actually behaves like a subset of $\Delta_{\alpha^{1/2}}(A)$, which means we get a much stronger increment of

$$[\alpha \lvert H\rvert, \dim(H)] = [1, 1].$$

This is more than enough, and performing this argument over Bohr sets results in the following.

Theorem 9 (Bloom-Sisask)**.** *If $A \subseteq \{1, \ldots, N\}$ has no non-trivial three-term arithmetic progressions then*

$$|A| \ll \frac{N}{(\log N)^{1+c}}$$

for some $c > 0$.

CHAPTER 6

Lecture Six

We have sketched a complicated spectral approach that uses density increment to get a little bit past the density threshold of $1/\log N$. Last year Kelley and Meka achieved an astounding breakthrough when they proved a much better bound – moreover, using hardly any spectral information at all!

Theorem 10 (Kelley-Meka)**.** *If $A \subseteq \{1, \ldots, N\}$ has no non-trivial three-term arithmetic progressions then*

$$|A| \ll \frac{N}{\exp(c(\log N)^{1/12})}$$

for some $c > 0$.

This is close to the Behrend lower bound, which has an exponent of 1/2 rather than 1/12. We will sketch the main ideas of their approach.

As usual, we will switch to the model setting of $\mathbb{F}_p^n$; you have already seen the tricks needed to convert this into the language of Bohr sets and hence obtain the original result (although the details of the conversion are quite tricky!). The driving force is the following.

Theorem 11. *If $A \subseteq \mathbb{F}_p^n$ has no non-trivial three-term arithmetic progressions then either $\alpha \ll N^{-1/2}$ or A has a density increment of strength $[1, \tilde{O}_\alpha(1)]$.*

We write $\tilde{O}_\alpha(1)$ to emphasise that this codimension loss is really $\log(1/\alpha)^{O(1)}$ – of course, if one cares about the exponent 1/12 then you need to track what this power actually is, but for our sketch we won't keep track, and will just ignore logarithmic factors.

Unlike all the approaches we have seen so far the argument of Kelley and Meka does not begin with writing things in Fourier space – indeed, Fourier space hardly enters their argument at all in any significant way. We will depart a little from their perpsective however, and begin in familiar territory.

11. From few 3APs to large L^p norm

Suppose that A has no non-trivial three-term arithmetic progressions. Then we have already seen that either $\alpha \ll N^{-1/2}$ or

$$\mathop{\mathbb{E}}_{\gamma} |\widehat{f_A}||\widehat{1_A}|^2 \gg \alpha \left|A\right|^2,$$

where $f_A = 1_A - \alpha$. Since $\widehat{f_A} = \widehat{1_A}$ except at the trivial character, where it is 0, we can add back in the trivial character to see that, for constant $c > 0$,

$$\mathop{\mathbb{E}}_{\gamma} |\widehat{1_A}(\gamma)|^3 \geq (1+c)\alpha \left|A\right|^2$$

We use the familiar sign trick to write the left-hand side as

$$\sum_{a\in A}\mathop{\mathbb{E}}_{\gamma} c_\gamma\gamma(a)\lvert\widehat{1_A}(\gamma)\rvert^2$$

for some signs $c_\gamma\in\mathbb{C}$ with $\lvert c_\gamma\rvert=1$. We now apply Hölder's inequality and orthogonality to deduce that, for any $m\geq 1$,

$$\alpha^{-1/2m}\left|A\right|\left(\mathop{\mathbb{E}}_{\gamma_1,\cdots,\gamma_{2m}} c_{\gamma_1}\cdots\overline{c_{\gamma_{2m}}}\lvert\widehat{1_A}(\gamma_1)\rvert^2\cdots\lvert\widehat{1_A}(\gamma_{2m})\rvert^2 1_{\gamma_1+\cdots-\gamma_{2m}=0}\right)^{1/2m}\geq(1+c)\alpha\left|A\right|^2.$$

If we choose $m\approx\log(1/\alpha)$ then $\alpha^{-1/2m}\leq 1+c/10$, say, and so (replacing c with a slightly smaller constant) using the triangle inequality we deduce that

$$\left(\mathop{\mathbb{E}}_{\gamma_1,\cdots,\gamma_p}\lvert\widehat{1_A}(\gamma_1)\rvert^2\cdots\lvert\widehat{1_A}(\gamma_{2m})\rvert^2 1_{\gamma_1+\cdots-\gamma_p=0}\right)^{1/p}\geq(1+c)\alpha\left|A\right|.$$

Note that this is exactly the same as the argument we used for Chang's lemma, except that now we're arguing with $\widehat{1_A}$ instead of 1_Δ. This has the advantage that we're not throwing away information anymore.

What now? We take advantage of the fact that $\lvert\widehat{1_A}\rvert^2$ is the Fourier transform of $1_A\circ 1_A$, so that

$$\lvert\widehat{1_A}(\gamma)\rvert^2=\sum_x 1_A\circ 1_A(x)\gamma(x).$$

. This means that the left-hand side can, by orthogonality, be written as

$$\sum_{x_1,\ldots,x_p}1_A\circ 1_A(x_1)\cdots 1_A\circ 1_A(x_p)\mathop{\mathbb{E}}_{\gamma_1,\ldots,\gamma_p}1_{\gamma_1+\cdots-\gamma_p=0}\gamma_1(x_1)\cdots\gamma_p(x_p).$$

By orthogonality this is

$$\mathop{\mathbb{E}}_{x\in G}1_A\circ 1_A(x)^p.$$

In other words, we have arrived at the pleasingly simple conclusion that (with $p\approx\log(1/\alpha)$) we have

$$\left(\mathop{\mathbb{E}}_{x\in G}1_A\circ 1_A(x)^p\right)^{1/p}\geq(1+c)\alpha\left|A\right|.$$

This is all the information that we will use to get a density increment. Kelley and Meka obtained it in an alternative way that uses less Fourier analysis, but this presentation shows the similarities to previous arguments.

For comparison, note that by Hölder's inequality

$$\left(\mathop{\mathbb{E}}_{x\in G}1_A\circ 1_A(x)^p\right)^{1/p}\geq\mathop{\mathbb{E}}_{x\in G}1_A\circ 1_A(x)=\alpha\left|A\right|.$$

We have found an improvement over this by a multiplicative factor of $1+c$. But what to do with it?

12. Sifting