ADDITIVE COMBINATORICS

THOMAS F. BLOOM

These are lecture notes for the Part III lecture course given in Lent Term 2021 at the University of Cambridge.

Two textbooks in particular have been useful in preparing these lecture notes:

- *Additive Combinatorics* by T. Tao and V. Vu, Cambridge University Press, 2006.
- *Combinatorial Number Theory and Additive Group Theory* by A. Geroldinger and I. Ruzsa, Birkhäuser Basel, 2009.

Much of the material in this course, however, has not appeared in any textbook, and is taken directly from the research literature. Any inaccuracies or mistakes are, of course, entirely my own.

If you have any questions, concerns, or corrections, please email me at bloom@maths.ox.ac.uk.

What is additive combinatorics?

Additive combinatorics is an old subject, although the term 'additive combinatorics' was first coined (to the best of my knowledge) by Tao and Vu in 2006, in their textbook of the same name. Results of an additive combinatorial flavour go back at least 100 years, although they were often thought of as belonging to additive/combinatorial number theory.

Additive combinatorics is, at heart, the study of combinatorial questions involving addition: What does counting information (counting the size of a set, or counting the number of sums that can be formed from that set) imply about algebraic structures (e.g. is the set a subgroup, or an arithmetic progression? Is it close to being such? Does it contain one?). It can be distinguished from number theory in that the assumptions we work from are often very mild - e.g. rather than dealing with special structured sets, such as the primes or the zeros of some polynomial, we are concerned with quite arbitrary sets.

Much of additive combinatorics can be viewed as the quest for understanding exactly what it means for a set (e.g. of integers) to be 'additively structured'. From the algebraic point of view, one might insist that a set be closed under addition, and hence be a subgroup. But there is a much richer collection of sets which are 'approximately structured'. For example, if we take some random 1% of a subgroup, is that still structured? How much?

There are a number of ways to measure how structured a set is, and we will explore the connection between these ways. Often the hard part is to go from some quite weak, statistical, measures (such as a set have many solutions to $a+b=c+d$) to those which embody much more rigid algebraic structure.

Structure of the course. Additive combinatorics uses many different methods from all across pure mathematics: combinatorics, probability, harmonic analysis, group theory, ergodic theory. We will not attempt to cover all, or even most, of the many different flavours of modern additive combinatorics. Two notable omissions, for example, are anything from the ergodic theoretic point of view, and anything concerning the theory of 'higher order Fourier analysis'.

Our aim is to present the 'blunt edge' of research in that part of modern additive combinatorics which uses analytic techniques (either Fourier analysis or analysis of an 'elementary' nature) to answer quantitative questions. Although we will not give the best-known results, we will present the same kind of methods and techniques that are used.

The material in this course has been divided into four chapters. Chapter 2 is essentially independent of the other three, but the other three have some interdependency.

1. Elementary techniques
2. Fourier analysis and Roth's theorem
3. Almost-periodicity
4. Inverse sumset results

Prerequisites. There are no formal prerequisites for this course, aside from a certain amount of mathematical maturity. We will be using finite Fourier analysis and basic probability theory, but both entirely over finite objects, and so there are no analytic technicalities to plague us here. Everything needed will be developed from scratch in a 'low-brow' way suitable for our applications – in particular, you should not avoid this course just because of an aversion to Fourier analysis as you might have seen it elsewhere!

CHAPTER 1

Elementary tools

In this chapter we will introduce some basic concepts of additive combinatorics, and develop some useful tools to work with them. All our methods will be elementary (in particular, there is no Fourier analysis in this section).

Before we begin properly, we give a quick overview of some our basic assumptions and notational choices.

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 $f \ll g \ll f$.

Objects. All sets will be assumed finite and non-empty, unless explicitly stated otherwise. Our usual setting will be an unspecified abelian group, which we denote by G. By default, when finite, the size of G will be denoted by N. We will write $\mathbb{Z}/N\mathbb{Z}$ for the cyclic group of order N, although we may write $\mathbb{F}_p$ (by which we understand the additive group of the field) when we want to emphasise that the order is prime.

Although much of what we do will be valid for arbitrary finite subset of an abelian group, it might help if you pick your favourite group as an example to use throughout. The most popular choices are either

$$\mathbb{F}_2^n \quad \text{if you're a computer scientist),}$$

$$\mathbb{Z} \quad \text{(if you're a number theorist), or}$$

$$\mathbb{Z}/N\mathbb{Z} \quad \text{(if you're a number theorist who's uncomfortable with infinity).}$$

In practice, for the questions we will be considering, the latter two settings are equivalent, but there are often fundamental differences between $\mathbb{F}_p^n$ (with p fixed and $n \to \infty$) and $\mathbb{Z}/N\mathbb{Z}$. Proofs are often much cleaner in the former, in large part due to the many different subgroups available.

We use $\subset$ to include the case of equality (where others may write $\subseteq$).

Functions. We will usually adopt an analytic point of view, in particular often viewing sets $A \subset G$ 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

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

1. Sum sets

Given any pair of sets A, B we can define their sum set

$$A + B = \{a + b : a \in A \text{ and } b \in B\},$$

or difference set

$$A - B = \{a - b : a \in A \text{ and } b \in B\}.$$

It is important to note that these are operations on sets, and do not obey any nice algebraic laws - in particular they do not cancel, so $(A + B) - B \neq A$ (in general). Rather, they should be viewed as combinatorial objects whose structure is influenced by the additive structure of A and B. Of course, the difference set is just a special case of the sum set construction, and could instead be viewed as $A + (-B)$.

We have the trivial inequalities

$$|A| \leq |A + B| \leq |A| \, |B| \, .$$

Indeed, for any fixed $b \in B$, the set $\{a + b : a \in A\}$ has size $|A|$ and is contained in $A + B$, and there is an obvious surjection from $A \times B \to A + B$ with $(a, b) \mapsto a + b$.

In the special case $A = B$ the upper bound can be improved since this surjection maps (a_1, a_2) and (a_2, a_1) to the same element, yielding

$$|A + A| \leq \frac{|A| \, (|A| + 1)}{2}.$$

Both these trivial lower and upper bounds are sharp. For example, if A is itself a group, then $A + A = A$, and so $|A + A| = |A|$. On the other hand, if A is such that all its pairwise sums are distinct, for example $A = \{1, 2, 4, \ldots, 2^{n-1}\}$, then the upper bound is sharp. The case for equality for the lower bound is easy to characterise.

Lemma 1. *$|A+A|=|A|$ if and only if A is a coset of a subgroup.*

Proof. Since both properties (the size of the sumset and being a coset) are invariant under translation, we can without loss of generality assume that $0\in A$, and so $A\subset A+A$, and so $A+A=A$. For any $a\in A$, there are $|A|$ many distinct translates $a+a'$ for $a'\in A$, all of which belong to $A+A=A$, and hence one of them must be 0, and so $a'=-a\in A$. Thus A is closed under sums and inverses, and is a subgroup. □

In groups where there are many non-trivial finite subgroups we have a plentiful supply of sets such that $|A+A|=|A|$. What about groups where there are no non-trivial finite subgroups? Then we know that if $|G|>|A|>1$ we must have $|A+A|>|A|$. It is a simple fact, but one with suprisingly far-reaching consequences, that for the integers something stronger can always be said.

Lemma 2. *If $A\subset\mathbb{Z}$ then*

$$|A+A|\geq 2\,|A|-1.$$

Equality holds if and only if A is an arithmetic progression.

Proof. Suppose the elements of A are ordered like $a_1<a_2<\cdots<a_n$. This induces an ordering on some elements of $A+A$:

$$2a_1<a_1+a_2<\cdots<a_1+a_n<a_2+a_n<\cdots<a_{n-1}+a_n<2a_n.$$

In particular, these $2n-1$ sums are all distinct, and hence $|A+A|\geq 2n-1$.

For the second part, it suffices to show that if equality holds then for any $1\leq i<n$, there is some $j>i$ such that $a_j-a_i=a_2-a_1$. Downwards induction on i then implies that in fact $j=i+1$, and hence if $d=a_2-a_1$ then $a_i=a_1+d(i-1)$, and A is an arithmetic progression as required. Consider, for some $2\leq i<n$, the sum $a_2+a_i\in A+A$. Since $|A+A|=2n-1$, the above $2n-1$ elements are the entirety of $A+A$, and so one of them must be a_2+a_i. Since $i<n$ we have $a_2+a_i<a_j+a_n$ for all $2\leq j\leq n$, and so $a_2+a_i=a_1+a_j$ for some $j>i$ as required. □

This proof is only valid for $\mathbb{Z}$, and uses the ordering of $\mathbb{Z}$ in a crucial way. It is natural to wonder if a similar result holds for other groups without non-trivial finite subgroups, in particular for $\mathbb{F}_p$. The answer is yes (once one incorporates the obvious constraint that the sumset cannot be larger than p), and is one of the early jewels of additive combinatorics.

Lemma 3 (Cauchy-Davenport[1])**.** *If $A,B\subset\mathbb{F}_p$ then*

$$|A+B|\geq\min(|A|+|B|-1,p).$$

Proof. 1We will show that the stated inequality is true for any fixed B and all $A\subset\mathbb{F}_p$ by induction on the size of B. The case $|B|=1$ is trivial. Suppose that $|B|\geq 2$, and let b_1,b_2 be two distinct elements of B, with $z=b_2-b_1\neq 0$. If $A+z\subset A$ then it immediately follows that $A+kz\subset A$ for all $k\geq 1$, and hence $A=\mathbb{F}_p$ (since any non-zero element of $\mathbb{F}_p$ additively generates the entire group), in which case the claim is trivial.

Thus $A+z\not\subset A$, and so there exists $a\in A$ such that $a+z\notin A$. It follows that, if $x=a-b_1$, then $B+x$ contains at least one element which is not in A (namely

[1] A strange title, given that Davenport was born 50 years after Cauchy died! This is its customary label; it was proved by Cauchy [2] in 1813 and independently rediscovered by Davenport [3] in 1935.

$b_2 + a - b_1$), and at least one element that is in A (namely $a = b_1 + a - b_1$ itself). Thus

$$1 \leq |A \cap (B + x)| < |B|.$$

We now note that, if $A' = (A - x) \cup B$ and $B' = A \cap (B + x)$, then

$$A' + B' \subset A + B.$$

(This transformation, from $(A, B) \mapsto (A', B')$, has found many more applications, and is sometimes called the 'Dyson e-transform'.)

Indeed, if $a' + b' \in A' + B'$ with $a' \in B$, then $b' \in A$ forces $a' + b' \in A + B$, while if $a' \in A - x$, then $b' \in B + x$ forces the same. Hence by the inductive hypothesis, valid since $1 \leq |B'| < |B|$,

$$|A + B| \geq |A' + B'| \geq \min(|A'| + |B'| - 1, p).$$

The proof is now completed by noting that

$$|A'| + |B'| = |A| + |B| - |A \cap (B + x)| + |B'| = |A| + |B|.$$

□

Just as in the integers, we can characterise the case of equality as arithmetic progressions: that is, if $|A + B| = |A| + |B| - 1$ then (aside from edge cases) A and B must both be arithmetic progressions of the same step. This is known as Vosper's theorem, and we leave a proof for the first examples sheet.

We say that A has 'small doubling' if $|A + A| \leq K|A|$ for some 'small' K (usually K being bounded by some absolute constant). We have now seen two examples of sets with small doubling: cosets of subgroups (where $K = 1$) and arithmetic progressions (where $K = 2 - \frac{1}{|A|}$).

Given any sets with small doubling it is easy to generate more, by two simple constructions: passing to a large subset and/or passing to a sumset with some small set. If $|A + A| \leq K|A|$ and $X \subset A$ then

$$|X + X| \leq |A + A| \leq \left(K\frac{|A|}{|X|}\right)|X|.$$

In particular, if $|X| \geq K^{-O(1)}|A|$, then A having doubling K implies X has doubling at most $K^{O(1)}$. Alternatively, if $X = A + Y$, then

$$|X + X| = |A + A + Y + Y| \leq |Y|^2 |A + A| \leq K|Y|^2 |A|.$$

Thus, if $|Y| \leq K^{O(1)}$, then A having doubling K implies $X = A + Y$ has doubling at most $K^{O(1)}$.

It is a remarkable fact, and one of the great achievements of additive combinatorics, that these trivial constructions account for all of the sets with small doubling, in a way which can be made quite quantitatively explicit. Results of this sort, that give some structural information about sets with small doubling, are known as inverse sumset results, and we have already seen a couple:

$$\text{if } |A + A| \leq |A| \text{ then } A \text{ is a coset of a subgroup,}$$

$$\text{if } A \subset \mathbb{Z} \text{ and } |A + A| < 2|A| \text{ then } A \text{ is an arithmetic progression.}$$

Unfortunately, these only given information when $K < 2$. The situation with larger K (e.g. $K = 100$) is much more complicated, but by the end of this course we will have proved an explicit inverse sumset result that gives non-trivial information about A even in the regime $K \approx \log|A|$, say.

1.1. **Sumset calculus.** There are a number of useful lemmas on sumsets available via elementary methods. At the heart of such proofs is often an additive triviality such as $a+b=b+a$ or $(a-c)+(c-b)=a-b$, but there is an enduring elegance in how these trivialities are exploited.

The collection of such lemmas is sometimes called the 'sumset calculus'. The two most important, and frequently used, tools in sumset calculus are Ruzsa's triangle inequality and Plünnecke's inequality. We begin with the former, which loosely says that if A and B both have small sumset compared to some other set C, then the sumset $A+B$ itself must be small.[2]

Lemma 4 (Ruzsa's triangle inequality [8])**.** *For any A, B, C,*

$$|A+B| \leq \frac{|A+C|\,|B-C|}{|C|}.$$

Proof. For each $x \in A+B$ fix an arbitrary representation $x = a_x + b_x$, and consider the map

$$(c, x) \mapsto (c + a_x, b_x - c).$$

This is a well-defined function from $C \times (A+B)$ to $(A+C) \times (B-C)$. Furthermore, since $(c+a_x)+(b_x-c) = a_x+b_x = x$, the x component can be uniquely recovered from its image. Then, however, we can recover a_x, and hence also recover c as $(c+a_x)-a_x$. Thus this function is an injection, and the proof is complete. □

Plünnecke's inequality addresses what happens if we iterate the sumset operation. That is, just as we can define $A+A$, we can define $A+A+A$, $A+A+A+A$, and so on, in the obvious fasion. In brief, we write kA for the k-fold sumset, so that, for example, $2A = A+A$. (Note that it does *not* mean A dilated by k!) If we know that A has small doubling then can we deduce that kA remains (relatively) small?

Plünnecke's inequality says yes we can! In its most usefully applied form, it says that if $|A+A| \leq K\,|A|$, then for any $k \geq 1$, $|kA| \leq K^k\,|A|$. In particular, if A has a small sumset, then it also has small iterated sumsets. The original proof of this by Plünnecke [7] is quite complicated, and can be found, for example, in [10, Section 6.5]. More recently, Petridis [6] has found an alternative beautiful proof, which is much shorter.

The key lemma in the approach of Petridis is the following.

Lemma 5 (Plünnecke-Petridis inequality [6])**.** *For any A and B there exists $X \subset A$ such that, for all C,*

$$\frac{|C+X+B|}{|C+X|} \leq \frac{|A+B|}{|A|}.$$

Proof. As a clue for where to start, observe that if the bound holds even when $|C| = 1$, then, whatever X is, it must satisfy

$$\frac{|X+B|}{|X|} \leq \frac{|A+B|}{|A|}.$$

It seems a reasonable choice, then, to choose some $X \subset A$ such that this inequality holds and the left hand side is as small as possible (note that the space of such X

[2] Note the obvious parallel with the metric triangle inequality, if one thinks of '$A+B$ is small' as a way of measuring the 'closeness' between A and B. This metric viewpoint is developed a little further in [10, Section 2.3].

is non-empty since $X = A$ suffices, so such a minimal X certainly exists). We will demonstrate that any such X works; in fact, we will prove the stronger claim that, for all C,

$$\frac{|C+X+B|}{|C+X|} \leq \frac{|X+B|}{|X|}.$$

To prove this, we use induction on the size of C. The case when $|C| = 1$ is trivial, so suppose that $|C| > 1$. Choose some arbitrary $c \in C$, and let $C' = C\backslash\{c\}$. Let $X_c \subset X$ be maximal such that $C' + X$ and $c + X_c$ are disjoint. By maximality, $c + (X\backslash X_c) \subset C' + X$, whence $c + (X\backslash X_c) + B \subset C' + X + B$ and

$$C + X + B \subset (C' + X + B) \cup ((c + X + B)\backslash(c + (X\backslash X_c) + B)).$$

Taking cardinalities of both sides, and using both the inductive hypothesis and minimality of $|X+B| / |X|$, we have

$$\begin{aligned} |C+X+B| &\leq |C'+X+B| + |X+B| - |(X\backslash X_c)+B| \\ &\leq \frac{|X+B|}{|X|}|C'+X| + |X+B| - \frac{|X+B|}{|X|}|X\backslash X_c| \\ &= \frac{|X+B|}{|X|}|C'+X| + \frac{|X+B|}{|X|}(|X| - |X\backslash X_c|) \\ &= \frac{|X+B|}{|X|}(|C'+X| + |X_c|) \\ &= \frac{|X+B|}{|X|}|C+X|, \end{aligned}$$

and the proof is complete. □

Since this holds for any set C, iterating it leads to the following useful corollary, and it is some form of this that is usually meant by invocations of Plünnecke's inequality.

Corollary 1 (Plünnecke's inequality)**.** *Suppose that $|A+B| \leq K|A|$. There exists $X \subset A$ such that, for all $h \geq 1$,*

$$|X + hB| \leq K^h |X|.$$

In particular,

$$|hB| \leq K^h |A|.$$

Proof. We will show by induction on h that the X given by Lemma 5 works. The case $h = 1$ is immediate from Lemma 5 taking C to be any singleton set. The general case follows by induction and Lemma 5 applied with $C = (h-1)B$.

The second conclusion follows immediately from the first since trivially $|hB| \leq |X + hB|$ and $|X| \leq |A|$. □

Ruzsa's triangle inequality immediately allows this result to be generalised to mixed sum and difference sets.

Corollary 2. *Suppose that $|A+B| \leq K|A|$. For any $k, l \in \mathbb{N}$ with $k + l \geq 2$,*

$$|kB - lB| \leq K^{k+l}|A|.$$

Proof. By Ruzsa's triangle inequality, with X being as in Corollary 1,

$$\left|kB-lB\right| \leq \frac{\left|X+kB\right|\left|X+lB\right|}{\left|X\right|} \leq \left(\frac{\left|A+B\right|}{\left|A\right|}\right)^{k+l}\left|X\right|.$$

□

2. Additive energy

The sumset is one useful way to measure structure, but not the only one. One disadvantage is that it is not very robust - for example, if we take a set A with small doubling, and add in just a few random elements, then the size of the sumset could increase massively.

It is often more convenient, especially in the kind of analytic arguments we will be using, to use instead the notion of additive energy.

Definition 1 (Additive Energy)**.** There are various ways to define this: one way is as a count of solutions to a symmetric linear equation in four variables:

$$E(A)=\#\{(a,b,c,d)\in A^4 : a+b=c+d\}.$$

Equivalently, one can write this as

$$E(A)=\sum_x\left(\sum_{a,b\in A}1_{a+b=x}\right)^2=\sum_x 1_A*1_A(x)^2.$$

That is, the additive energy is just the (squared) L^2 norm of the convolution 1_A*1_A. One can imagine using other norms here – indeed, the size of the sumset is precisely the size of the support of 1_A*1_A, which is sometimes called the L^0 norm (although it's not really a norm).

It is also useful to note that (since $a+b=c+d$ if and only if $a-c=d-b$)

$$E(A)=\#\{(a,b,c,d)\in A^4 : a-c=d-b\}=\sum_x 1_A\circ 1_A(x)^2.$$

Roughly speaking, just as "small doubling constant" (i.e. $\left|A+A\right|/\left|A\right|$ small) is one way to quantify how structured a set is, "large additive energy" serves the same purpose. Intuitively, we expect these properties to correlate, since if there are many 'collisions' of $a+b=c+d$ then we expect there to be few distinct sums $a+b$ inside the sumset $A+A$, and vice versa.

One direction of this is a straightforward consequence of the Cauchy-Schwarz inequality, which we record as a lemma, along with the trivial size bounds on the additive energy.

Lemma 6.

$$\left|A\right|^2 \leq E(A) \leq \left|A\right|^3.$$

Furthermore,

$$E(A)\geq \frac{\left|A\right|^4}{\left|A+A\right|}.$$

Proof. The lower bound comes from counting those solutions to $a+b=c+d$ where $a=c$ and $b=d$. The upper bound is true since once we have fixed any three $a,b,c\in A$ the choice of d such that $a+b=c+d$ is fixed.

To connect it with the size of the sumset, we use the Cauchy-Schwarz inequality. Namely, we note that

$$\left|A\right|^2 = \sum_x 1_A * 1_A(x),$$

since every pair $(a,b) \in A^2$ is a solution to exactly one equation of the type $a+b=x$. Since $1_A * 1_A$ is supported on $A+A$, by the Cauchy-Schwarz inequality,

$$\left|A\right|^4 \leq \left|A+A\right| \sum_x 1_A * 1_A(x)^2 = \left|A+A\right| E(A)$$

as required. □

In particular, if A has small doubling then it has large energy. Despite what naive intuition might suggest, the converse does not hold, in that large energy does not necessarily force a small sumset - indeed, if we take A to be the union of an arithmetic progression and a geometric progression of equal sizes, then the sumset will be very large, $\left|A+A\right| \gg \left|A\right|^2$ (from the geometric progression) but the additive energy will also be very large, $E(A) \gg \left|A\right|^3$ (from the arithmetic progression).

Fortunately, we can show that this is the only obstruction, and that if $E(A)$ is large then there is a relatively large subset $A' \subset A$ such that $\left|A'-A'\right|$ is small. This is known as the Balog-Szemerédi-Gowers lemma[3], of which we will prove the following version. This is a reasonably strong form of the lemma, and we will follow closely the proof of Schoen [9].

Lemma 7 (Balog-Szemerédi-Gowers)**.** *If $E(A) \geq K^{-1}\left|A\right|^3$ then there exists a subset $A' \subset A$ such that $\left|A'\right| \gg K^{-1}\left|A\right|$ and*

$$\left|A'-A'\right| \ll K^5 \left|A\right|.$$

The precise exponents of K here are rarely important in practice - what matters is that both the size of A' and the size of $A'-A'$ have only a polynomial dependence on K. As we will see throughout this course, structural results with polynomial dependencies are both rare and useful! Some ways to improve the exponents here are given on the examples sheet (though the exact best exponent possible remains an open problem).

Digression: The First Moment Method. Before we prove the Balog-Szemerédi-Gowers lemma, we digress briefly to discuss the first moment method, which we shall use in the proof. This is a simple application of probability, but this style of proof is extremely useful in combinatorics.

Put succinctly, the first moment method that we require is the observation that, for any real-valued random variable X, we have $X \geq \mathbb{E}\, X$ with positive probability. This is coupled with the observation that expectation is linear, which makes it very straightforward to calculate. (More generally, one might use the first moment method to refer to more quantitative uses of the expectation, involving things like Markov's inequality.)

Since we'll just be dealing with probability measures on finite sets, you don't need to worry about any serious use of probability theory. Indeed, using the language of

[3] Note the lack of alphabetical order! Some form of this, with very poor dependency on K, was proved by Balog and Szemerédi [1], and then Gowers [4] found an alternative approach which delivered polynomial-type bounds, which were needed for his application to Szemerédi's theorem.

probability is just for convenience, and one could phrase these arguments as purely counting arguments. But the probabilistic viewpoint is a useful and suggestive one.

Back to the proof. The proof will be done in two stages. For the first, we will find a large subset $X \subset A$ such that 'many' differences in $X - X$ have many different representations as elements of $A - A$. In the second stage we will leverage this to find some large $X' \subset X$ such that $X' - X'$ is small, as required. (Again, this proof is a slight weakening of one due to Schoen [9], which proves a slightly stronger version of the Balog-Szemerédi-Gowers lemma, with the best known exponents at the time of writing.)

Lemma 8. *If $E(A) \geq K^{-1} \left|A\right|^3$ then for any $0 < c < 1$ there is some $X \subset A$ such that $\left|X\right| \gg K^{-1} \left|A\right|$ and for all but at most $c \left|X\right|^2$ many pairs $(a, b) \in X^2$,*

$$1_A \circ 1_A(a - b) \gg cK^{-2} \left|A\right|.$$

Proof. The set X will be of the form $A \cap (A + s)$, where s is a random element of $A - A$ (which is necessary or else X will be empty). The simplest way to choose such an s is uniformly, with probability $1/\left|A - A\right|$ – the problem there is that we don't have control over the size of $A - A$, so want to avoid it appearing in our calculations.

Instead, we choose the next most natural way to choose a random element of $A - A$, which is to note that $1_A \circ 1_A$ is a function we already know something about (by control of $E(A)$) and whose support is $A - A$. Thus we choose s with probability $1_A \circ 1_A(s)/\left|A\right|^2$. The expected size of X is

$$\begin{aligned}\mathbb{E}\left|X\right| &= \frac{1}{\left|A\right|^2} \sum_{a \in A} \sum_s 1_A \circ 1_A(s) 1_A(a - s)\\ &= \frac{\sum_s 1_A \circ 1_A(s)^2}{\left|A\right|^2}.\\ &= E(A)/\left|A\right|^2\end{aligned}$$

For any $G \subset A^2$, the expected number of pairs of X^2 in G is

$$\begin{aligned}\mathbb{E}\left|X^2 \cap G\right| &= \sum_{(a,b) \in G} \mathbb{P}(a, b \in X)\\ &= \frac{1}{\left|A\right|^2} \sum_{(a,b) \in G} \sum_s 1_A \circ 1_A(s) 1_A(a - s) 1_A(b - s).\end{aligned}$$

The innermost sum we bound using the trivial observation that $1_A \circ 1_A(s) \leq \left|A\right|$:

$$\begin{aligned}\sum_s 1_A \circ 1_A(s) 1_A(a - s) 1_A(b - s) &\leq \left|A\right| \sum_x 1_A(a - x) 1_A(b - x)\\ &= \left|A\right| 1_A \circ 1_A(a - b).\end{aligned}$$

It follows that

$$\mathbb{E}\left|X^2 \cap G\right| \leq \frac{1}{\left|A\right|} \sum_{(a,b) \in G} 1_A \circ 1_A(a - b).$$

In particular, if G is the set of pairs where

$$1_A\circ 1_A(a-b)\leq \frac{c}{2}\frac{E(A)^2}{\lvert A\rvert^5},$$

then (using the trivial bound $\lvert G\rvert\leq \lvert A\rvert^2$)

$$\mathbb{E}\left\lvert X^2\cap G\right\rvert\leq \frac{c}{2}\frac{E(A)^2}{\lvert A\rvert^4}.$$

It follows that, since by the Cauchy-Schwarz inequality we have $\mathbb{E}\lvert X\rvert^2\geq (\mathbb{E}\lvert X\rvert)^2$,

$$\mathbb{E}\left(\lvert X\rvert^2-\tfrac{1}{c}\left\lvert X^2\cap G\right\rvert\right)\geq \frac{1}{2}\frac{E(A)^2}{\lvert A\rvert^4}.$$

By the first moment method there is some X such that

$$\lvert X\rvert^2-\tfrac{1}{c}\left\lvert X^2\cap G\right\rvert\geq \frac{1}{2}\frac{E(A)^2}{\lvert A\rvert^4}.$$

In particular, such an X must satisfy

$$\lvert X\rvert\geq \frac{E(A)}{2^{1/2}\lvert A\rvert^2}$$

and $\left\lvert X^2\cap G\right\rvert\leq c\lvert X\rvert^2$. By our choice of G, this means that all but at most $c\lvert X\rvert^2$ many pairs $a,b\in X$ satisfy

$$1_A\circ 1_A(a-b)\geq \frac{c}{2}\frac{E(A)^2}{\lvert A\rvert^5},$$

and the lemma is proved. □

We have found a large subset of X with a lot of structure, in particular that many elements of $X-X$ are well-represented by differences from $A-A$. It does not follow immediately that $\lvert X-X\rvert$ is small, however - for this we need to refine X a little further.

Proof of Theorem 7. We apply Lemma 8 with $c=1/8$. Noting that $1_A\circ 1_A(a-b)=1_A\circ 1_A(b-a)$, we can define a graph G with vertex set X such that a,b are connected by an edge in G if and only if $a\neq b$ and $1_A\circ 1_A(a-b)\gg K^{-2}\lvert A\rvert$. By Lemma 8 there are at most $\frac{1}{8}\lvert X\rvert^2$ many pairs $(a,b)\in X^2$ where this lower bound fails, and hence at least $\binom{\lvert X\rvert}{2}-\frac{1}{16}\lvert X\rvert^2$ many edges in G.

In particular, if $d(x)$ denotes the degree of a vertex x in G, then

$$\sum_{x\in X}d(x)\geq \tfrac{7}{8}\lvert X\rvert^2-\lvert X\rvert. \tag{1}$$

Let A' be the subset of X consisting of those elements of degree at least $\frac{3}{4}\lvert X\rvert$ in G. The contribution to (1) from those $x\in X\backslash A'$ is at most $\frac{3}{4}\lvert X\rvert^2$. In particular,

$$\lvert X\rvert\lvert A'\rvert\geq \sum_{x\in A'}d(x)\geq \tfrac{1}{8}\lvert X\rvert^2-\lvert X\rvert,$$

whence $\lvert A'\rvert\gg\lvert X\rvert$ (note that we can certainly assume $\lvert X\rvert\geq 10$, for example, or else the conclusion holds trivially).

We now claim that for any $x \in A'-A'$ there are $\gg K^{-4}\left|A\right|^2\left|X\right|$ many quadruples $(a_1,a_2,a_3,a_4)\in A^4$ such that $x=a_1-a_2+a_3-a_4$. Assuming this for the moment, since the total number of such quadruples is trivially at most $\left|A\right|^4$, we have

$$\left|A\right|^4 \gg \left|A'-A'\right| K^{-4}\left|A\right|^2\left|X\right|.$$

Recalling that $\left|A'\right| \gg \left|X\right| \gg K^{-1}\left|A\right|$, the result follows.

It remains to prove the claimed lower bound on the number of quadruples. Fix some[4] $a,b\in A'$ such that $x=a-b$. By choice of A', there must be $\geq \frac{1}{2}\left|X\right|$ many elements $c\in X$ such that both (a,c) and (b,c) are edges in G, whence there are $\gg K^{-2}\left|A\right|$ many $(a_1,a_2)\in A^2$ such that $a_1-a_2=a-c$, and similarly many $(a_3,a_4)\in A^2$ such that $a_3-a_4=b-c$. Any choice of such representations gives a quadruple such that $a_1-a_2-a_3+a_4=x$, since

$$x=a-b=(a-c)-(b-c)=(a_1-a_2)-(a_3-a_4).$$

Finally, note that different c must give rise to different quadruples, since a and b are fixed with x, and hence c can be recovered from the quadruple. There are $\gg \left|X\right|$ many choices for c, and each c gives rise to $\gg K^{-4}\left|A\right|^2$ many different quadruples, whence there are $\gg K^{-4}\left|A\right|^2\left|X\right|$ many quadruples in total, as required. □

3. Covering lemmas

Definition 2 (Covering)**.** A set A is K-covered by B if there is a set X with $\left|X\right|\leq K$ such that $A\subset X+B$ – that is, A is contained in the union of at most K translates of B.

This is a weak, but very useful notion of structure, which we have already seen informally. For example, if A is K-covered by another set B of comparable size, so $\left|A\right|\gg\left|B\right|$, and B has small doubling $\left|B+B\right|\ll\left|B\right|$, then A also has small doubling, since

$$\left|A+A\right|\leq\left|X+B+X+B\right|\leq K^2\left|B+B\right|\ll_K\left|B\right|.$$

Covering is a more refined structural property than doubling constant, however, and one can often get a lot more out of it. For example, if we can efficiently cover a sumset $A+B$ by A itself then we can immediately control the iterated sumsets as well, since if $A+B\subset A+X$ then $A+nB\subset A+nX$, by induction on n. We will see a demonstration of this idea in practice shortly.

First, however, we need to show how to produce an efficient covering in the first place! The first idea, due to Ruzsa, has a remarkably simple proof, but already this form of covering suffices for many applications.

Lemma 9 (Ruzsa covering lemma)**.** *If $\left|A+B\right|\leq K\left|B\right|$ then A is K-covered by $B-B$. That is, there is a set $X\subset A$ such that $\left|X\right|\leq K$ and $A\subset X+B-B$.*

Proof. Let $X\subset A$ be maximal such that the translates $\cup_{x\in X}(t+B)$ are all disjoint. Note that some such X certainly exists, since any singleton member of A satisfies this requirement. Furthermore,

$$\left|X\right|\left|B\right|=\left|X+B\right|\leq\left|A+B\right|\leq K\left|B\right|,$$

[4] Note that we are not using many such $a,b\in A'$ with $x=a-b$, just fixing one such pair – the multiplicity in this proof comes from the popularity of a and b, not in the popularity of x itself!

and hence $|X| \leq K$. Finally, if $x \in X$ then certainly $x \in X+B-B$, and if $x \in A\backslash X$ then, by maximality of X, there exists some $b_1, b_2 \in B$ and $y \in X$ such that $x + b_1 = y + b_2$, and so $x = y + b_2 - b_1 \in X + B - B$. □

Note that we are covering A by $B-B$ rather than B itself. Intuitively, the difference set $B-B$ is much 'smoother' than B itself, and enjoys much richer structure. A classic picture to have in mind is a random large subset of a group, say $B \subset G$ with $|B| \approx \frac{3}{4}|G|$. The randomness of B is an obstacle to many desirable structural properties (for example, one cannot cover G itself using only a few translates of B), but it is easy to show[5] that $B - B = G$, which is as structured as one could hope for! The difference set operation has smoothed out the set, and filled in the gaps.

Before giving more elaborate types of covering lemmas, we present a striking application of Ruzsa's covering lemma and the other elementary tools proven thus far to establish our first non-trivial inverse sumset result. For this we must specialise our group G to be a group of small torsion, which for the sake of concreteness, we will assume is of the form $\mathbb{F}_p^n$, for some fixed prime p.

As we saw earlier, if A is a subset of a coset of a subgroup H, with $|A| \geq K^{-1}|H|$, then $|A+A| \leq K|A|$. In particular, if A is a large subset of a coset of a subgroup, then A must have small doubling. We will now show that this is the *only* way that a set can have small doubling.

Unfortunately, our proof only works in groups of small torsion. This is a demonstration of what was alluded to earlier, that additive combinatorics in groups such as $\mathbb{F}_p^n$ tends to be much easier than the general case.

Theorem 1 (Freiman-Ruzsa inverse theorem for bounded torsion)**.** *If $A \subset \mathbb{F}_p^n$ is such that $|A+A| \leq K|A|$ then A is contained in a coset of some subgroup H such that $|A| \gg_{p,K} |H|$.*

Roughly speaking, the idea is that if we just take all the additive span of A, then this itself generates some finite group which clearly contains A (assuming $0 \in A$). The difficulty is in obtaining a suitable upper bound for the size of this group – trivially all we can say is that $|H| \leq p^{|A|}$. We need to use the small doubling hypothesis to improve this to $|H| \ll |A|$.

Proof. Without loss of generality, we may suppose that $0 \in A$, since both hypothesis and conclusion are invariant under translation. The obvious way to apply Ruzsa's covering lemma is to cover A by translates of $A-A$, but this does not lead anywhere useful – obviously A is covered by a single translate of $A-A$ anyway, without any small doubling assumption!

Instead of covering A by $A-A$, we will cover $2A-A$ – this is non-trivial, and much more powerful, since $A-A$ is a 'simpler' set than $2A-A$. We first use Plünnecke's inequality to see that

$$|(2A-A)+A| = |3A-A| \leq K^4|A|.$$

We may now apply Ruzsa's covering lemma to find some $X \subset 2A-A$ such that $|X| \leq K^4$ and

$$2A - A \subset X + A - A.$$

[5] Indeed, any subset $A \subset G$ of size $|A| > \frac{1}{2}|G|$ has $A - A = G$, since for any $x \in G$ the sets A and $x + A$ must intersect.

That is, $A + (A - A) \subset X + A - A$. But then also

$$A + A + (A - A) \subset A + X + A - A = X + 2A - A \subset 2X + A - A,$$

and so on. In general, by induction, for any $n \geq 1$, we have $nA + (A - A) \subset nX + A - A$.

Let H be the subgroup generated by A and H_0 be the subgroup generated by X. If $h \in H$ then $h \in nA$ for some $n \geq 1$. Since

$$nA \subset nA + (A - A) \subset nX + A - A \subset H_0 + A - A$$

for all $n \geq 1$, we have $H \subset H_0 + A - A$. We may now bound the size of H by

$$|H| \leq |H_0| \, |A - A| .$$

We can easily bound both of these factors: $|H_0| \ll_{p,K} 1$, since it is generated by $O_K(1)$ many elements in a group with torsion p, and by the Ruzsa triangle inequality (or Plünnecke's inequality) we have $|A - A| \leq K^2 |A|$. Hence $|H| \ll_{p,K} |A|$ as required. □

Note the essential use of bounded torsion here, in bounding the size of H_0 – if we were to try the same proof for subsets of $\mathbb{Z}$ or $\mathbb{Z}/N\mathbb{Z}$ then we'd get nowhere, since even if X is small the group generated by X could be very large!

If one tracks the dependency of the constants on p and K, then one gets $|H| \leq K^2 p^{K^4} |A|$. This dependency has been improved (as we will explore in Chapter 4), and now an upper bound of $p^{O(K)}$, is known, which is essentially optimal. For example, suppose we take A to be K linearly independent basis vectors in $\mathbb{F}_p^n$. Then trivially $|A + A| \leq |A|^2 = K |A|$, and the smallest coset which contains all of A has size p^K.

In part the reason for this exponential dependence on K is that we are insisting that all of A is contained in the same coset of H. Using the covering terminology, we might say that A has to be 1-covered by H. If we relax this slightly, and allow instead for A to be just efficiently covered by H, then we expect to be able to do much better, with polynomial bounds instead of exponential. (For example, note that in the linearly independent example above, A is trivially K-covered by the trivial subgroup.)

Conjecture 1 (Marton's conjecture)**.** *If $A \subset \mathbb{F}_p^n$ is such that $|A + A| \leq K |A|$ then A is $O_p(K^{O_p(1)})$-covered by a subgroup H with $|H| \leq |A|$.*

This immediately implies the following, by the pigeonhole principle.

Conjecture 2 (Polynomial Freiman-Ruzsa conjecture for $\mathbb{F}_p^n$)**.** *If $A \subset \mathbb{F}_p^n$ is such that $|A + A| \leq K |A|$ then there is a coset of a subgroup H with $|H| \ll_p K^{O_p(1)} |A|$ such that $|A \cap H| \gg_p K^{-O_p(1)} |A|$.*

A similar result is conjectured for any abelian groups, even subsets of $\mathbb{Z}$, provided one replaces subgroups by (generalised) arithmetic progressions, and in general is known as the Polynomial Freiman-Ruzsa Conjecture. We will return to this topic in Chapter 4.

For some applications, we don't necessarily need the set of covering translates X to be small in size, as long as it is spanned by a small number of elements (i.e. small in complexity). This suggests the following definition.

Definition 3 (Span covering)**.** A set A is K-span covered by B if there is a multiset X (i.e. a set where an element may occur more than once) of size $|X| \leq K$ such that $A \subset \mathrm{Span}(X) + B$, where

$$\mathrm{Span}(X) = \left\{ \sum_{x \in X} c_x x : c_x \in \{-1, 0, 1\} \right\}.$$

The following lemma is similar to Ruzsa's covering lemma, but for span covering. The main point is that the dependence on $|A + B| / |B|$ is now only logarithmic, provided $|A + A| / |A|$ is small. Note that this result does not imply a stronger result for the previous notion of covering, since the size of $\mathrm{Span}(X)$ itself may be exponential in $|X|$, but we have obtained much information about what the set of covering translates looks like.

Lemma 10. *Suppose $|A + A| \leq K |A|$ and $|A + B| \leq K' |B|$. Then A is $O(K \log(KK'))$-span covered by $B - B$.*

Proof. Instead of producing some X by a single greedy construction, as in the proof of Ruzsa's covering lemma, we will use a more subtle iterative construction.

We will construct a sequence of sets $B_0, B_1, B_2, \ldots$ as follows. Let $B_0 = B$. Suppose that B_n is given, for some $n \geq 0$. There are two cases:

1. If there are $2K$ many disjoint translates $a + B_n$ with $a \in A$, then we let B_{n+1} be the union of these translates, so that $B_{n+1} = A'_n + B_n$ for some $A'_n \subset A$ of size $|A'_n| = 2K$, and $|B_{n+1}| = 2K |B_n|$.
2. Otherwise, if it is not possible to choose $2K$ such translates, then we choose some maximal set $A' \subset A$ such that the translates $a + B_n$ are disjoint, and let $B_{n+1} = A' + B_n$.

We will first argue that this process terminates eventually. Note that, by induction, we must have $B_n \subset B + nA$ for all $n \geq 0$, and so

$$(2K)^n |B| = |B_n| \leq |B + nA|.$$

By Ruzsa's triangle inequality and Plünnecke's inequality, we have

$$|B + nA| \leq \frac{|B + A| |nA - A|}{|A|} \leq K' K^{n+1} |B|.$$

Comparing these lower and upper bounds, we have

$$(2K)^n \leq K' K^{n+1}$$

and so $2^n \leq K'K$, or $n \ll \log(K'K)$. Therefore this constructive process must halt at some $n \ll \log(K'K)$.

We now claim that if this process halts at the construction of B_{n+1}, then the span of $X = A_0 \cup \cdots \cup A_n$ (viewed as a multiset, so that some $a \in X$ may occur more than once) is a suitable set of covering translates, and we are done, since

$$|X| = \sum_{i=0}^{n} |A_i| \leq (n + 1)2K \ll K \log(K'K).$$

Suppose that $a \in A$. We know that $a + B_n$ is not disjoint from B_{n+1}, by maximality of the set of translates in our choice of B_{n+1}. Therefore $a \in B_{n+1} - B_n$.

But we know what each of these sets are: by induction

$$B_n = A_{n-1} + \cdots + A_0 + B$$

and

$$B_{n+1} = A_n + \cdots + A_0 + B.$$

Therefore

$$\begin{aligned} A &\subset B_n - B_{n-1} \\ &= B - B + (A_0 - A_0) + \cdots + (A_{n-1} - A_{n-1}) + A_n \\ &\subset B - B + \mathrm{Span}(A_0 \cup \cdots \cup A_n) \end{aligned}$$

as required. □

Finding arithmetic progressions

Thus far we have discussed mainly 'global' problems – given a set A, what kind of structure does it have, and how do the different measures of structure relate.

Another rich area of problems comes from looking at 'local' problems – what kind of smaller structures must A contain? In particular, we have one of the guiding meta-questions of additive combinatorics:

What kind of conditions on a set of integers guarantee the presence of an arithmetic progression?

To avoid confusion, a k-term arithmetic progression is a set of the shape $\{a, a+d, a+2d, \ldots, a+(k-1)d\}$ and we say that it is non-trivial if $d \neq 0$.

Theorem 2 (van der Waerden)**.** *For any $r, k \geq 1$, if the integers are coloured in r colours then there is some monochromatic non-trivial k-term arithmetic progression.*

This theorem is proved, using purely combinatorial methods, in Imre Leader's Part III course on Ramsey Theory. This was first conjectured (in the case of 2 colours) by Baudet in the early 1920s, and it floated around conferences for a while, before it reached the ears of van der Waerden. In 1926, in a conversation after lunch, he, Emil Artin, and Otto Schreier, worked out the essentials of the proof, which van der Waerden put together and published.

After this, it is reasonable to wonder exactly what kind of conditions on a colour class suffice. Since, by the pigeonhole principle, there must exist some colour containing at least $1/r$ proportion of the integers, one might ask whether this 'large' colour class must contain long arithmetic progressions? More generally, if we have any positive density subset of the integers, must it contain arbitrarily long arithmetic progressions?

This conjecture was made informally since at least the 1930s, but it was not until 1975 that a full proof was found by Szemerédi. We state the result in the following finitary form.

Theorem 3 (Szemerédi's theorem)**.** *For any $\delta > 0$ and $k \geq 1$ there exists $N \ll_{\delta,k} 1$ such that if $A \subset \{1, \ldots, N\}$ has size $|A| \geq \delta N$ then A contains a non-trivial k-term arithmetic progression.*

This is one of the great achievements of 20th centry combinatorics. The original proof by Szemerédi was very combinatorial, but since then a number of alternative proofs have been found. The most effective method so far is that of Gowers. To state it most clearly, let $r_k(N)$ denote the largest subset of $\{1, \ldots, N\}$ with no non-trivial k-term arithmetic progressions. Using this language, we can state Szemerédi's theorem as $r_k(N)/N \to 0$ as $N \to \infty$. Gowers' method gives an explicit rate of decay of this function, which remains the best known for general k.

Theorem 4 (Gowers)**.**

$$r_k(N) \ll \frac{N}{(\log\log N)^{c_k}},$$

where $c_k > 0$ is some constant depending only on k.

Unfortunately we will not have time to prove Szemerédi's theorem or give a proper account of the method of Gowers (which will, however, be discussed in Julia Wolf's non-examinable course 'Higher-order uniformity and applications' in Easter term).

In this course we will focus on the easiest non-trivial case of $k = 3$. We will omit the subscript, so $r(N)$ denotes the size of the largest subset of $\{1, \ldots, N\}$ which contains no non-trivial three-term arithmetic progressions.

This special case of Szemerédi's theorem was proved earlier, by Roth in 1953.

Theorem 5 (Roth 1953)**.** *For any $\delta > 0$ there exists $N \ll_\delta 1$ such that if $A \subset \{1, \ldots, N\}$ has $|A| \geq \delta N$ then A contains a non-trivial three-term arithmetic progression.*

Equivalently, $r(N)/N \to 0$. In fact Roth proved the explicit estimate

$$r(N) \ll \frac{N}{\log\log N}.$$

The proof of Roth used Fourier analysis, which will be the topic of the next chapter. At the end of the next chapter we will use the tools we will develop to prove the following quantitative improvement of Roth's theorem.

Theorem 6 (Bourgain 1999)**.**

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

CHAPTER 2

Fourier analysis

In this chapter we will develop the Fourier analysis we require. We will keep things simple by only discussing finite abelian groups in this chapter.

4. Basic concepts

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 $\widehat{G}$.

Since multiplication in $\mathbb{C}$ is commutative, it is immediate that $\widehat{G}$ is abelian. Furthermore, if $\gamma \in \widehat{G}$ then, since γ is a homomorphism, we must have $\gamma(1) = 1$, and furthermore $\gamma(x)$ must be a $|G|$th root of unity, for any $x \in G$. In particular, $\widehat{G}$ is finite.

Thus it is trivial that $\widehat{G}$ is also a finite abelian group. What is less trivial, is that in fact $\widehat{G}$ is isomorphic to G itself.

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 11. *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.)*

This is straightforward to prove once one has the classification of finite abelian groups (note that it is immediate for cyclic groups, and all finite abelian groups are the direct product of cyclic groups), but we will not take that digression here, and simply state it without proof.

We will often identify elements of G and $\widehat{G}$ under this isomorphism. This identification is straightforward for our G of special interest. 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

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

For example, the additive energy can be written as

$$E(A) = \sum_{x} 1_A * 1_A(x)^2 = \langle 1_A * 1_A, 1_A * 1_A\rangle.$$

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.

The key identities which are at the heart of Fourier analysis are the orthogonality relationships:

Lemma 12 (Orthogonality)**.** *For any $\gamma \in \widehat{G}$,*

$$\sum_{x\in G}\gamma(x) = \begin{cases} |G| & \textit{if } \gamma = \mathbf{1} \textit{ and} \\ 0 & \textit{otherwise.}\end{cases}$$

Similarly, for any $x \in G$,

$$\mathop{\mathbb{E}}_{\gamma\in\widehat{G}}\gamma(x) = \begin{cases} 1 & \textit{if } x = 0 \textit{ and} \\ 0 & \textit{otherwise.}\end{cases}$$

Proof. The first equality in both claims is trivial, since $\mathbf{1}(x) = 1$ for all $x \in G$, and $\gamma(0) = 1$ for all $\gamma \in \widehat{G}$. For the other equality in the first claim, let $\gamma \neq \mathbf{1}$, so that there exists some $y \in G$ such that $\gamma(y) \neq 0$. Since $G + y = G$, we have

$$\gamma(y)\sum_{x\in G}\gamma(x) = \sum_{x\in G}\gamma(x+y) = \sum_{z\in G}\gamma(z).$$

Since $\gamma(y) \neq 0$ the only way this is possible is if $\sum_{x\in G}\gamma(x) = 0$. The second claim is proved similarly, using the existence of some $\lambda \in \widehat{G}$ such that $\lambda(x) \neq 1$. Such a λ exists, for otherwise $\widehat{G}$ would act trivially on the group generated by x, and hence would also be the dual group for $G/\langle x\rangle$, but we know this has size $|G/\langle x\rangle| < |G| = |\widehat{G}|$. □

Definition 4. 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 13 (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 14 (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 $A \subset G$ 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 15. *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).

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

Our final goal in this chapter will be to prove an upper bound for the size of the largest $A \subset \mathbb{Z}/N\mathbb{Z}$ without 3APs. In this section we will prove a similar upper bound for the size of subsets of $\mathbb{F}_p^n$ without 3APs. This setting is much simpler, and it is much easier to see what's going on. Our subsequent proof for the case of $A \subset \mathbb{Z}/N\mathbb{Z}$ will be essentially translating this proof and seeing what needs to change when we change $\mathbb{F}_p^n$ to $\mathbb{Z}/N\mathbb{Z}$.

Fix some odd prime $p \geq 3$. (This is needed in the proof, basically because a 3AP is degenerate in the case $p = 2$ – the progression $x, x+d, x+2d$ is just $x, x+d, x$, which always exists provided $|A| \geq 2$.)

Before the proof, we'll give a big picture sketch. We are given a set $A \subset \mathbb{F}_p^n$, and all we know about it is its size – or, equivalently, its density $\alpha = |A|/p^n$. We want to know that, provided α is large, and with no other information about A at all, that A contains a 3AP. We make two observations:

1. If A is very structured, e.g. if A is a subspace, then A contains many 3APs, since if $x, x+d \in A$ are any two distinct elements, then $d \in A$, and hence $x + 2d \in A$ also.
2. On the other hand, if A is a random subset of density α, then the expected number of 3APs is $\alpha^3 p^{2n}$. This includes the trivial 3APs (with $d = 0$), but there are only $|A| = \alpha p^n$ of such. So provided $\alpha^3 p^{2n} \gg \alpha p^n$, i.e. $\alpha \gg p^{-n/2}$, this count is negligible, and we're done.

So we're done, in the case α large, if A is either very structured or very random. This certainly supports our belief that it should hold for all A (viewing structure/random as two opposing extremes).

But how to turn this into a proof? Core idea is 'density increment', goes back to Roth 1953. What Tao calls the 'randomness vs. structure' dichotomy. Suppose $A \subset \mathbb{F}_p^n$ has no 3APs. We want to show $\alpha = |A|/p^n$ is small. 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 7 (Meshulam)**.** *If $A \subset \mathbb{F}_p^n$ 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 16. *Let V be an n-dimensional vector space over $\mathbb{F}_p$, and let $A \subset V$ 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 7. Let $A \subset \mathbb{F}_p^n$ 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. $A_i \subset V_i$,
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 16 tells us, applied to $A_k \subset V_k$. By maximality of k, the second condition of Lemma 16 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 16. 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 16. 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. We first note that (in any group) 3APs are exactly those sets $\{x, y, z\}$ which are solutions to the linear equation $x + y = 2z$. Indeed, given a 3AP $x, x + d, x + 2d$, we see that letting $y = x + 2d$ and $z = x + d$, we have a solution to this equation:

$$x + (x + 2d) = 2(x + d).$$

On the other hand, if $x + y = 2z$, then letting $d = z - x$, the equation forces $y = x + 2d$ and $z = x + d$, and so $\{x, y, z\} = \{x, x + d, x + 2d\}$. In this language, the trivial 3APs with $d = 0$ are those trivial solutions where $x = y = z$.

This means that 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 $\alpha 1_G$:

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

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

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

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 (2) 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 $v_1 + V', \dots, v_p + V'$ 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 $V_1', \dots, V_p'$ 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 $x_1, \dots, x_p$. 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 $\geq \frac{1}{2}\alpha^2 p^n$, we have

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

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.

6. 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 5 (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
$$\left|1-\gamma(x_1+x_2)\right| = \left|\gamma(-x_1)-\gamma(x_2)\right| \leq \left|1-\gamma(x_1)\right| + \left|1-\gamma(x_2)\right|.$$
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
$$\left|1-e(\theta)\right| = \left|e^{-\pi i\theta} - e^{\pi i\theta}\right| = 2\left|\sin(\pi\theta)\right|.$$
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
$$4\|\theta\| \leq \left|1-e(\theta)\right| \leq 2\pi\|\theta\|.$$

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 $\left|1-\gamma(x)\right| \leq \rho$ 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 17. *If B is a Bohr set of rank d and width $\rho \in (0, 1]$ then*

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

Furthermore,

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

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

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

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 $X - X \subset B$ 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

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

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 $X' - X' \subset B_{1/2}$, 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,

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

as required. □

Bohr sets are, in general, not even approximately group-like, and may grow exponentially under addition. Indeed, recall that $\mathrm{Bohr}(\Gamma; \rho) + \mathrm{Bohr}(\Gamma; \rho) \subset \mathrm{Bohr}(\Gamma; 2\rho)$. 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 6 (Regularity[a])**.** A Bohr set B of rank d is regular if for all $0 \leq \delta \leq 1/200d$ we have

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

and

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

[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 $\Gamma \subset \mathbb{F}_2^n$ 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 18 (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 18, 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 19. *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 $(a,b) \subset I \cup J$. 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 17. 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 18. 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 19, 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 17, 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 20. *If B is a regular Bohr set of rank d and $B'\subset B_\delta$, 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 $A\subset B$, 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

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

We now note that $B_{1-\delta}\subset B+y$ – 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,

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

and the proof is complete. □

7. 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 8 (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 9 (Bourgain 1999). *Let G be a finite abelian group of odd order N. If $A \subset G$ 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 8 from Theorem 9.

Proof of Theorem 8 assuming Theorem 9. Suppose $A \subset \{1, \ldots, N\}$ 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 9 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 9 follows from it.

Lemma 21. *Let B be a regular Bohr set of rank d and width ρ. Let $A \subset B$ 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* $B' \subset B$ *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 9 by repeated applications of Lemma 21.

Proof. Let $A \subset G$ 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 $A_0, \ldots, A_k$ and associated Bohr sets $B_0, \ldots, B_k$, with ranks $d_0, \ldots, d_k$ and widths $\rho_0 \geq \cdots \geq \rho_k$, 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) $A_i \subset B_i$,

(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 21,

(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 7, part (4) implies that $k \ll \alpha^{-1}$.

We now apply Lemma 21 to $A_k \subset B_k$. By maximality of k, the second condition of Lemma 21 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 17, we have

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

Therefore,

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

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 21, 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.

Lemma 22. *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 $A \subset B$ with density $\alpha = \left|A\right| / \left|B\right|$. Either*

1. *(many 'progressions') $\langle 1_A * 1_A, 1_{2\cdot B_\delta}\rangle \geq \frac{1}{2}\alpha^2 \left|B\right| \left|B_\delta\right|$ 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*
$$\frac{\left|(A-x)\cap B'\right|}{\left|B'\right|} \geq (1+1/256)\alpha.$$

Proof. If the first condition fails then

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

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 $A = A_{\text{large}} \sqcup A_{\text{small}}$, where

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

We have

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

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

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 $z \in B_{c\delta^2}$,

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

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

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

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

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

We therefore let $B_{\text{small}} = B \cap (A_{\text{small}} + 2\cdot B_{c\delta^2})$. 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 $|B_{\text{small}}| < \frac{1}{16}|B|$. In this case we consider the convolution $\langle 1_{A_{\text{small}}} * 1_{2\cdot B_{c\delta^2}}, 1_B\rangle$. By regularity, and noting that $2\cdot B_{c\delta^2} \subset B_{c\delta^2}+B_{c\delta^2} \subset B_{2c\delta^2}$,

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

(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

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

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

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

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 $|(2\cdot B)_{1+\delta}| = |2\cdot B_{1+\delta}| = |B_{1+\delta}|$.

Case 2: Suppose that $|B_{\text{small}}| \geq \frac{1}{16}|B|$. 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

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

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

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

Combining these upper bounds, we deduce that

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

Comparing our lower and upper bounds and simplifying yields

$$(\tfrac{1}{8}+\tfrac{1}{256})\,|B_{\text{small}}| \leq \tfrac{1}{128}\,|B|\,,$$

which contradicts our lower bound on $|B_{\text{small}}|$, 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 $A \subset B$ 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 23. *Let B be a regular Bohr set of rank d and suppose $A \subset B$ has density $\alpha = |A| / |B|$. Suppose that $B', B'' \subset B_\delta$ 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 $|(A - x) \cap B'| \geq (1-2\epsilon)\alpha |B'|$ and $|(A - x) \cap B''| \geq (1-2\epsilon)\alpha |B''|$, or*
2. *there is an x such that*

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

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

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

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

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

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 21. The overall structure of the proof is very similar to the simpler case in $\mathbb{F}_p^n$, Lemma 16, but there are complications due to having to work with Bohr sets of different widths.

Proof of Lemma 21. Let $A \subset B$ 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 23 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 22 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 17 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 16) 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 16, 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

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

As in the proof of Lemma 16, 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

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

In the proof of Lemma 16 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

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

and hence

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

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,

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

Once again, provided we have chosen $c_3 > 0$ small enough, this right-hand side is at least $\frac{1}{2}c_4\alpha_1 \left|A_2\right| \left|B'\right|$. Taking the real parts and averaging over all $x \in B^{(2)}$, as in the proof of Lemma 16, we deduce that there exists some $x \in B^{(2)}$ such that

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

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

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

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 3

Almost-periodicity

We use τ_t to denote the translation by t operator, so that if $f : G \to \mathbb{C}$ then $\tau_t f(x) = f(x-t)$ for any $x \in G$. A period of a function is some t such that $\tau_t f = f$. For example,

(1) every function has 0 as a period;
(2) $\sin(x)$ has period 2π; and
(3) if $H \leq G$ is a subgroup then any $t \in H$ is a period for 1_H.

Asking for an exact period is a very rigid constraint on t, which is difficult to achieve in practice, and most functions will not have any non-trivial periods. This is similar to the situation with small sumsets we explored in Chapter 1: the exact condition $|A+A| = |A|$ is very rarely achieved except for very structured sets, so we instead relaxed the condition and studied the behaviour of sets A with $|A+A| \ll |A|$ instead.

In this section we will do the same for the notion of periods, relaxing the definition to 'almost-periods', which is a much more flexible notion – roughly speaking, the idea is that instead of asking for $\tau_t f = f$, we weaken this to $\tau_t f \approx f$, for a precise notion of $\approx$. We will show how to find large sets of almost-periods, and then show how this can be applied to prove several interesting results quite quickly.

We will be particularly interested in finding almost-periods for convolutions – since a sumset is exactly the support of a convolution. For example, if we know that $x \in A+B$, then $1_A * 1_B(x) > 0$, and if t is such that $\tau_t(1_A * 1_B) \approx 1_A * 1_B$, we we expect $1_A * 1_B(x+t) > 0$ also, and so $x + T \subset A + B$, where T is the set of almost-periods. Thus the study of almost-periods is a very useful tool in finding structures inside sumsets.

A useful way to think of almost-periods is as a coarse type of 'continuity'. Recall that a function f being uniformly continuous means that, for all x, the difference $|f(x+t) - f(x)|$ is very small for small t, or put another way, $\|\tau_t f - f\|_\infty$ is small. Thus if a function $f : \mathbb{R} \to \mathbb{R}$ is uniformly continuous then, for any ϵ, we can find a ball B_ϵ around the origin such that for all $t \in B_\epsilon$, we have $\|\tau_t f - f\|_\infty \leq \epsilon$. Almost-periodicity is a similar property for functions $f : G \to \mathbb{C}$, but where we have the additional flexibility of replacing the L^∞ norm by some L^p norm. Recall that $\|f\|_\infty = \sup_x |f(x)|$ and for any $1 \leq p < \infty$ we define the L^p norm on functions $f : G \to \mathbb{C}$ by

$$\|f\|_p = \left(\sum_{x \in G} |f(x)|^p \right)^{1/p}.$$

These are norms, and in particular satisfy the triangle inequality. It is also obvious that they are invariant under the translation operator, so that for any $t \in G$, we have $\|\tau_t f\|_p = \|f\|_p$. The idea is that t is an L^p-almost period if $\|\tau_t f - f\|_p$ is small. More precisely, we have the following.

Definition 7 (Almost-periods)**.** Let $m \geq 1$ be an integer and $X > 0$. For any $1 \leq p \leq \infty$ and $f : G \to \mathbb{C}$ we define the set of L^p-almost periods of f with error X by

$$\mathrm{AP}_p(f; X) = \{t \in G : \|\tau_t f - f\|_p \leq X\}.$$

Before we state our main theorems, we observe a couple of trivial properties.

- For any p, f, X, the set $AP_p(f; X)$ is a symmetric set containing 0. Indeed, we have $\tau_0 f - f \equiv 0$, and if $t \in \mathrm{AP}_p(f; X)$ then
$$\|\tau_{-t} f - f\|_p = \|f - \tau_{-t} f\|_p = \|\tau_t(f - \tau_{-t} f)\|_p = \|\tau_t f - f\|_p.$$
- Almost-period sets are increasing in X, in that if $X' \geq X$ then $\mathrm{AP}_p(f; X) \subseteq \mathrm{AP}_p(f; X')$.
- For any p, f, for large enough X the set of almost-periods is the entire group: $\mathrm{AP}_p(f; 2\|f\|_p) = G$. This follows from the triangle inequality,
$$\|\tau_t f - f\|_p \leq \|\tau_t f\|_p + \|f\|_p \leq 2\|f\|_p.$$
- Addition of sets of almost-periods adds the errors, in the sense that
$$\mathrm{AP}_p(f; X_1) + \mathrm{AP}_p(f; X_2) \subset \mathrm{AP}_p(f; X_1 + X_2).$$
Again, this is an immediate consequence of the triangle inequality, since
$$\begin{aligned}\|\tau_{t_1+t_2} f - f\|_p &\leq \|\tau_{t_1+t_2} f - \tau_{t_2} f\|_p + \|\tau_{t_2} f - f\|_p \\ &= \|\tau_{t_1} f - f\|_p + \|\tau_{t_2} f - f\|_p.\end{aligned}$$

Our first general result on almost-periods says that for any function f we can find a reasonably large Bohr set in the set of almost-periods where the error is proportionate to

$$\|\widehat{f}\|_1 = \mathop{\mathbb{E}}_{\gamma} \left|\widehat{f}(\gamma)\right|.$$

Theorem 10 (Large Bohr sets of almost-periods)**.** *Let $\epsilon \in (0, 1)$ and $m \geq 1$. For any function $f : G \to \mathbb{C}$ the set of L^{2m}-almost-periods of f with error $\epsilon\|\widehat{f}\|_1 N^{1/2m}$ contains a Bohr set B of rank $O(m\epsilon^{-2})$ and width $\gg \epsilon$.*

Our second type of result is just a lower bound on the number of almost-periods, and this only works for functions which are convolutions $1_A * 1_B$, where A is some structured set (but B need not be!).

Theorem 11 (Convolutions have many almost-periods; Dense set version)**.** *Let $\epsilon > 0$ and $m \geq 1$, and G be some finite abelian group of order N. Suppose that $A \subset G$ with density $\alpha = |A|/N$. For any set B,*

$$\left|\mathrm{AP}_{2m}(1_A * 1_B; \epsilon\,|A|\,|B|^{1/2m})\right| \geq \alpha^{O(m\epsilon^{-2})} N.$$

Theorem 12 (Convolutions have many almost-periods; Small doubling version)**.** *Let $\epsilon > 0$ and $m \geq 1$. Suppose that $|A + S| \leq K\,|A|$. Then, for any set B,*

$$\left|\mathrm{AP}_{2m}(1_A * 1_B; \epsilon\,|A|\,|B|^{1/2m})\right| \geq K^{-O(m\epsilon^{-2})}\,|S|\,.$$

We will first give two applications of these almost-periodicity results, and then prove them. The proof of both will be probabilistic.

8. Applications of almost-periodicity

An essential tool in our applications of almost-periodicity is Hölder's inequality, which we have already used several times in this course. As a reminder, Hölder's inequality states that, for any reals $a_i, b_i \in \mathbb{R}$, and any $p \geq 1$,

$$\left|\sum a_i b_i\right| \leq \|a_i\|_p \|b_i\|_{p/(p-1)} = \left(\sum |a_i|^p\right)^{1/p} \left(\sum |b_i|^{\frac{p}{p-1}}\right)^{1-1/p}.$$

For example, the Cauchy-Schwarz inequality is Hölder's inequality with $p = 2$. The case $p = 1$ (where $\|b_i\|_{p/(p-1)}$ should be read as $\|b_i\|_\infty$) is obvious by the triangle inequality. There are many different proofs available for Hölder's inequality, and if you don't know one, I encourage you to try and discover one for yourself. (One approach is to deduce Hölder's inequality from the Cauchy-Schwarz inequality. Another is to first show the simpler inequality that for any $x, y \geq 0$, we have $xy \leq \frac{1}{p}x^p + (1 - \frac{1}{p})y^{p/(p-1)}$.)

Hölder's inequality implies, in particular, that for any function $f : G \to \mathbb{C}$ and any set A, for any $m \geq 1$,

$$|\langle 1_A, f\rangle| \leq |A|^{1-1/2m} \|f\|_{2m},$$

which we will use frequently. Part of the power of almost-periodicity comes from the fact that we will use such estimates for large m, in particular large enough so that $\|f\|_{2m} \approx \|f\|_\infty$.

Our first application is to show that if A is a set with small doubling then we can find a lot of structure inside $A+A-A-A$ – in particular, this four-fold sumset in fact contains some k-fold sumset kT where k can be very large indeed, and T is also reasonably large. This will play a crucial role in our proof of the inverse sumset results of the next chapter.

Theorem 13. *If $|A + A| \leq K|A|$ then for any $k \geq 1$ there is a symmetric set T such that $0 \in T$ and*

$$|T| \geq \exp(-O(k^2(\log K)^2))\,|A|$$

and

$$kT \subset A + A - A - A.$$

Proof. We apply Theorem 12 with $S = A$ and $B = A - A$, and $\epsilon > 0$ and $m \geq 1$ to be chosen soon. Let T be the corresponding set of almost-periods, which in particular is a symmetric set containing 0. Since addition of almost-periods adds the errors, kT is a subset of the set of L^{2m}-almost-periods for $1_A * 1_{A-A}$ with error $k\epsilon |A| |A - A|^{1/2m}$, that is, for any $t \in kT$, we have

$$\|\tau_t(1_A * 1_{A-A}) - 1_A * 1_{A-A}\|_{2m} \leq k\epsilon |A| |A - A|^{1/2m}.$$

Suppose for a contradiction that there is some $t \in kT$ such that $t \notin A+A-A-A$. Then, in particular, $1_A * 1_{A-A} \circ 1_A(t) = 0$, and so

$$\begin{aligned} 0 &= \sum_{a,b\in A} \sum_{c\in A-A} 1_{a-b-c=t} \\ &= \sum_{a\in A} \sum_{b\in A} \sum_{c\in A-A} 1_{b+c=a-t} \\ &= \langle 1_A, \tau_t(1_A * 1_{A-A})\rangle. \end{aligned}$$

On other hand, we have

$$\langle 1_A, 1_A * 1_{A-A}\rangle = \sum_{a,b\in A}\sum_{c\in A-A} 1_{a-b=c} = |A|^2 .$$

Taking the difference,

$$|\langle 1_A, (\tau_t(1_A * 1_{A-A}) - 1_A * 1_{A-A})\rangle| = |A|^2 .$$

On the other hand, by Hölder's inequality, the left-hand side is at most

$$|A|^{1-1/2m} \|\tau_t(1_A * 1_{A-A}) - 1_A * 1_{A-A}\|_{2m} \leq k\epsilon |A|^2 (|A-A| / |A|)^{1/2m}.$$

By Plünnecke's inequality $|A-A| \leq K^2 |A|$, and hence if we choose $m = \lceil \log K \rceil$, then the right-hand side is at most $ek\epsilon |A|^2$. Choosing $\epsilon = 1/2ek$, say, gives a contradiction.

Thus the theorem is proved, since the size of T is

$$|T| \geq K^{-O(m\epsilon^{-2})} |A| \geq \exp(-O(k^2(\log K)^2)) |A| .$$

□

For our second application, we will use Theorem 10, which guarantees a large Bohr set of almost-periods. We will use this to answer the following natural question: given a reasonably large subset A of $\{1, \ldots, N\}$, how long an arithmetic progression must $A + A$ contain?

Theorem 14. *There are constants $c > 0$ and $C \geq 1$ such that the following is true for all large N. Suppose that $A \subset \{1, \ldots, N\}$ has size $|A| = \alpha N$, where $\alpha \geq C\frac{\log\log N}{\sqrt{\log N}}$. Then $A + A$ contains an arithmetic progression of length*

$$\gg \exp(c\alpha\sqrt{\log N}).$$

Note that this is already interesting and non-trivial when $\alpha = 1/4$, for example. Furthermore, the lower bound becomes trivial when $\alpha \ll \log\log N/\sqrt{\log N}$, so this is only saying something for reasonably dense sets.

The proof we give here is due to Croot, Łaba, and Sisask. The first proof of Theorem 14 was given (by quite different, pre-almost-periodicity methods) by Ben Green. The dependence on N here (with $\alpha = 1/4$, say) is still the best-known, but it is an open problem to find the optimal dependence. The best known upper bound is a construction of Ruzsa, which in particular yields a subset A of $\{1, \ldots, N\}$ of size $\geq \frac{1}{4}N$ such that the longest arithmetic progression in $A + A$ is of length $\exp(O(\log N)^{2/3} \log\log N)$.

Before the proof, we note the following simple fact about Bohr sets which we will use.

Lemma 24. *If N is a prime and $B \subset \mathbb{Z}/N\mathbb{Z}$ is a Bohr set of rank d and width ρ then B contains an arithmetic progression of length $\gg \rho N^{1/d}$.*

Proof. Let $\ell \geq 1$ be maximal such that $(\rho/8\ell)^d \geq 2/N$. It is clear that $\ell \gg \rho N^{1/d}$. Our lower bound for the size of Bohr sets implies that $|B_{1/\ell}| \geq 2$, and in particular $B_{1/\ell}$ contains some non-zero element x. By the triangle inequality we have $\ell B_{1/\ell} \subset B$, and in particular $P = \ell\{0, x\}$ is an arithmetic progression of length ℓ contained inside B (note that since N is prime P does indeed have size ℓ). □

Proof. We first note that, just as when searching for three-term arithmetic progressions in $\{1,\dots,N\}$, it suffices to prove the same result with $\{1,\dots,N\}$ replaced by $\mathbb{Z}/N\mathbb{Z}$. Indeed, let $4N < M \leq 8N$ be prime, and note that $A+A \subset \{1,\dots,2N\}$. We claim that any arithmetic progression modulo M inside $A+A$ is a genuine arithmetic progression in the integers (of the same length).

Let P be an arithmetic progression modulo M inside $A+A$, that is, a set $\{x_0,\dots,x_k\} \subset A+A$ where there is $d \in \mathbb{Z}$ such that $x_i \equiv x+id \pmod{M}$. Without loss of generality, since $d \equiv x_1 - x_0 \pmod{M}$, we can assume that $d = x_1 - x_0$. By induction we then have that in fact that $x_i = x+id$ for all $0 \leq i \leq k$ – indeed, suppose that $i \geq 2$ is minimal such that this fails. Then

$$x_i - (x+id) = x_i - (x_{i-1}+d) = x_i + x_0 - x_{i-1} - x_1 \in \{-4N,\dots,4N\},$$

and in particular $|x_i - (x+id)| < M$, and so in fact $x_i \equiv x+id \pmod{M}$ forces $x_i = x+id$, as required. Thus $P \subset A+A$ is not only an arithmetic progression modulo M, but also a genuine arithmetic progression of the same length. We can therefore concentrate on proving the result with $\mathbb{Z}/N\mathbb{Z}$.

The idea is to use Theorem 10 to find some large Bohr set of almost-periods for $1_A * 1_A$, and then find a long progression P inside this Bohr set, and use almost-periodicity to show that some translate of P (not necessarily P itself!) is inside $A+A$.

Let T be the set of L^{2m}-almost periods of $1_A * 1_A$ with error $\frac{\alpha}{4}\left|A\right| N^{1/2m}$. We first observe that this is exactly the set of almost-periods addressed by Theorem 10 with $\epsilon = \alpha/4$, since

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

By Hölder's inequality, for any $f : G \times G \to \mathbb{C}$, and any $P \subset G$,

$$\begin{aligned}\sum_{x\in G}\sup_{t\in P}|f(x,t)| &\leq \sum_{x\in G}\left(\sum_{t\in P}\left|f(x,t)\right|^{2m}\right)^{1/2m}\\ &\leq N^{1-1/2m}\left(\sum_{t\in P}\sum_{x\in G}\left|f(x,t)\right|^{2m}\right)^{1/2m}\\ &\leq N^{1-1/2m}\left|P\right|^{1/2m}\sup_{t\in P}\|f(\cdot,t)\|_{2m}.\end{aligned}$$

In particular, for any $P \subset T$,

$$\begin{aligned}\sum_{x\in G}\sup_{t\in P}|1_A * 1_A(x+t) - 1_A * 1_A(x)| &\leq \left|P\right|^{1/2m} N^{1-1/2m}\max_{t\in P}\|\tau_t(1_A * 1_A) - 1_A * 1_A\|_{2m}\\ &\leq \frac{1}{4}\left|A\right|^2\left|P\right|^{1/2m}.\end{aligned}$$

In particular, provided $|P| < 4^{2m}$, this is less than $\left|A\right|^2 = \sum_{x\in G} 1_A * 1_A(x)$, and hence there is some $x \in G$ such that for all $t \in P$, we have

$$\sup_{t\in P}|1_A * 1_A(x+t) - 1_A * 1_A(x)| < 1_A * 1_A(x).$$

But this means that $1_A * 1_A(x+t) \neq 0$ for all $t \in P$, whence $x + P \subset A + A$.

It remains to choose m so that our set of almost-periods contains a progression P of appropriate length. By Theorem 10, the almost-periods contain a Bohr set of rank $O(m\alpha^{-2})$ and width $\gg \alpha$, and hence a progression P of length $\geq c_1\alpha N^{c_2\alpha^2/m}$ for some small constants $1 > c_1, c_2 > 0$.

If we choose $m = \lfloor \alpha\sqrt{\log N}\rfloor$ then by assumption we have $m \geq 1$ and we can find a progression P of length $\ell = \lfloor c_1\alpha\exp(c_2\alpha\sqrt{\log N})\rfloor$. In particular

$$\ell \leq \exp(c_2\alpha\sqrt{\log N}) \leq e^{2m} < 4^{2m}$$

as required, and so we have found an arithmetic progression of length ℓ inside $A+A$. (The final shape of the bound comes from the fact that $\lfloor x\rfloor \geq x/2$ provided $x \geq 1$, and $\alpha \geq C\log\log N/\sqrt{\log N}$ for large enough constant C (depending on c_1 and c_2) guarantees that $c_1\alpha\exp(c_2\alpha\sqrt{\log N}) \geq \exp(\frac{c_2}{2}\alpha\sqrt{\log N}) \geq 1$, say.) □

9. Probabilistic inequalities

In this section the $\mathbb{E}$ symbol denotes the probabilistic expectation of a random variable. All random variables in this section should be read as taken over some finite probability space.

Suppose we have n independent random variables $X_1, \ldots, X_n$. Consider their sum $X_1 + \cdots + X_n$. Can we control the L^{2m} norm of the sum by the L^{2m} norms of the individual X_i? Hölder's inequality and the triangle inequality immediately give such control:

$$\mathbb{E}\left|\sum_{i=1}^n X_i\right|^{2m} \leq n^{2m-1}\sum_{i=1}^n \mathbb{E}\lvert X_i\rvert^{2m}.$$

It is a very useful fact that, when the random variables are balanced by subtracting their mean, this estimate can be improved by a factor of $(4m/n)^m$ – which, for fixed m and $n \to \infty$, becomes very significant.

Lemma 25 (Marcinkiewicz-Zygmund inequality). *Let $m \geq 1$. If $X_1, \ldots, X_n$ are independent complex-valued random variables with mean zero then*

$$\mathbb{E}\left|\sum_{i=1}^n X_i\right|^{2m} \leq (4m)^m\,\mathbb{E}\left(\sum_{i=1}^n \lvert X_i\rvert^2\right)^m.$$

In particular,

$$\mathbb{E}\left|\sum_{i=1}^n X_i\right|^{2m} \leq (4m)^m n^{m-1}\,\mathbb{E}\sum_{i=1}^n \lvert X_i\rvert^{2m}.$$

Proof. The second inequality follows immediately from the first using Hölder's inequality. We will concentrate on proving the first. Fix some $m \geq 1$. We want to understand the average

$$S = \mathbb{E}\left|\sum_i X_i\right|^{2m}.$$

We first introduce a new family of random variables by letting $(Y_i)_{1\leq i\leq n}$ be new random variables distributed identically to the respective X_i (although completely independently sampled). Since each has $\mathbb{E}\,Y_i = 0$ we can write, by linearity of expectation, the triangle inequality, and Hölder's inequality,

$$S = \mathop{\mathbb{E}}_{X_i}\left|\sum_i X_i - \mathop{\mathbb{E}}_{Y_i} Y_i\right|^{2m} = \mathop{\mathbb{E}}_{X_i}\left|\mathop{\mathbb{E}}_{Y_i}\left(\sum_i X_i - Y_i\right)\right|^{2m} \leq \mathop{\mathbb{E}}_{X_i,Y_i}\left|\sum_i X_i - Y_i\right|^{2m}.$$

Next, we make the crucial observation: that since X_i and Y_i are identically distributed, $X_i - Y_i$ has the same distribution as $Y_i - X_i$. Thus, for any $\epsilon_i \in \{-1, +1\}$,

$$S \leq \mathop{\mathbb{E}}_{X_i, Y_i} \left| \sum_i \epsilon_i (X_i - Y_i) \right|^{2m}.$$

In particular, if we sample $\epsilon_i \in \{-1, +1\}$ uniformly at random, then

$$S \leq \mathop{\mathbb{E}}_{\epsilon_i} \mathop{\mathbb{E}}_{X_i, Y_i} \left| \sum_i \epsilon_i (X_i - Y_i) \right|^{2m}.$$

We now change the order of expectation and consider the expectation over just ϵ_i, viewing the $X_i - Y_i = x_i$, say, as fixed quantities. For any x_i we can expand $\mathbb{E}_{\epsilon_i} |\sum_i \epsilon_i x_i|^{2m}$ and then bound it from above, using the triangle inequality, by

$$\sum_{k_1 + \cdots + k_n = 2m} \binom{2m}{k_1, \ldots, k_n} |x_1|^{k_1} \cdots |x_n|^{k_n} \left| \mathbb{E}\, \epsilon_1^{k_1} \cdots \epsilon_n^{k_n} \right|.$$

The inner expectation vanishes unless each k_i is even, when it is trivially one. Therefore the above quantity is exactly

$$\sum_{l_1 + \cdots + l_n = m} \binom{2m}{2l_1, \ldots, 2l_n} |x_1|^{2l_1} \cdots |x_n|^{2l_n} \leq m^m \left(\sum_{i=1}^n |x_i|^2 \right)^m,$$

since for any $l_1 + \cdots + l_n = m$,

$$\binom{2m}{2l_1, \ldots, 2l_n} \leq m^m \binom{m}{l_1, \ldots, l_n}.$$

This can be seen, for example, by writing both sides out using factorials, yielding

$$\frac{(2m)!}{(2l_1)! \cdots (2l_n)!} \leq \frac{(2m)!}{2^m m!} \frac{m!}{l_1! \cdots l_n!} \leq m^m \frac{m!}{l_1! \cdots l_n!}.$$

In particular,

$$S \leq m^m \mathop{\mathbb{E}}_{X_i, Y_i} \left(\sum_i |X_i - Y_i|^2 \right)^m.$$

We now apply Hölder's inequality, first in the form $|a - b|^2 \leq 2(|a|^2 + |b|^2)$ to get

$$S \leq 2^m m^m \mathop{\mathbb{E}}_{X_i, Y_i} \left(\sum_i |X_i|^2 + \sum_i |Y_i|^2 \right)^m,$$

and secondly in the form $(a + b)^m \leq 2^{m-1}(a^m + b^m)$ to get

$$S \leq 4^m m^m \mathbb{E} \left(\sum_i |X_i|^2 \right)^m$$

as required. □

10. Almost-periodicity via random sampling in Fourier space

In this section we will prove Theorem 10, which finds a large Bohr sets of almost-periods for a function f with small $\|\widehat{f}\|_1$. The proof breaks down into two stages: we will first find some function g such that $f \approx g$ (in the L^{2m} sense) and g is defined using only a small number of characters. We then observe that if t is taken from the Bohr set that approximately annihilates these characters then t must be an almost-period for g (since then $\gamma(x+t) \approx \gamma(x)$ for all x), and hence also for f, since then

$$\tau_t f \approx \tau_t g \approx g \approx f.$$

Lemma 26. *Let $\epsilon \in (0,1)$ and $m \geq 1$. For any function $f : G \to \mathbb{C}$ there is $k = O(m\epsilon^{-2})$ and (not necessarily distinct) $\gamma_1, \ldots, \gamma_k \in \widehat{G}$ together with $c_1, \ldots, c_k \in \mathbb{C}$ with $|c_i| = 1$ such that, if*

$$g(x) = \frac{\|\widehat{f}\|_1}{k} \sum_{i=1}^{k} c_i \gamma_i(x),$$

then

$$\|g - f\|_{2m} \leq \epsilon \|\widehat{f}\|_1 N^{1/2m}.$$

Proof. Dilating f by a constant if necessary, without loss of generality, we can assume that $\|\widehat{f}\|_1 = 1/N$, so that $\sum_\gamma \left|\widehat{f}(\gamma)\right| = 1$. This naturally suggests a probability distribution on the space of all characters, where we choose $\gamma \in \widehat{G}$ with probability $\left|\widehat{f}(\gamma)\right|$.

We consider the random function $h(x) = \frac{1}{N} c_\gamma \gamma(x)$ where $\gamma \in \widehat{G}$ is chosen as above, and $c_\gamma \in \mathbb{C}$ is such that $\widehat{f}(\gamma) = c_\gamma \left|\widehat{f}(\gamma)\right|$. The key observation is that, for any $x \in G$, the expectation of $h(x)$ under this probability distribution is just $f(x)$, since

$$\mathbb{E}\, h(x) = \frac{1}{N} \sum_\gamma \left|\widehat{f}(\gamma)\right| c_\gamma \gamma(x) = \frac{1}{N} \sum_\gamma \widehat{f}(\gamma)\gamma(x) = f(x).$$

In particular, for any fixed x, the random variable $h(x) - f(x)$ has mean zero. Thus we can apply the Marcinkiewicz-Zygmund inequality, applied to k independently sampled such $h_1, \ldots, h_k$. Thus we have, for any fixed x,

$$\mathbb{E} \left| \frac{1}{k} \sum_{i=1}^{k} (h_i(x) - f(x)) \right|^{2m} \leq (16m/k)^m\, \mathbb{E}\, \frac{1}{k} \sum_{i=1}^{k} \left|h_i(x) - f(x)\right|^{2m}.$$

Note that, for any x, by the triangle inequality,

$$|f(x)| = \frac{1}{N} \left| \sum_\gamma \widehat{f}(\gamma)\gamma(x) \right| \leq 1/N,$$

and for any h_i we have $|h_i(x)| = 1/N$, and so

$$\mathbb{E} \left| \frac{1}{k} \sum_{i=1}^{k} h_i(x) - f(x) \right|^{2m} \leq (64m/k)^m N^{-2m}.$$

Summing both sides over $x \in G$ and changing the order of expectation and sum yields, with $g = \frac{1}{k}\sum_{i=1}^{k} h_i$,

$$\mathbb{E}\sum_{x\in G}\left|\frac{1}{k}\sum_{i=1}^{k} h_i(x) - f(x)\right|^{2m} = \mathbb{E}\|g-f\|_{2m}^{2m}$$

$$\leq (64m/k)^m N^{-2m+1}.$$

In particular, by the pigeonhole principle, there must exist some $\gamma_1,\ldots,\gamma_k$ (not necessarily distinct) such that, if

$$g(x) = \frac{1}{Nk}\sum_{i=1}^{k} c_{\gamma_i}\gamma_i(x)$$

then

$$\|g-f\|_{2m} \leq (64m/k)^{1/2}N^{-1+1/2m} = (64m/k)^{1/2}\|\widehat{f}\|_1 N^{1/2m}.$$

Choosing $k = \lceil 64m/\epsilon^2\rceil$ completes the proof. □

Theorem 15. *Let $\epsilon \in (0,1)$ and $m \geq 1$. For any function $f : G \to \mathbb{C}$ there is a Bohr set B of rank $O(m\epsilon^{-2})$ and radius $\Omega(\epsilon)$ such that, for all $t \in B$,*

$$\|\tau_t f - f\|_{2m} \leq \epsilon\|\widehat{f}\|_1 N^{1/2m}.$$

Proof. By Lemma 26 there exist $\gamma_1,\ldots,\gamma_k \in \widehat{G}$ with $k \ll m\epsilon^{-2}$ and $c_i \in \mathbb{C}$ with $|c_i|$ such that the g defined there satisfies

$$\|g-f\|_{2m} \leq \tfrac{1}{3}\epsilon\|\widehat{f}\|_1 N^{1/2m}.$$

Let $B = \mathrm{Bohr}(\{\gamma_1,\ldots,\gamma_k\};\epsilon/3)$. For any $t \in B$ and $x \in G$, by the triangle inequality,

$$|g(x+t)-g(x)| \leq \frac{\|\widehat{f}\|_1}{k}\sum_{i=1}^{k}|1-\gamma_i(t)| \leq \frac{\epsilon}{3}\|\widehat{f}\|_{\ell^1}.$$

In particular, for any $t \in B$,

$$\|\tau_t g - g\|_{2m} \leq N^{1/2m}\|\tau_t g - g\|_\infty \leq \frac{\epsilon}{3}\|\widehat{f}\|_{\ell^1}N^{1/2m}.$$

By the triangle inequality, therefore,

$$\|\tau_t f - f\|_{2m} \leq \|\tau_t f - \tau_t g\|_{2m} + \|\tau_t g - g\|_{2m} + \|g-f\|_{2m} \leq \epsilon\|\widehat{f}\|_1 N^{1/2m}.$$

□

11. Almost-periodicity via random sampling in physical space

In this section we will prove our other main almost-periodicity result, Theorem 12, which finds a large set of almost-periods for $1_A * 1_B$. The first part of the proof is very similar to the previous section: by random sampling we will find some function g such that $1_A * 1_B \approx g$ (in an L^{2m}-sense). Instead of finding g by randomly sampling Fourier space, however, we will randomly sampling physical space. This is suggested by writing

$$1_A * 1_B(x) = \sum_{a\in A} 1_B(x-a).$$

This suggests that if $A' \subset A$ is a random subset of A, then we might expect

$$1_A * 1_B(x) \approx \frac{|A|}{|A'|}\sum_{a\in A'} 1_B(x-a).$$

We will use the Marcinkiewicz-Zygmund inequality to show that this is true. It is slightly easier to work with tuples from A^k rather than subsets of A (in particular so that we allow repetitions in A').

We will also require something slightly stronger than we needed in the previous section: not just that there exists some such randomly chosen g with $1_A * 1_B \approx g$, but even more, that this is true for almost-all such g.

Let $k \geq 1$. If $\vec{a} \in A^k$ then, for any function f, we write

$$\mu_{\vec{a}} * f(x) = \frac{1}{k}\sum_{i=1}^{k} f(x-a_i).$$

It is also convenient to write $\mu_A = \frac{1}{|A|}1_A$. The following well-known inequality, a special case of Young's inequality for convolutions, will be useful.

Lemma 27 (Young's inequality)**.** *For any $f, g : G \to \mathbb{C}$ and any $p \geq 1$,*

$$\|f * g\|_p \leq \|f\|_p\|g\|_1.$$

Proof. We have

$$\|f * g\|_p^p = \sum_{x\in G} |f * g(x)|^p = \sum_{x\in G}\left|\sum_{y\in G} f(y)g(x-y)\right|^p.$$

By Hölder's inequality, (applied to the product $|f(y)|\,|g(x-y)|^{1/p} \cdot |g(x-y)|^{1-1/p}$)

$$\left|\sum_{y\in G} f(y)g(x-y)\right| \leq \left(\sum_{y\in G}|f(y)|^p\,|g(x-y)|\right)^{1/p}\left(\sum_{y\in G}|g(x-y)|\right)^{1-1/p},$$

and so

$$\|f * g\|_p^p \leq \|g\|_1^{p-1}\sum_{x,y\in G}|f(y)|^p\,|g(x-y)| = \|g\|_1^p\|f\|_p^p$$

as required. □

Lemma 28. *Let $\epsilon > 0$ and $m \geq 1$. Let $A \subset G$ and $f : G \to [0,1]$. If $k \geq 256m\epsilon^{-2}$ then the set*

$$\{\vec{a} \in A^k : \|\mu_{\vec{a}} * f - \mu_A * f\|_{2m} \leq \epsilon\|f\|_{2m}\}$$

has size $\geq (1-2^{-2m})\,|A|^k$.

Proof. We will show that if $\vec{a} \in A^k$ is sampled uniformly at random then, if $k \geq 256m\epsilon^{-2}$, we have

$$\mathbb{E}\|\mu_{\vec{a}} * f - \mu_A * f\|_{2m}^{2m} \leq (\epsilon/2)^{2m}\|f\|_{2m}^{2m}.$$

The lemma then follows from this by averaging (also known as Markov's inequality): if L is the set in question, then the contribution to this expectation from those $\vec{a} \notin L$ is at least

$$\left(1 - \frac{|L|}{|A|^k}\right)(\epsilon\|f\|_{2m})^{2m},$$

and hence comparing this to our upper bound we have

$$\left(1 - \frac{|L|}{|A|^k}\right) \leq 2^{-2m}$$

as required.

Note that if $a \in A$ is chosen uniformly at random then, for any fixed $x \in G$,

$$\mathbb{E} f(x - a_i) = \frac{1}{|A|} \sum_{a \in A} f(x - a) = \frac{1}{|A|} 1_A * f(x) = \mu_A * f(x).$$

Therefore, if we choose $a_1, \ldots, a_k \in A$ independently uniformly at random, for any fixed $x \in G$ and $1 \leq i \leq k$, the random variable $f(x - a_i) - f * \mu_A(x)$ has mean zero. By the Marcinkiewicz-Zygmund inequality Lemma 25, therefore,

$$\mathbb{E}\left|\frac{1}{k}\sum_i f(x - a_i) - f * \mu_A(x)\right|^{2m} \leq$$

$$(16m/k)^m k^{-1} \mathbb{E} \sum_i |f(x - a_i) - f * \mu_A(x)|^{2m}.$$

We now sum both sides over all $x \in G$. By the triangle inequality, for any fixed $1 \leq i \leq k$ and $a_i \in A$,

$$\begin{aligned}\sum_{x \in G} |f(x - a_i) - f * \mu_A(x)|^{2m} &\leq 2^{2m-1} \sum_{x \in G} |f(x - a_i)|^{2m} + \sum_{x \in G} |f * \mu_A(x)|^{2m} \\ &\leq 2^{2m-1}\left(\|f\|_{2m}^{2m} + \|f * \mu_A\|_{2m}^{2m}\right).\end{aligned}$$

We note that $\|\mu_A\|_1 = \frac{1}{|A|}\sum_{x \in A} 1_A(x) = |A| / |A| = 1$, and hence by Young's inequality, $\|f * \mu_A\|_{2m} \leq \|f\|_{2m}$, and so

$$\sum_{x \in G} |f(x - a_i) - f * \mu_A(x)|^{2m} \leq 2^{2m} \|f\|_{2m}^{2m}.$$

It follows that

$$\mathop{\mathbb{E}}_{a_1, \ldots, a_k \in A} \|\frac{1}{k}\sum_i \tau_{a_i} f - f * \mu_A\|_{2m}^{2m} \leq (64m/k)^m \|f\|_{2m}^{2m}.$$

In particular, if $k \geq 256\epsilon^{-2}m$ then the right-hand side is at most $(\frac{\epsilon}{2}\|f\|_{2m})^{2m}$ as required. □

We have do a little more work than in the Fourier case, since there is no obvious large set of almost-periods available for $\mu_{\vec{a}} * f$, even if $\vec{a}$ is a short vector. We will have to use, in an essential way, both the small doubling assumption $|A + S| \leq K|A|$, and also the fact that we have found *many* 'good' approximations $\mu_{\vec{a}} * f$. We will combine these to show that the set of 'good' vectors in A^k is approximately closed under diagonal translations, in the sense that there are many diagonal vectors $(t, \ldots, t)$ for which both $\vec{a}$ and $\vec{a} + (t, \ldots, t)$ are 'good'. But then it follows that

$$\mu_{\vec{a}} * f \approx \mu_A * f \approx \mu_{\vec{a}+(t,\ldots,t)} * f = \tau_t(\mu_{\vec{a}} * f),$$

and so t is an almost-period for $\mu_{\vec{a}} * f$, and hence for $\mu_A * f$.

Theorem 16. *Let $\epsilon > 0$ and $m \geq 1$. Suppose that A and S are such that $|A+S| \leq K|A|$. Then, for any function f, if T is the set of L^{2m}-almost periods of $\mu_A * f$ with error $\epsilon\|f\|_{2m}$, we have*

$$\langle 1_S \circ 1_S, 1_T \rangle \gg K^{-O(m\epsilon^{-2})} |S|^2 .$$

We will first show how to prove Theorem 12 from this, which is almost immediate.

Proof of Theorem 12. Let $f = 1_B$ in Theorem 16, so that $\|f\|_{2m} = |B|^{1/2m}$. If T is the set of L^{2m}-almost-periods of $\mu_A * 1_B$ with error $\epsilon |B|^{1/2m}$, then by Theorem 16,

$$\sum_{t\in T} 1_S \circ 1_S(t) = \langle 1_S \circ 1_S, 1_T \rangle \gg K^{-O(m\epsilon^{-2})} |S|^2 .$$

But trivially $1_S \circ 1_S(t) \leq |S|$, and so $|T| \gg K^{-O(m\epsilon^{-2})} |S|$. Finally, we note that if by dilation of norms, for any $t \in T$,

$$\|\tau_t(1_A * 1_B) - 1_A * 1_B\|_{2m} = |A| \, \|\tau_t(\mu_A * 1_B) - \mu_A * 1_B\|_{2m} \leq \epsilon |A| |B|^{1/2m}$$

as required. □

Proof of Theorem 16. Let $k = \lceil 1024 m\epsilon^{-2} \rceil$. By Lemma 28, if $L \subset A^k$ is the set of $\vec{a} \in A^k$ such that

$$\|\mu_{\vec{a}} * f - f * \mu_A\|_{2m} \leq \tfrac{1}{2}\epsilon\|f\|_{2m},$$

then $|L| \geq (1 - 2^{-2m}) |A|^k$.

The key observation is that if $(t, \ldots, t) \in L - L$, or in other words, there is $\vec{a}(t) \in L$ such that $\vec{a}(t) + (t, \ldots, t) \in L$, then such a $t \in T$. Indeed,

$$\begin{aligned}\|\tau_t(\mu_{\vec{a}(t)} * f) - f * \mu_A\|_{2m} &= \|\mu_{\vec{a}(t)+(t,\ldots,t)} * f - f * \mu_A\|_{2m} \\ &\leq \tfrac{1}{2}\epsilon\|f\|_{2m},\end{aligned}$$

since $\vec{a}(t) + (t, \ldots, t) \in L$, and similarly

$$\begin{aligned}\|\tau_t(\mu_{\vec{a}(t)} * f) - \tau_t(f * \mu_A)\|_{2m} &= \|\mu_{\vec{a}(t)} * f - f * \mu_A\|_{2m} \\ &\leq \tfrac{1}{2}\epsilon\|f\|_{2m},\end{aligned}$$

since $\vec{a}(t) \in L$. It follows by the triangle inequality that

$$\|\tau_t(f * \mu_A) - f * \mu_A\|_{2m} \leq \epsilon\|f\|_{2m}$$

as required.

So understanding T becomes understanding which diagonal vectors appear in $L - L$. Since $L \subset A^k$ is very large, in fact almost all of all possible k-tuples in A^k, one might expect that in the difference set $L - L$ we could find many diagonal vectors $(t, \ldots, t)$. This is not true in general, since A might be very sparse – for example, if $A = \{1, 2, 4, \ldots, 2^m\}$ then the only non-zero diagonal vectors in $A^k - A^k$ come from differences of diagonal vectors $(a, \ldots, a)$. Since $L \subset A^k$ could be very large while not containing any of the $|A|$ diagonal vectors, we cannot be sure that $L - L$ contains any non-zero diagonal vectors at all.

This is where the small doubling assumption $|A + S| \leq K|A|$ comes in handy. Let $\Delta = \{(s, \ldots, s) : s \in S\} \subset S^k$. Note that $\Delta - \Delta = \{(s', \ldots, s') : s' \in S - S\}$ is also a set of diagonal vectors, and that $1_\Delta \circ 1_\Delta(s', \ldots, s') = 1_S \circ 1_S(s')$.

Since $L + \Delta \subset (A + S)^k$, we have $|L + \Delta| \leq K^k |A|^k \leq 2K^k |L|$, say. By the Cauchy-Schwarz inequality, therefore, used just as when we bounded the additive energy from below in terms of the sumset, we have

$$\langle 1_L \circ 1_L, 1_\Delta \circ 1_\Delta \rangle = \|1_L * 1_\Delta\|_2^2 \geq \frac{|L|^2 |\Delta|^2}{|L + \Delta|} \geq \frac{|L| |S|^2}{2K^k}.$$

The left-hand side is equal to

$$\sum_{\vec{a} \in \Delta - \Delta} 1_L \circ 1_L(\vec{a}) 1_\Delta \circ 1_\Delta(\vec{a}).$$

Since $\Delta - \Delta$ is only supported on diagonal vectors, however, this is equal to

$$\sum_{t \in S - S} 1_L \circ 1_L(t, \dots, t) 1_S \circ 1_S(t).$$

As noted above, if $(t, \dots, t) \in L - L$ then $t \in T$, and since trivially $1_L \circ 1_L \leq |L|$, this is at most

$$|L| \sum_{t \in S - S} 1_T(t) 1_S \circ 1_S(t) = |T| \langle 1_T, 1_S \circ 1_S \rangle,$$

whence

$$\langle 1_T, 1_S \circ 1_S \rangle \geq \frac{1}{2K^k} |S|^2,$$

and the proof is complete, recalling our choice of k. □

CHAPTER 4

Inverse sumset results

In this final chapter we will prove one of the cornerstone results of additive combinatorics, the Freiman-Ruzsa inverse sumset theorem. Roughly speaking, this says that every $A \subset \mathbb{Z}$ with small doubling must be efficiently contained in a (generalised) arithmetic progression.

We will state precisely what this means, and the theorem we will prove, below. First, note that we have already proved something similar for subsets of $\mathbb{F}_p^n$ in Chapter 1: Theorem 1 says that if $A \subset \mathbb{F}_p^n$ has doubling $|A+A| \leq K|A|$ then there is some coset of a subgroup H such that $A \subset H$ and $|H| \ll_{p,K} |A|$.

This proof does not work in $\mathbb{Z}$, or even in $\mathbb{Z}/N\mathbb{Z}$ for N prime, because we crucially used the fact that $\mathbb{F}_p^n$ has bounded torsion. We will have to work a little harder to obtain a result over the integers.

Inverse sumset results of this type, that say "if A has small doubling then we can efficiently contain A in some structured object" are often referred to as Freiman-Ruzsa results. The reason is that, for the integers, such results were first obtained by Freiman in the 1960s, in work that was mostly overlooked at the time. Ruzsa rediscovered inverse sumset results in the 1990s with a simpler proof that was much more generalisable, and Ruzsa's papers were among those that started the modern age of additive combinatorics.

Before we state precisely the inverse result we will prove, we need to define what the structured objects we will use are. We saw at the beginning of Chapter 1 that arithmetic progressions are classic examples of sets of integers with small doubling. This quickly leads to other sets with small doubling, since as we have seen, the sumset of two sets with small doubling also has small doubling. Thus the correct notion of 'structured sets' is to consider all sets generated by taking sums of arithmetic progressions, which leads to the notion of an 'generalised arithmetic progression'.

Just as an arithmetic progression is a translated and scaled copy of a 1-dimensional interval in $\mathbb{Z}$, a generalised arithmetic progression (often abbreviated to GAP) of rank d is a translated and scaled copy of a d-dimensional cuboid.

Definition 8 (Generalised Arithmetic Progressions)**.** Let G be any abelian group. A generalised arithmetic progression (GAP) of rank d is the sum set of d arithmetic progressions $P_1 + \cdots + P_d$.
Equivalently, a GAP of rank d is a set P of the shape

$$P = \{a + n_1 v_1 + \cdots + n_d v_d : 0 \leq n_i < N_i \text{ for } 1 \leq i \leq d\}$$

for some $N_1, \dots, N_d \geq 1$ and $a, v_1, \dots, v_d \in G$. The volume of P is $\prod_i N_i$, and we say that P is proper if $|P| = \prod_i N_i$. We say that P is symmetric if $P = -P$.

We will first show that the space of GAPs is closed under addition and multiplication, and that GAPs (of bounded rank) do indeed have small doubling.

Lemma 29. *If P is a GAP of rank d then for any $k, l \geq 0$, the set $kP - lP$ is also a GAP of rank d, and*

$$|kP - lP| \leq (k+l)^d \, |P| .$$

In particular, $|P + P| \leq 2^d \, |P|$.

Proof. We can write explicitly

$$kP - lP = \{(k-l)a - l((N_1 - 1)v_1 + \cdots + (N_d - 1)v_d) + n_1 v_1 + \cdots + n_d v_d : \\ 0 \leq n_i < (k+l)N_i - (k+l) + 1\}.$$

In particular, this is also a GAP of rank d. It is clear from this that $kP - lP$ is contained in the union of the translates

$$(k - l - 1)a - l(N_1 v_1 + \cdots + N_d v_d) + v' + P,$$

where v' ranges over all $(k+l)^d$ many sums of the form

$$\sum_{i=1}^{d} c_i N_i v_i \text{ where } c_i \in \{0, 1, \dots, k+l-1\}.$$

In particular, $|kP - lP| \leq (k+l)^d \, |P|$ as required. □

In particular, we have the following trivial result.

Theorem 17. *If P is a GAP of rank d and $A \subset P$ with size $|A| \geq K^{-1} |P|$ then $|A + A| \ll_{K,d} |A|$.*

Proof. This is just

$$|A + A| \leq |P + P| \leq 2^d \, |P| \leq 2^d K \, |A| .$$

□

Our goal is the following converse result to this, that says that being contained in a small GAP of low rank is in fact the *only* way a set of integers can have small doubling!

Theorem 18 (Freiman-Ruzsa inverse theorem)**.** *Let $K \geq 4$. If $A \subset \mathbb{Z}$ is a finite set and $|A + A| \leq K \, |A|$ then there is a generalised arithmetic progression P of rank $O_K(1)$ and size $|P| \ll_K |A|$ such that $A \subset P$.*

We will prove the following strong quantitative version of this fact, which is due to Sanders.

Theorem 19 (Freiman-Ruzsa-Sanders quantitative inverse theorem)**.** *Let $K \geq 4$. If $A \subset \mathbb{Z}$ is a finite set and $|A+A| \leq K|A|$ then $A \subset P$ for some GAP P of rank at most $K(\log K)^{O(1)}$, where*

$$|A| \geq 2^{-K(\log K)^{O(1)}}|P|.$$

The bounds we give here, due to Sanders, are essentially the best known (although we have not bothered to keep track of the constant in the exponent of $\log K$ – Sanders has shown one can take this to be arbitrarily close to 3). They are in fact not too far off from the best possible bounds. Consider, for example, the case when $A = \{1, 2, \ldots, 2^{K-1}\}$. Since $|A| = K$, we trivially have $|A+A| \leq K|A|$. There is not really any non-trivial way to contain A in a GAP, due to its geometric growth. The obvious choice is the progression of rank 1 $\{1, 2, 3, \ldots, 2^{K-1}\}$, whjich has size $|P| \gg (2^K/K)|A|$. The bounds in Theorem 19 are only a power of $\log K$ away from these best possible bounds. It is conjectured that one can replace these $(\log K)^{O(1)}$ factors by just an error $O(1)$. This is one of the most important open problems in additive combinatorics.

We will deduce Theorem 19 from the following result, which finds a large GAP inside $4A - 4A$. The elementary techniques from Chapter 1 will allow us to quickly deduce the full Theorem 19.

Lemma 30 (Bogolyubov-Ruzsa Lemma)**.** *If $K \geq 4$ and $A \subset \mathbb{Z}$ has $|A+A| \leq K|A|$ then $4A - 4A$ contains a proper GAP of rank $O((\log K)^{O(1)})$ and size $\gg \exp(-O(\log K)^{O(1)})|A|$.*

Proof of Theorem 19 assuming Lemma 30. Let $|A+A| \leq K|A|$. By Lemma 30 there is a progression P of rank $d \ll (\log K)^{O(1)}$ and size $|P| \gg \exp(-(\log K)^{O(1)})|A|$ such that $P \subset 4A - 4A$.

In particular, we also have a good upper bound for the size of P, since $|P| \leq |4A - 4A| \leq K^8|A|$, by Plünnecke's inequality. We will now show that A is efficiently span-covered by $P - P$, using Lemma 10. For this we need to control $|A+P|$, which is easily done by Plünnecke again:

$$|A+P| \leq |5A - 4A| \leq K^9|A| \leq \exp((\log K)^{O(1)})|P|.$$

In particular, by Lemma 10, A is $O(K(\log K)^{O(1)})$-span covered by $P - P$. That is, there are $r \ll K(\log K)^{O(1)}$ many $x_1, \ldots, x_r$ (not necessarily distinct) such that

$$A \subset P - P + \{c_1x_1 + \cdots + c_rx_r : -1 \leq c_i \leq 1\} = Q,$$

say. It is easy to see that, since $P - P$ is a GAP of rank d, the right-hand side Q is also a GAP of rank $\leq d + r$, and

$$|Q| \leq |P - P|\,3^r \leq 3^r2^d|P| \leq \exp(K(\log K)^{O(1)})|A|$$

as required. (Note that since $K \geq 4$ we can write $O((\log K)^{O(1)})$ as $\leq (\log K)^{O(1)}$ by increasing the constant in the exponent if necessary.) □

The rest of this chapter will be spent developing the tools to prove Lemma 30, which will use both almost-periodicity and Bohr sets. We will do this in three stages:

(1) Reduce to a similar statement where instead of considering $A \subset \mathbb{Z}$ with $|A+A| \leq K|A|$ we consider $A \subset G$ with G a finite abelian group of order N and $|A| \geq K^{-1}N$.

(2) Use almost-periodicity and manipulations with the Fourier transform to show that under these hypotheses $4A-4A$ contains a Bohr set of low rank.
(3) Use geometry of numbers to show that a Bohr set of low rank must contain a GAP of low rank.

12. Freiman homomorphisms

When studying arithmetic progressions, we have already seen the necessity to move from considering subsets of $\{1,\ldots,N\}$ to subsets of $\mathbb{Z}/M\mathbb{Z}$, in part so that we could use Fourier analysis over finite abelian groups. We will need to do something similar, but more involved, here.

We first introduce the notion of a 'Freiman homomorphism'. Just as homomorphisms in group theory preserve the group structure (and so two isomorphic groups are considered equivalent from the point of view of group theory), Freiman homomorphisms preserve the kind of 'approximate structure' that is the subject of additive combinatorics. We will then show that every set of integers with small doubling is Freiman isomorphic to a large subset of a cyclic group.

Definition 9 (Freiman homomorphism)**.** Let $s \geq 1$. If A, B are subsets of some (possibly distinct) abelian groups, then we say that a function $\phi : A \to B$ is a Freiman s-homomorphism if for any $x_1,\ldots,x_s,y_1,\ldots,y_s \in A$, if

$$x_1+\cdots+x_s = y_1+\cdots+y_s$$

then

$$\phi(x_1)+\cdots+\phi(x_s) = \phi(y_1)+\cdots+\phi(y_s).$$

We say that an s-homomorphism is an s-isomorphism, and then that A and B are s-isomorphic, if ϕ is a bijection and its inverse is also an s-homomorphism, so that the "if...then" above can be upgraded into an "if and only if".

- Every function is a 1-homomorphism. Every bijection is a 1-isomorphism.
- An s-homomorphism is automatically a t-homomorphism for any $t \leq s$.
- If ϕ is an s-homomorphism then $\phi(x)+t$ is also an s-homomorphism, for any t.
- Any constant shift of a group homomorphism is an s-homomorphism for all $s \geq 1$.
- If A and B are r-isomorphic with $r = s(k+l)$ then $kA-lA$ and $kB-lB$ are s-isomorphic.
- The property of being k-isomorphic is translation invariant, in that if A and B are k-isomorphic then so are $A+x$ and $B+y$ for any x, y.

An important example of a Freiman isomorphism is given by the quotient map, which we have already used a couple of times implicitly. Let $M \geq 1$, and consider the quotient map $\phi : \mathbb{Z} \to \mathbb{Z}/M\mathbb{Z}$. This is a group homomorphism, and hence in particular is a s-homomorphism for any $s \geq 1$.

More significantly, if $M \geq kN$ then ϕ is a k-isomorphism on $\{1,\ldots,N\}$. Indeed, if $a_1,\ldots,a_k,b_1,\ldots,b_k \in \{1,\ldots,N\}$ and $a_1+\cdots+a_k \equiv b_1+\cdots+b_k \pmod{M}$ then $a_1+\cdots-b_k$ is divisible by M, but it is in $(-kN,kN)$, and hence is less than M in absolute value, so must be equal to 0.

In particular, any subset of $\{1,\ldots,N\}$ is k-isomorphic to a subset of $\mathbb{Z}/M\mathbb{Z}$ for any $M \geq kN$.

Lemma 31. *If A and B are r-isomorphic with $r = s(k+l)$ then $kA - lA$ and $kB - lB$ are s-isomorphic.*

Proof. Let $\phi : A \to B$ be a r-isomorphism. We define a new map $\phi' : kA - lA \to kB - lB$ by

$$\phi'(a_1 + \cdots + a_k - a_{k+1} - \cdots - a_{k+l}) = \phi(a_1) + \cdots - \phi(a_{k+l}).$$

This is well-defined since if $a_1 + \cdots - a_{k+l} = a_1' + \cdots - a_{k+l}'$ then $\phi(a_1) + \cdots - \phi(a_{k+l}) = \phi(a_1') + \cdots - \phi(a_{k+l}')$ since ϕ is, in particular a $(k+l)$-isomorphism. It is straightforward to similarly check that ϕ' is also a s-isomorphism. □

Lemma 32. *If $\phi : A \to B$ is a 2-homomorphism and $P \subset A$ is a GAP then $\phi(P)$ is also a GAP of the same rank and volume. If ϕ is a 2-isomorphism then $\phi(P)$ also has the same size. In particular, if P is proper then $\phi(P)$ is also proper.*

Proof. Let

$$P = \{a + n_1v_1 + \cdots + n_dv_d : 0 \leq n_i < N_i\}.$$

Since ϕ is a 2-homomorphism, for any $1 \leq j \leq d$, whenever $a + x + v_j, a + x \in A$, we have

$$\phi(a + x + v_j) = \phi(a + x) + (\phi(a + v_j) - \phi(a)).$$

since we always have $a, a + v_j \in A$ as $P \subset A$. By induction, therefore, for any $0 \leq n_i < n_i$,

$$\phi(a + n_1v_1 + \cdots + n_dv_d) = \phi(a) + n_1(\phi(a + v_1) - \phi(a)) + \cdots + n_d(\phi(a + v_d) - \phi(a)).$$

Therefore,

$$\phi(P) = \{\phi(a) + n_1w_1 + \cdots + n_dw_d : 0 \leq n_i < N_i\}$$

where $w_j = \phi(a + v_j) - \phi(a)$. This is a progression with the same rank and volume. If ϕ is an 2-isomorphism, then in particular it is a bijection, and so $|\phi(P)| = |P|$. □

Any finite $A \subset \mathbb{Z}$ is k-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$ for some N – indeed, if $A \subset [-M, M]$, then by translation invariance and the above reduction map, we see that A is k-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$ for any $N \geq k(2M+1)$. This might be a very poor bound in general, however – if A is very widely scattered amongst the integers then M might be much larger than the size of A. It is far more useful if we can find some N such that A is k-isomorphic to a *large* subset of $\mathbb{Z}/N\mathbb{Z}$.

The following modelling lemma, due to Ruzsa, gives us such an isomorphism, provided we have some control over the size of the sumsets of A, and further provided that we are willing to pass to some reasonably large subset of A.

Lemma 33 (Ruzsa modelling lemma)**.** *Let $A \subset \mathbb{Z}$ be a finite subset and $k \geq 2$. Suppose that $|A + A| \leq K|A|$. Then there is some prime N and $A' \subset A$ with $|A'| \geq |A|/k$ such that A' is k-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$ of size $\geq N/(2kK^{2k})$.*

Proof. We will show that for any $N \geq |kA - kA|$ there is a set $A' \subset A$ with $|A'| \geq |A|/k$ such that A' is k-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$. The lemma then follows from Plünnecke's inequality, which implies that $|kA - kA| \leq K^{2k}|A|$, and Bertrand's postulate.

Our isomorphism will be $\phi : A \to \mathbb{Z}/N\mathbb{Z}$ defined by

$$\phi(x) = \lfloor \xi x \rfloor \pmod{N},$$

where $\xi \in [0, N]$ is some fixed real number and $\lfloor \cdot \rfloor$ is the floor function, rounding down to the nearest integer. The set A' will be one of

$$A_j = \left\{ x \in A : \{\xi a\} \in \left[\frac{j-1}{k}, \frac{j}{k} \right) \right\}$$

for $1 \leq j \leq k$, where $\{t\} = t - \lfloor t \rfloor$ is the fractional part. We will show that there exists ξ such that ϕ restricted to each A_j is a k-isomorphism, and clearly at least one of them has at least $\left| A \right| / k$ members.

To show that ϕ is a k-isomorphism, we require that (with $1 \leq j \leq k$ fixed) for any $a_1, \ldots, b_k \in A_j$,

$$a_1 + \cdots + a_k = b_1 + \cdots + b_k \tag{3}$$

if and only if

$$\lfloor \xi a_1 \rfloor + \cdots + \lfloor \xi a_k \rfloor \equiv \lfloor \xi b_1 \rfloor + \cdots + \lfloor \xi b_k \rfloor \pmod{N}. \tag{4}$$

We first note that, for any $\xi \in \mathbb{R}$, since $t = \lfloor t \rfloor + \{t\}$ for any $t \in \mathbb{R}$, we have

$$\sum_{1 \leq i \leq k} (\lfloor \xi a_i \rfloor - \lfloor \xi b_i \rfloor) = \xi \sum_{1 \leq i \leq k} (a_i - b_i) - \sum_{1 \leq i \leq k} (\{\xi a_i\} - \{\xi b_i\}).$$

Furthermore, since all $a_i, b_i \in A_j$ for some fixed j, all of the fractional parts on the right-hand side lie in the same interval $[u, u + 1/k)$ for some u. In particular, the sum of the fractional parts must lie in $(-1, 1)$, and so we have

$$\sum_{1 \leq i \leq k} (\lfloor \xi a_i \rfloor - \lfloor \xi b_i \rfloor) = \xi \sum_{1 \leq i \leq k} (a_i - b_i) + \delta$$

for some $\delta \in (-1, 1)$. One direction of the isomorphism is now immediate: if (3) holds then the first sum vanishes and the right-hand side is just δ, but since the left-hand side is an integer we must have $\delta = 0$, and hence (4) holds. Note that this is true for any choice of ξ.

On the other hand, if (4) holds, then we have $mN = \xi t + \delta$ for some $t \in kA - kA$ and $m \in \mathbb{Z}$. If $t = 0$ then (3) holds as required. Otherwise, $\xi = (mN - \delta)/t$. It therefore suffices to choose any $\xi \in \mathbb{R}$ that lies outside the union of

$$\bigcup_{t \in (kA - kA) \setminus \{0\}} \bigcup_{m \in \mathbb{Z}} \left(\frac{mN - 1}{t}, \frac{mN + 1}{t} \right).$$

We will in fact show that there is some $\xi \in [0, N]$ which is left uncovered.

We first estimate the measure of what is excluded for any fixed $t > 0$. If $1 \leq m < t$ then trivially the interval $(mN - 1/t, mN + 1/t)$ excludes at most $2/t$ from $[0, N]$. For $m = 0$, note that the measure of $(-1/t, 1/t) \cap [0, N]$ is $1/t$, and similarly for $m = t$. Otherwise, if $m < 0$ or $m > t$ then the interval $(mN - 1/t, mN + 1/t)$ has empty intersection with $[0, N]$. In total, therefore, the amount excluded from $[0, N]$ has measure at most

$$(t - 1)\frac{2}{t} + \frac{1}{t} + \frac{1}{t} = 2.$$

Finally, we note that $kA - kA$ is symmetric, and the set of excluded intervals from $-t$ is identical to that of t. Therefore the total mass excluded from $[0, N]$ is at most (summing the total excluded measure for all $t \in (kA - kA) \cap \mathbb{R}_{>0}$)

$$\frac{\left| kA - kA \right| - 1}{2} \cdot 2 < N,$$

by assumption, and hence there is some $\xi \in [0, N]$ left uncovered, Any such ξ yields a suitable k-isomorphism. □

We will now use Ruzsa's modelling lemma to show that, when proving the Bogolyubov-Ruzsa lemma, Lemma 30, instead of considering $A \subset \mathbb{Z}$ with $|A+A| \ll |A|$, we can instead consider $A \subset \mathbb{Z}/N\mathbb{Z}$ with N prime and $|A| \gg N$. Namely, we will actually prove the following.

Lemma 34 (Bogolyubov-Ruzsa Lemma, Dense version). *Let $K \geq 4$ and N be prime. Suppose that $A \subset \mathbb{Z}/N\mathbb{Z}$ has size $|A| \geq K^{-1}N$. Then $4A - 4A$ contains a proper GAP of rank $O((\log K)^{O(1)})$ and size $\gg \exp(-(\log K)^{O(1)})N$.*

We will prove this in the next two sections, and we now use Ruzsa's modelling lemma to show how Lemma 30 follows.

Proof of Lemma 30 assuming Lemma 34. Let $A \subset \mathbb{Z}$ be such that $|A+A| \leq K|A|$. By Lemma 33 there is some prime N and $A' \subset A$ with $|A'| \geq |A|/16$ such that A' is 16-isomorphic to a subset of $\mathbb{Z}/N\mathbb{Z}$, say B, where $|B| \geq N/32K^{32}$.

We now apply Lemma 34 to B (with K replaced by $32K^{32}$). In particular, there is a proper GAP P inside $4B - 4B$ of rank $O((\log K)^{O(1)})$ and size $\gg \exp(-(\log K)^{O(1)})N$.

By Lemma 31, however, $4B-4B$ is 2-isomorphic to $4A'-4A'$, and 2-isomorphisms preserve both the rank and size of GAPs by Lemma 32, and so the image of P under this isomorphism yields a proper GAP of the same rank and size inside $4A' - 4A' \subset 4A - 4A$, as required. □

13. Finding a large Bohr set inside $2A - 2A$

We now need to do the following: given $A \subset \mathbb{Z}/N\mathbb{Z}$ which is large $|A| \gg N$, find a large GAP inside $4A + 4A$. We will do this in two stages: first find a large Bohr set (with small rank) inside $4A - 4A$, and then show that Bohr sets contain large GAPS. In this section we address the first task, in which almost-periodicity plays a starring role.

We will first use almost-periodicity to find a large set X such that every element of kX is a 'popular' element of $2A - 2A$. (Compare this to Theorem 13, which showed, under the assumption of small doubling, that $kX \subset 2A - 2A$ - the proof here is a simple modification.)

Lemma 35. *Let $k \geq 1$ and $K \geq 4$. If $A \subset \mathbb{Z}/N\mathbb{Z}$ with $|A| \geq N/K$ then there is some set X such that*

$$|X| \geq \exp(-O(k^2(\log K)^2))N$$

and if $x \in kX$ then

$$1_{A-A} * 1_A \circ 1_A(x) \geq \tfrac{1}{2}|A|^2 .$$

In particular, $kX \subset 2A-2A$. (In other words, not only is every $t \in kX$ an element of $A + A - A - A$, but there are many triples $a \in A - A$ and $b, c \in A$ such that $a + b - c = t$.)

Proof. We apply Theorem 11 with $B = A - A$, and $\epsilon > 0$ and $m \geq 1$ to be chosen soon. Let X be the corresponding set of almost-periods. Since addition of almost-periods adds the errors, kX is a subset of the set of L^{2m}-almost-periods for $1_A * 1_{A-A}$ with error $k\epsilon |A| |A-A|^{1/2m}$, that is, for any $x \in kX$, we have

$$\|\tau_x(1_A * 1_{A-A}) - 1_A * 1_{A-A}\|_{2m} \leq k\epsilon |A| |A-A|^{1/2m} .$$

Suppose for a contradiction that there is some $x \in kX$ such that

$$1_{A-A} * 1_A \circ 1_A(x) < \tfrac{1}{2}\left|A\right|^2.$$

Then

$$\begin{aligned}\tfrac{1}{2}\left|A\right|^2 &> \sum_{a,b\in A}\sum_{c\in A-A} 1_{a-b-c=x}\\ &= \sum_{a\in A}\sum_{b\in A}\sum_{c\in A-A} 1_{b+c=a-x}\\ &= \langle 1_A, \tau_x(1_A * 1_{A-A})\rangle.\end{aligned}$$

On other hand, we have

$$\langle 1_A, 1_A * 1_{A-A}\rangle = \sum_{a,b\in A}\sum_{c\in A-A} 1_{a-b=c} = \left|A\right|^2.$$

Taking the difference,

$$\left|\langle 1_A, (\tau_x(1_A * 1_{A-A}) - 1_A * 1_{A-A})\rangle\right| \geq \tfrac{1}{2}\left|A\right|^2.$$

On the other hand, by Hölder's inequality, the left-hand side is at most

$$\left|A\right|^{1-1/2m}\left\|\tau_x(1_A * 1_{A-A}) - 1_A * 1_{A-A}\right\|_{2m} \leq k\epsilon\left|A\right|^2(\left|A-A\right|/\left|A\right|)^{1/2m}.$$

By Plünnecke's inequality $\left|A-A\right| \leq K^2\left|A\right|$, and hence if we choose $m = \lceil \log K\rceil$, then the right-hand side is at most $ek\epsilon\left|A\right|^2$. Choosing $\epsilon = 1/4ek$, say, gives a contradiction.

Thus the theorem is proved, since the size of X is

$$\left|X\right| \geq K^{-O(m\epsilon^{-2})}N \geq \exp(-O(k^2(\log K)^2))N.$$

□

The idea will be to choose the Bohr set to annihilate those characters where $|\widehat{1_X}|$ is large (say $\geq \frac{1}{2}\left|X\right|$). To ensure that this is low rank, we first need to make sure we control the 'dimension' of the set of such characters. The following lemma was first proved (actually in a slightly stronger form) by Chang.

Lemma 36 (Weak Chang dimension bound)**.** *Let N be prime. Let $X \subset \mathbb{Z}/N\mathbb{Z}$ with density $\delta = \left|X\right|/N$. There is a multiset Γ such that $\left|\Gamma\right| \ll (\log(1/\delta))^3$ and*

$$\{\gamma : |\widehat{1_X}(\gamma)| \geq \tfrac{1}{2}\left|X\right|\} \subset \mathrm{Span}(\Gamma).$$

For comparison, note that by Parseval's identity, if Δ is the set of characters in question, then

$$\tfrac{1}{4}\left|X\right|^2\frac{\left|\Delta\right|}{N} \leq \mathop{\mathbb{E}}_{\gamma}|\widehat{1_X}(\gamma)|^2 = \left|X\right|$$

by Parseval's identity, whence $\left|\Delta\right| \leq 4\delta^{-1}$. Chang's lemma tells us that if we're only concerned about the 'dimension' of Δ, rather than its size, we can replace this δ^{-1} by a power of $\log(1/\delta)$. (The full Chang's lemma in fact allows one to take $\left|\Gamma\right| \ll \log(1/\delta)$.)

Proof. Let $m \geq 1$ be some large integer, to be chosen later. Let Δ be the set of characters in question, and suppose $\Gamma_0 \subset \Delta$ is maximal such that the only solutions to

$$\gamma_1 + \cdots + \gamma_m = \gamma_1' + \cdots + \gamma_m' \tag{5}$$

are the trivial ones with $\{\gamma_1,\ldots,\gamma_m\}=\{\gamma_1',\ldots,\gamma_m'\}$. Then if $E_{2m}(\Gamma_0)$ counts the number of solutions to (5) with the variables coming from Γ_0, we have $E_{2m}(\Gamma_0)\leq m!\left|\Gamma_0\right|^m$.

We will compare this to a lower bound on $E_{2m}(\Gamma_0)$ coming from Hölder's inequality. We have

$$\tfrac{1}{2}\left|X\right|\left|\Gamma_0\right|\leq\sum_{\gamma\in\Gamma_0}\lvert\widehat{1_X}(\gamma)\rvert=\sum_{\gamma\in\Gamma_0}c_\gamma\sum_{x\in X}\overline{\gamma(x)}=\sum_{x\in X}\mathop{\mathbb{E}}_{\gamma}c_\gamma\overline{\gamma(x)},$$

where $c_\gamma\in\mathbb{C}$ are some suitable signs chosen such that $\lvert\widehat{1_X}(\gamma)\rvert=c_\gamma\widehat{1_X}(\gamma)$. By Hölder's inequality, the right-hand side is at most

$$\left|X\right|^{1-1/2m}\left(\sum_x\left|\sum_{\gamma\in\Gamma_0}c_\gamma\overline{\gamma(x)}\right|^{2m}\right)^{1/2m}.$$

Expanding out the power and changing the order of summation,

$$\sum_x\left|\mathop{\mathbb{E}}_{\gamma}c_\gamma\overline{\gamma(x)}\right|^{2m}=\sum_{\gamma_1,\ldots,\gamma_{2m}\in\Gamma_0}c_{\gamma_1}\cdots\overline{c_{\gamma_{2m}}}\sum_x(\gamma_1+\cdots-\gamma_{2m})(-x).$$

By orthogonality, the sum over x is 0 unless $\gamma_1+\cdots-\gamma_{2m}=0$, when it is $=N$. By the triangle inequality, therefore, (since $\left|c_\gamma\right|=1$),

$$\tfrac{1}{2}\left|X\right|\left|\Gamma_0\right|\leq\left|X\right|^{1-1/2m}\left(\sum_{\gamma_1,\ldots,\gamma_{2m}\in\Gamma_0}1_{\gamma_1+\cdots-\gamma_{2m}=0}\right)^{1/2m}N^{1/2m}.$$

The sum over γ_i is exactly $E_{2m}(\Gamma_0)$. Hence, rearranging this inequality, and using the above upper bound, we have

$$2^{-2m}\delta\left|\Gamma_0\right|^{2m}\leq E_{2m}(\Gamma_0)\leq m!\left|\Gamma_0\right|^m\leq m^m\left|\Gamma_0\right|^m.$$

In particular,

$$\left|\Gamma_0\right|\leq 4m\delta^{-1/m}\ll\log(1/\delta)$$

if we choose $m=\lceil\log(1/\delta)\rceil$.

If $\lambda\in\Delta\backslash\Gamma_0$, then by maximality, there is some solution to

$$\gamma_1+\cdots+\gamma_m=\gamma_1'+\cdots+\gamma_m'$$

with $\gamma_i,\gamma_i'\in\Gamma_0\cup\{\lambda\}$ where the right-hand side is not a permutation of the left-hand side. If there is an equal number of λ on both sides of this equation then, by cancelling them and replacing them by some arbitrary $\gamma\in\Gamma_0$, we have a non-trivial solution to (5) with variables all in Γ_0, which is a contradiction.

Hence, by rearranging this, we have

$$t\lambda=\sum_{\gamma\in\Gamma_0}a_\gamma\gamma$$

where $1\leq t\leq m$ and $a_\gamma\in\mathbb{Z}$ satisfy $\left|a_\gamma\right|\leq m$. Since $t\leq m\leq\lceil\log(1/\delta)\rceil\leq\lceil\log N\rceil<N$, and N is prime, we can multiply both sides by t^{-1} to see that

$$\lambda=\sum_{\gamma\in\Gamma_0}a_\gamma(t^{-1}\gamma).$$

The proof is complete, if we let Γ be the union of all $a_\gamma t^{-1}\gamma$ as γ ranges over Γ_0. Since there are $O(m^2)$ many choices for a_γ and t^{-1}, we have $|\Gamma| \ll m^2\log(1/\delta) \ll \log^3(1/\delta)$ as required. □

Lemma 37. *If N is prime and $K \geq 4$, and $A \subset \mathbb{Z}/N\mathbb{Z}$ with $|A| \geq N/K$ then there is a Bohr set B of rank $O((\log K)^{O(1)})$ and width $\gg K^{-2}$ such that $B \subset 4A-4A$.*

Proof. Let X be the set given by Lemma 35 with $k \geq 1$ some parameter to be chosen later, so that

$$|X| \geq \exp(-O(k^2(\log K)^2))N$$

and if $x \in kX$ then

$$1_{A-A} * 1_A \circ 1_A(x) \geq \tfrac{1}{2}|A|^2 .$$

In particular, since $1_X^{(k)}$ (the k-fold iterated convolution) is supported on kX, we have

$$\langle 1_{A-A} * 1_A \circ 1_A, 1_X^{(k)}\rangle \geq \tfrac{1}{2}|A|^2 \sum_x 1_X^{(k)}(x) = \tfrac{1}{2}|A|^2|X|^k . \tag{6}$$

Let $\delta = |X|/N$ and $\Delta = \{\gamma : |\widehat{1_X}(\gamma)| \geq \frac{1}{2}|X|\}$, and let Γ be a multiset of size $O(\log^3(1/\delta))$ such that $\Delta \subset \mathrm{Span}(\Gamma)$, as given by Lemma 36. Let $B = \mathrm{Bohr}(\Gamma; \rho/|\Gamma|)$ (where we replace the multiset Γ with just its underlying set), with $\rho > 0$ some parameter which we will choose later (it will end up as $\rho = 1/8K$). We note that for any $\lambda_1, \lambda_2 \in \widehat{G}$ and any $t \in G$, by the triangle inequality,

$$|1-(\lambda_1+\lambda_2)(t)| \leq |1-\lambda_1(t)| + |\lambda_1(t)-\lambda_1(t)\lambda_2(t)| = |1-\lambda_1(t)| + |1-\lambda_2(t)|.$$

It follows by induction on k that for any $\lambda_1, \ldots, \lambda_k$ and t

$$\left|1-\left(\sum_i \lambda_i\right)(t)\right| \leq \sum_i |1-\lambda_i(t)| .$$

In particular, for any $\gamma \in \Delta$, writing

$$\gamma = \sum_{\lambda\in\Gamma} c_\lambda \lambda$$

for some $c_\lambda \in \{-1,0,1\}$, and any $t \in B$,

$$|1-\gamma(t)| \leq \sum_{\lambda\in\Gamma} |1-\lambda(t)| \leq \rho.$$

We claim that $B \subset 2A-2A+kX$. Indeed, by Fourier inversion,

$$1_{A-A} * 1_A \circ 1_A * 1_X^{(k)}(t) = \mathop{\mathbb{E}}_\gamma \widehat{1_{A-A}}(\gamma)\left|\widehat{1_A}(\gamma)\right|^2 \widehat{1_X}(\gamma)^k\gamma(t).$$

Without the $\gamma(t)$, this is

$$\mathop{\mathbb{E}}_\gamma \widehat{1_{A-A}}(\gamma)\left|\widehat{1_A}(\gamma)\right|^2 \widehat{1_X}(\gamma)^k = \langle 1_{A-A} * 1_A \circ 1_A, 1_X^{(k)}\rangle \geq \tfrac{1}{2}|A|^2|X|^k .$$

It follows that

$$\left|1_{A-A} * 1_A \circ 1_A * 1_X^{(k)}(t) - \tfrac{1}{2}|A|^2|X|^k\right| \leq \mathop{\mathbb{E}}_\gamma \left|\widehat{1_{A-A}}(\gamma)\right|\left|\widehat{1_A}(\gamma)\right|^2\left|\widehat{1_X}(\gamma)\right|^k|\gamma(t)-1| .$$

In particular, if $t \notin 2A - 2A + kX$ then, trivially bounding $\left|\widehat{1_{A-A}}\right| \leq |A-A| \leq N \leq K|A|$, we have

$$\tfrac{1}{2K}|A||X|^k \leq \mathop{\mathbb{E}}_{\gamma} \left|\widehat{1_A}(\gamma)\right|^2 \left|\widehat{1_X}(\gamma)\right|^k |\gamma(t)-1|.$$

We will obtain a contradiction by upper bounding the right-hand side. Firstly, consider the contribution from $\gamma \notin \Delta$. Then $\left|\widehat{1_X}(\gamma)\right| \leq |X|/2$ for such γ, and so this part contributes (using the trivial $|\gamma(t)-1| \leq 2$)

$$\leq 2^{1-k}|X|^k \mathop{\mathbb{E}}_{\gamma} \left|\widehat{1_A}(\gamma)\right|^2 = 2^{1-k}|X|^k|A|,$$

by Parseval's identity. On the other hand, if $\gamma \in \Delta$, then as discussed above, we have $|\gamma(t)-1| \leq \rho$, and so this part contributes (using again a trivial bound $\left|\widehat{1_X}(\gamma)\right| \leq |X|$)

$$\leq \rho|X|^k|A|.$$

Putting our two upper bounds together, we have

$$\tfrac{1}{2K}|A||X|^k \leq (2^{1-k}+\rho)|X|^k|A|.$$

Thus we get a contradiction if we choose $\rho = 1/8K$ and $k = 100\lceil \log K\rceil$, say. In particular we have a Bohr set B with $B \subset 2A - 2A + kX \subset 4A - 4A$ as required, where the rank of B is

$$\ll \log(1/\delta)^3 \ll k^6(\log K)^6 \ll (\log K)^{12}$$

and the width of B is

$$\rho/|\Gamma| \gg K^{-1}(\log K)^{-12} \gg K^{-2}.$$

□

We have made excellent progress towards proving the Freiman-Ruzsa-Sanders inverse theorem. We have found a large Bohr set B inside $4A-4A$. It remains to show that $4A-4A$ contains a large GAP, which we will do so by showing that, in general, Bohr sets contain large GAPs. For this we will undertake a brief digression into the geometry of numbers.

14. Geometry of Numbers and Progressions in Bohr sets

We have already seen that a Bohr set in $\mathbb{Z}/N\mathbb{Z}$ of rank d and width ρ contains an arithmetic progression of length $\gg \rho N^{1/d}$. This is just a 1-dimensional object, and we might hope that if we're considering generalised arithmetic progressions we should be able to do much better. Indeed, since a Bohr set of rank d is the inverse image of a d-dimensional cube, it is natural to search for a d-dimensional GAP inside a Bohr set of rank d.

The proof is a little delicate, however, and in particular a simple 'greedy' construction will not work. The best way to proceed is via the geometry of numbers, and old and fascinating subject.

Definition 10 (Lattices)**.** A lattice $L \subset \mathbb{R}^d$ of rank k is a set of the form

$$L = \left\{ \sum_{i=1}^{k} a_i v_i : a_i \in \mathbb{Z} \right\}$$

where $v_1, \ldots, v_k$ are some k linearly independent vectors in $\mathbb{R}^d$. (One can show this is equivalent to defining a lattice to be any discrete additive subgroup of $\mathbb{R}^d$, but we will not need this here.) Given a lattice L we define its fundamental parallelepiped to be

$$FP(L) = \left\{ \sum_{i=1}^{k} c_i v_i : c_i \in [0,1) \right\}.$$

The (Lebesgue) measure of $FP(L)$ is called the covolume of L, and is denoted by $\mu(\mathbb{R}^d/L)$. (Note that the covolume of L is zero if $k < d$.)

Lemma 38 (Blichfeldt's Lemma)**.** *If $L \subset \mathbb{R}^d$ is a lattice of rank d and $V \subset \mathbb{R}^d$ has $\mu(V) > \mu(\mathbb{R}^d/L)$ then there are distinct $x, y \in V$ such that $x - y \in L$.*

Proof. Let $Q = FP(L)$. For $x \in L$ consider the set $V \cap (Q + x)$. Since all translates $(Q + x)_{x \in L}$ are disjoint, and cover all of $\mathbb{R}^d$, we have

$$\sum_{x \in L} \mu(Q \cap (V - x)) = \sum_{x \in L} \mu(V \cap (Q + x)) = \mu(V) > \mu(Q).$$

It follows that the translates $Q \cap (V - x)$ cannot be disjoint, and hence in particular two of the translates $V - x$ must overlap, and there are (distinct) $v_1, v_2 \in V$ and $x_1, x_2 \in L$ such that $v_1 - x_1 = v_2 - x_2$, where $x_1 \neq x_2$. Then $v_1 - v_2 = x_1 - x_2 \in L$ as required. □

Lemma 39 (Minkowski's First Theorem)**.** *If V is a symmetric convex set and $L \subset \mathbb{R}^d$ is a lattice of rank d then, provided $\mu(V) > 2^d \mu(\mathbb{R}^d/L)$, the set V must contain a non-zero point of L.*

Proof. We apply Blichfeldt's lemma to the set $\frac{1}{2} \cdot V$, which has measure $2^{-d}\mu(V)$. This gives us $v_1, v_2 \in V$ such that $\frac{1}{2}(v_1 - v_2) \in L \backslash \{0\}$. By the symmetry and convexity of V, $\frac{1}{2}(v_1 - v_2) \in V$, and we are done. □

Minkowski's first theorem is very useful in finding a single lattice vector inside some given symmetric convex set. In particular, it tells us what radius ball around the origin we need before we can guarantee a single non-zero lattice vector. What if we want more? In particular, how large a ball around the origin do we need to take before we can find d linearly independent vectors in L?

The answer is given by Minkowski's second theorem, and the concept of successive minima. Roughly speaking, these measure how large a ball we need to take around the origin before we are guaranteed to find first one, then two, and so on, linearly independent lattice vectors.

Definition 11 (Successive minima)**.** If $L \subset \mathbb{R}^d$ is a lattice of rank d then we define the successive minima $0 < \lambda_1 \leq \cdots \leq \lambda_d < \infty$ with respect to L by setting λ_k to be the infimum of all $\lambda > 0$ such that there k linearly independent $v_1, \ldots, v_k \in L$ such that $|v_i| < \lambda$.
It is easy to check that this definition makes sense, in that the successive minima are finite and non-zero. (In fact one can define successive minima with respect to L for any convex body B, but we will only explore the simplest case presented here, which corresponds to taking B to be the unit sphere.)

For example, Minkowski's first theorem tells us that $\lambda_1^d \mu(B_d) \leq 2^d \mu(\mathbb{R}^d/L)$, where B_d is the unit ball around the origin in d dimensions. The volume of the unit ball can, of course, be calculated explicitly, but since we do not care too much about precise bounds, we use the simple fact that

$$\mu(B_d) \geq 2^d d^{-d/2},$$

which follows from the fact that it contains the cube $[-1/\sqrt{d}, 1/\sqrt{d}]^d$. We deduce that $\lambda_1 \leq d^{1/2}\mu(\mathbb{R}^d/L)^{1/d}$. Obtaining similar upper bounds on the individual λ_i is more difficult, but Minkowski's second theorem gives us, quite incredibly, the same upper bound on their geometric mean $(\lambda_1 \cdots \lambda_d)^{1/d}$. The basic idea of the proof is to transform the lattice (roughly by dilating each direction by λ_i^{-1}) into another for which the unit ball contains no non-zero lattice vectors, and apply Minkowski's first theorem.

Lemma 40 (Minkowski's second theorem)**.** *If $L \subset \mathbb{R}^d$ is a lattice of rank d and $0 < \lambda_1 \leq \cdots \leq \lambda_d$ are the successive minima with respect to L, then*

(1) *there exist d linearly independent vectors (called the directional basis) $v_1, \ldots, v_d \in L$ such that $|v_j| = \lambda_j$ and if $x \in L$ satisifes $|x| < \lambda_j$ then x is in the $\mathbb{R}$-span of $\{v_1, \ldots, v_{j-1}\}$, and*

(2)
$$\lambda_1 \cdots \lambda_d \leq 2^d \frac{\mu(\mathbb{R}^d/L)}{\mu(B_d)},$$
where B_d is the unit sphere in $\mathbb{R}^d$.

In particular, $\lambda_1 \cdots \lambda_d \leq d^{d/2}\mu(\mathbb{R}^d/L)$.

Proof. The first part follows almost immediately from the definition of successive minima. Indeed, suppose that $\mu_1 < \cdots < \mu_l$ are the distinct values taken on by the successive minima, and that $1 \leq k_1 < \cdot < k_l = d$ are such that $\mu_i = \lambda_{k_{i-1}+1} = \cdots = \lambda_{k_i}$. We claim that by induction on i we can find linearly independent $v_{k_{i-1}+1}, \ldots, v_{k_i} \in L$ such that (with $k_0 = 0$ and $k_{l+1} = \infty$)

(1) $\left|v_{k_{i-1}+1}\right| = \cdots = |v_{k_i}| = \mu_i$ and
(2) $x \in L$ that satisfies $|x| < \mu_{i+1}$ must be in the span of $v_1, \ldots, v_{k_i}$.

The case $i = 1$ is clear, since for any $\lambda < \mu_1$, there are no non-zero elements $x \in L$ such that $|x| < \lambda$, but for any $\mu_2 > \lambda > \mu_1$ there are least k_1 linearly independent vectors $x \in L$ such that $|x| < \lambda$. It follows immediately that there must be at least k_1 linearly independent vectors $v_1, \ldots, v_{k_1}$ such that $|v_1| = \cdots = |v_{k_1}| = \mu_1$. Point (2) follows since otherwise we could find $k_1 + 1$ linearly independent vectors $x \in L$

satisfying $|x| < \mu_2 = \lambda_{k_1+1}$, which would contradict the definition of λ_{k_1+1}. The general case follows similarly, and thus we have the first part.

By the Gram-Schmidt process (or general first year linear algebra) there exists an orthonormal basis $w_1, \ldots, w_d$ of $\mathbb{R}^d$ such that v_j is in the $\mathbb{R}$-span of $\{w_1, \ldots, w_j\}$. Suppose

$$v_j = \sum_{i=1}^{j} t_{ji} w_i$$

for some $t_{ji} \in \mathbb{R}$. The key observation is that if we define

$$v_j' = \sum_{i \leq j} t_{ji} \lambda_i^{-1} w_i$$

then for any integers $u_1, \ldots, u_d$ not all zero,

$$\left|\sum u_j v_j'\right|^2 = \sum_i \left(\sum_{j \geq i} u_j \lambda_i^{-1} t_{ji} \right)^2 \geq 1, \tag{7}$$

and so in particular the lattice L' generated by these v_i' (which is clearly of full rank) has no non-zero vectors inside the unit ball B_d. By Minkowski's first theorem, therefore, $\mu(\mathbb{R}^d/L') \geq 2^{-d}\mu(B_d)$. On the other hand, the fundamental parallelepiped spanned by v_i' is obtained by transforming that spanned by v_i under a linear transformation similar to the diagonal matrix with entries $\lambda_1^{-1}, \ldots, \lambda_d^{-1}$, and hence

$$2^{-d}\mu(B_d) \leq \mu(\mathbb{R}^d/L') = (\lambda_1 \cdots \lambda_d)^{-1} \mu(\mathbb{R}^d/L)$$

as required.

It remains to check (7). Let $u_1, \ldots, u_d$ be any integers, not all zero. Let $1 \leq J \leq d$ be such that $u_J \neq 0$ and $u_j = 0$ for $j > J$. Then $u_1 v_1 + \cdots + u_d v_d$ is a vector in L which is linearly independent of $v_1, \ldots, v_{J-1}$, and so

$$|u_1 v_1 + \cdots + u_d v_d| \geq \lambda_J.$$

It follows that, using the orthonormality of the w_i,

$$\begin{aligned}
\left|\sum u_j v_j'\right|^2 &= \left|\sum_{j=1}^{d}\sum_{i\leq j} u_j t_{ji}\lambda_i^{-1} w_i\right|^2\\
&= \sum_i \lambda_i^{-2}\left(\sum_{j\geq i} u_j t_{ji}\right)^2\\
&= \sum_{i\leq J} \lambda_i^{-2}\left(\sum_{j\geq i} u_j t_{ji}\right)^2\\
&\geq \sum_{i\leq J} \lambda_J^{-2}\left(\sum_{j\geq i} u_j t_{ji}\right)^2\\
&= \lambda_J^{-2}\left|\sum_j u_j v_j\right|^2\\
&\geq 1
\end{aligned}$$

as required. □

With Minkowski's second theorem, we can now prove that Bohr sets in $\mathbb{Z}/N\mathbb{Z}$ (for N prime) contain large GAPs.

Lemma 41. *If N is prime and B is a Bohr set in $\mathbb{Z}/N\mathbb{Z}$ of rank $d \geq 1$ and width ρ, then B contains a symmetric proper progression P of rank at most d and cardinality*

$$\lvert P\rvert \geq \left(\frac{\rho}{d}\right)^{O(d)} N.$$

We remark that in fact one can prove a sharper lower bound of $(\rho/cd)^d N$ for some constant $c > 0$, if one uses a slightly stronger version of Minkowski's second theorem, but this version is sufficient for our purposes.

Proof. Let $B = \mathrm{Bohr}(\Gamma, \rho)$, where Γ is a set of characters that we can identify with elements of $\mathbb{Z}/N\mathbb{Z}$. Without loss of generality we can assume that $0 \notin \Gamma$, and furthermore we may assume that $1 \in \Gamma$, since dilating the set of characters corresponds to dilating B, and the rank and size of a GAP is preserved under dilation (we are using crucially that N is prime here).

Let $\Gamma = \{\gamma_1, \ldots, \gamma_d\}$, therefore, where $\gamma_1, \ldots, \gamma_d \in \{1, \ldots, N\}$ and $\gamma_1 = 1$. We begin by recalling that, in the cyclic group $\mathbb{Z}/N\mathbb{Z}$, using the fact that $\lvert e(x) - 1\rvert \leq 2\pi\lVert x\rVert$, where $\lVert x\rVert$ is the distance of x from the nearest integer, and the explicit description of characters in $\mathbb{Z}/n\mathbb{Z}$ as corresponding to $x \mapsto e^{2\pi i \frac{\gamma x}{N}}$, we know that B contains

$$B' = \{x : x\gamma_i \equiv m_i \pmod{N} \text{ for some } \lvert m_i\rvert \leq \tfrac{\rho}{2\pi}N\}.$$

We will therefore concentrate on finding a large GAP inside B'.

Let L be the lattice in $\mathbb{R}^d$ spanned by the columns of

$$\begin{pmatrix} \gamma_1/N & 0 & \cdots & 0 \\ \gamma_2/N & 1 & \cdots & 0 \\ \vdots & & \ddots & \\ \gamma_d/N & 0 & \cdots & 1 \end{pmatrix}.$$

Since

$$\begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} = N \begin{pmatrix} \gamma_1/N \\ \gamma_2/N \\ \vdots \\ \gamma_d/N \end{pmatrix} - \gamma_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} - \cdots - \gamma_d \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix},$$

it is clear that in fact $L = \mathbb{Z}^d + \mathbb{Z} \cdot (\gamma_1/N, \ldots, \gamma_d/N)$. Let $0 < \lambda_1 \leq \cdots \leq \lambda_d$ be the successive minima with respect to L, with corresponding directional basis $v_1, \ldots, v_d$. A straightforward determinant calculation shows that $\mu(\mathbb{R}^d/L) = 1/N$, and therefore by Minkowski's second theorem,

$$\lambda_1 \cdots \lambda_d \leq d^{d/2} N^{-1}.$$

(Note also that clearly the unit ball contains d linearly independent vectors from L, and so $\lambda_i \leq 1$ for all i.)

For each $1 \leq i \leq d$, since $v_i \in L$, there must exist some $w_i \in \mathbb{Z}/N\mathbb{Z}$ such that $v_i \in \frac{w_i}{N}(\gamma_1, \ldots, \gamma_d) + \mathbb{Z}^d$. Let $N_i = \lfloor \rho/2\pi d \lambda_i \rfloor$, and define the progression P to be

$$P = \{n_1 w_1 + \cdots + n_d w_d : |n_i| \leq N_i\}.$$

This is a generalised arithmetic progression of rank d and volume $\prod_{i=1}^d (2N_i + 1)$. To get a lower bound on the volume of P, we note that for all $x \geq 0$, we have $2\lfloor x \rfloor + 1 \geq x$. Thus the volume of P is

$$\prod_i (2N_i + 1) \geq (\rho/2\pi d)^d \prod_i \lambda_i^{-1} \geq (\rho/2\pi d^{3/2})^d N \geq (\rho/d)^{3d} N,$$

say.

It remains to check that $P \subset B'$ and that P is proper (whence its volume will equal its size). For the first, suppose that $x = n_1 w_1 + \cdots + n_d w_d \in P$. Let $\gamma_j \in \Gamma$. Then

$$\gamma_j x = n_1(\gamma_j w_1) + \cdots + n_d(\gamma_j w_d).$$

We know that for all $1 \leq i \leq d$, $|v_i| = \lambda_i$, and so there is some integer vector $\mathbf{a} \in \mathbb{Z}^d$ such that

$$|(\gamma_1 w_j, \ldots, \gamma_d w_j) - \mathbf{a}N| = N\,|v_j| = \lambda_j N.$$

In particular, since $|(x_1, \ldots, x_d)| \geq |x_i|$, for any $1 \leq i, j \leq d$, the integer $\gamma_i w_j$ is congruent modulo N to some m_{ij} such that $|m_{ij}| \leq \lambda_j N$. It follows that $\gamma_i x$ is congruent to some m_i such that

$$|m_i| = |n_1 m_{i1} + \cdots + n_d m_{id}| \leq \sum_{j=1}^d N_j\,|m_{ij}| \leq \sum_{j=1}^d (\rho/2\pi d \lambda_j)\lambda_j N \leq \tfrac{\rho}{2\pi} N,$$

and hence $x \in B'$ as required.

Finally, we need to check that P is proper. If not, there must exist some n_i with $|n_i| \leq 2N_i$, not all zero, such that

$$x = n_1 w_1 + \cdots + n_d w_d \equiv 0 \pmod{N}.$$

It follows that, if $v = n_1v_1 + \cdots + n_dv_d$,

$$v \in \frac{n_1w_1 + \cdots + n_dw_d}{N}(\gamma_1, \ldots, \gamma_d) + \mathbb{Z}^d \in \mathbb{Z}^d.$$

On the other hand, we have

$$|v| \leq \sum_i |n_i| |v_i| \leq 2 \sum_i N_i\lambda_i \leq \rho/\pi < 1,$$

and so we have $v = 0$. This means that $n_1v_1 + \cdots + n_dv_d = 0$, which contradicts the linear independence of the v_i. Thus P is proper, and the proof is complete. □

We can now finish the proof of the dense Bogolyubov-Ruzsa lemma, and thus the proof of the Freiman-Ruzsa-Sanders inverse theorem.

Proof of Lemma 34. Let $A \subset \mathbb{Z}/N\mathbb{Z}$ with $|A| \geq N/K$. By Lemma 37 there is a Bohr set B of rank $d \ll (\log K)^{O(1)}$ and width $\rho \gg K^{-O(1)}$ such that $B \subset 4A - 4A$. By Lemma 41 B contains a proper GAP of rank $\ll (\log K)^{O(1)}$ and cardinality

$$\geq (\rho/d)^{O(d)}N \gg \exp(-O((\log K)^{O(1)})N$$

as required. □

We have now finished the proof of the Freiman-Ruzsa-Sanders inverse result. Let's summarise the route. We began with some $A \subset \mathbb{Z}$ such that $|A + A| \leq K |A|$.

(1) We then applied the Ruzsa modelling lemma to find some $A' \subset \mathbb{Z}/N\mathbb{Z}$ (where $N \ll K^{O(1)} |A|$ is prime) such that $|A'| \geq N/K^{O(1)}$ and $4A' - 4A'$ is 2-isomorphic to a subset of $4A - 4A$.

(2) We then applied almost-periodicity to find some X such that

$$|X| \gg \exp(-O((\log K)^{O(1)}))N$$

and kX is contained 'popularly' inside $2A' - 2A'$, where $k \approx \log K$.

(3) We then used a Fourier argument to show that the Bohr set which with frequency set $\Delta = \{\gamma : |\widehat{1_X}(\gamma)| \geq \frac{1}{2} |X|\}$ and width $\gg K^{-O(1)}$ must be inside $2A' - 2A' + kX \subset 4A' - 4A'$.

(4) Chang's dimension bound tells us this Bohr set has rank $\ll (\log |X| /N)^{O(1)} \ll (\log K)^{O(1)}$, so we have found a Bohr set with rank $\ll (\log K)^{O(1)}$ and width $\gg K^{-O(1)}$ inside $4A' - 4A'$.

(5) Finally, the geometry of numbers allows us to find inside this Bohr set (and hence inside $4A' - 4A'$) a GAP with rank $\ll (\log K)^{O(1)}$ and size $\gg \exp(-(\log K)^{O(1)}))N \gg \exp(-(\log K)^{O(1)})) |A|$.

(6) This is inside $4A' - 4A'$, which is 2-isomorphic to a subset of $4A - 4A$, and hence taking the image of the progression under this 2-isomorphism finds a GAP P of the same rank and size inside $4A - 4A$.

(7) Finally, elementary sumset inequalities imply that $|P + A| \leq \exp(O(\log K)^{O(1)}) |P|$, and so there are many A is $O(K(\log K)^{O(1)}$ span-covered by $P - P$. Taking $P - P$ together with this span yields a GAP Q of rank $O(K(\log K)^{O(1)})$ and size $\leq \exp(K(\log K)^{O(1)}) |A|$ such that $A \subset Q$, which is the inverse sumset theorem.

If instead we apply Ruzsa's covering lemma in step 7, we can find a progression which efficiently covers A. By the pigeonhole principle at least one translate of this progression has a large intersection with A. More precisely, we have the following.

Theorem 20 (Freiman-Ruzsa-Sanders inverse theorem, Version 2)**.** *If $A \subset \mathbb{Z}$ has $|A+A| \leq K|A|$ then there is a proper GAP P of rank $\ll (\log K)^{O(1)}$ and size $|P| \ll K^{O(1)}|A|$ such that*

$$|A \cap P| \gg \exp(-O((\log K)^{O(1)}))\,|A|\,.$$

Proof. By Lemma 30 there is a proper GAP P inside $4A-4A$ of rank $O((\log K)^{O(1)})$ and size $\gg \exp(-O((\log K)^{O(1)}))|A|$.

If

$$P = \{a + n_1v_1 + \cdots + n_dv_d : 0 \leq n_i < N_i\}$$

then let P' be the same progression with N_i replaced by $\lfloor N_i/2 \rfloor$. It is easy to check that P being proper guarantees that $P' - P'$ is also proper, and moreover $|P'| \geq 2^{-d}|P|$.

Note that by the Plünnecke inequality $|P'| \leq |4A - 4A| \leq K^8|A|$. Furthermore,

$$|A + P'| \leq |5A - 4A| \leq K^9|A| \leq \exp((\log K)^{O(1)})\,|P'|\,.$$

It follows that A is $\exp((\log K)^{O(1)})$-covered by $P' - P'$, which is a proper GAP of the required rank and size.

That is, if $Q = P' - P'$ there is some X of size $X \leq \exp((\log K)^{O(1)})$ such that $A \subset Q + X$. By the pigeonhole principle there exists some $x \in X$ such that

$$|A \cap (Q+x)| \geq \frac{1}{|X|}|A| \geq \exp(-(\log K)^{O(1)})\,|A|$$

as required. □

It is tempting to conjecture that the $O(1)$ in the exponent of $\log K$ to be 1 – but this turns out to be false in this formulation! This was only recently shown by Lovett and Regev in 2017. The 'correct' conjecture is probably to allow for an expanded notion of GAP where we replace the 'cube constraint' $0 \leq n_i < N_i$ by a more flexible $(n_1, \ldots, n_d) \in B$ for some convex body B.

We conclude this chapter with a sample application of the inverse theorem, yet another demonstration of how various notions of 'additively structured' are related.

Theorem 21. *Let $\delta > 0$ and $k \geq 1$. If $|A|$ is sufficiently large (depending on δ and k) and A contains at least $\delta|A|^2$ many three-term arithmetic progressions then A contains a (non-trivial) k-term arithmetic progression.*

In the proof we will require another deep result we have already mentioned, but sadly have not had the time to prove!

Theorem 22 (Szemerédi)**.** *For any $\delta > 0$ and $k \geq 1$ if $N \geq N_0(\delta, k)$ is sufficiently large and $A \subset \{1, \ldots, N\}$ has size $|A| \geq \delta N$ then A contains a (non-trivial) k-term arithmetic progression.*

Proof. We begin with the Cauchy-Schwarz inequality: the number of 3APs inside A can be written as $\langle 1_A * 1_A, 1_{2\cdot A}\rangle$, and so

$$\delta|A|^2 \leq \langle 1_A * 1_A, 1_{2\cdot A}\rangle \leq \|1_A * 1_A\|_2\,|A|^{1/2} = E(A)^{1/2}\,|A|^{1/2}\,.$$

It follows that $E(A) \gg_\delta |A|^3$. By the Balog-Szemerédi-Gowers lemma there exists $A' \subset A$ such that $|A'| \gg_\delta |A|$ and $|A' + A'| \ll_\delta |A'|$.

We can now apply the Freiman-Ruzsa inverse theorem in the second form. This produces some proper arithmetic progression P of rank $d \ll_\delta 1$ such that

$$|A| \gg_\delta |P| \geq |A \cap P| \gg_\delta |A|\,.$$

Suppose that $P = P_1 + \cdots + P_d$, where each P_i is an arithmetic progression of rank 1. By the pigeonhole principle, at least one P_i has size $\gg_\delta |A|^{1/d}$. It follows, by considering the different translates of P_i, that we can partition P into arithmetic progressions of length $\gg |A|^{1/d}$.

Since we know that $|A \cap P| \gg_\delta |P|$, by averaging the above decomposition there exists some progression Q of rank 1 such that $|Q| \gg |A|^{1/d}$ and $|A \cap Q| \gg_\delta |Q|$.

But a progression of length $|Q|$ is nothing more than a translated and dilated copy of $\{1, \ldots, |Q|\}$, and both translation and dilation preserve arithmetic progressions. It follows from Szemereédi's theorem that, provided $|Q|$ is sufficiently large depending on ϵ and k, whenever $|A \cap Q| \geq \epsilon |Q|$, we have that $A \cap Q$ contains a k-term arithmetic progression.

Now we're done, we just need to check quantifiers. Here $\epsilon \gg_\delta$, so for fixed δ there is some fixed lower bound for ϵ, and hence some fixed threshold for how large $|Q|$ needs to be. Since $|Q| \geq c_1 |A|^{c_2}$ for some constants c_1, c_2 that depend only on δ, we can ensure that $|Q|$ is as large as need provided $|A|$ is large enough. □

References

[1] A. Balog and E. Szemerédi, A statistical theorem of set addition, *Combinatorica* 14 (1994), 263-268.

[2] A. L. Cauchy, Recherches sur les nombres, *J. École Polytech.* 9 (1813), 99-116.

[3] H. Davenport, On the addition of residue classes, *J. London Math. Soc.* 10 (1935), 30-32.

[4] T. Gowers, A new proof of Szemerédi's theorem, *GAFA* 11 (2001), 465-588.

[5] B. Green, "On Triples in Arithmetic Progression", available at `http://people.maths.ox.ac.uk/greenbj/papers/bourgain-roth.pdf`

[6] G. Petridis, New proofs of Plünnecke-type estimates for product sets in groups, *Combinatorica* 32 (2012), 721-733.

[7] H. Plünnecke, *Eigenschaften und Abschätzungen von Wirkingsfunktionen*, BMwF-GMD-22 Gesellschaft für Mathematik und Datenverarbeitung, Bonn 1969.

[8] I. Ruzsa, Sums of finite sets, in *Number Theory: New York Seminar*, D. V. Chudnovsky, G. V. Chudnovsky and M. B. Nathanson (eds), Springer-Verlag (1996), 281-293.

[9] T. Schoen, New bounds in Balog-Szemerédi-Gowers theorem, *Combinatorica* 35 (2015), no. 6: 695–701.

[10] T. Tao and V. Vu, *Additive Combinatorics*, Cambridge University Press 2006.