LINEAR METHODS IN ADDITIVE COMBINATORICS

THOMAS F. BLOOM

Preliminaries

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

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

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

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

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

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

It's convenient to define the difference convolution by

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

We also define the inner product by

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

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

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

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

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

We define the L^p norms for $1 \leq p < \infty$ as

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

and

$$\|f\|_\infty = \sup_{x\in G} |f(x)|.$$

The translation operator τ is defined as

$$\tau_t f(x) = f(x-t).$$

Note that L^p norms are invariant under this operator, in that

$$\|\tau_t f\|_p = \|f\|_p.$$

I will write $\mathcal{L}(\alpha)$ to mean $\log(4/\alpha)$ (the 4 is just there to ensure that $4/\alpha \geq 4$ for all $\alpha \in (0,1]$, so $\mathcal{L}(\alpha) > 1$.)

CHAPTER 1

Fourier analysis

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

In particular, $(\gamma + \lambda)(x)$ does *not* mean the function defined by $\gamma(x) + \lambda(x)$, but rather the multiplication $\gamma(x)\lambda(x)$. (This makes sense once you realise the values of $\gamma(x)$ are restricted to the unit circle in $\mathbb{C}$, on which multiplication is canonically identical to addition on $\mathbb{R}/\mathbb{Z}$.)

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

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

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

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

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

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

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

Thus the natural inner product for functions on G is

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

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

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

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

The fundamental fact underlying all of Fourier analysis is that the characters are an orthogonal basis for the set of all functions $f : G \to \mathbb{C}$; that is, we have the following orthogonality relationships.

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

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

Lemma 2. (1) *If $\gamma \in \widehat{G}$ then*

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

(2) *If $x \in G$ then*

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

(3) *For any $f : G \to \mathbb{C}$ and any $x \in G$,*

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

Proof. We will prove these in order. For the first, the case $\gamma \equiv 1$ is trivial. Suppose that $\gamma \neq 1$, then there exists some $y \in G$ such that $\gamma(y) \neq 1$. If $S = \sum_x \gamma(x)$ then

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

and hence $S = 0$.

For the second, again it is trivial when $x = 0$, so suppose $x \neq 0$. We can construct an explicit character $\gamma \in \widehat{G}$ such that $\gamma(x) \neq 1$: let $H = \langle x\rangle$ and $|H| = n$, say. If $Hy_1, \dots, Hy_m$ are the cosets of H in G then any $z \in G$ can be written uniquely as $x^h y_i$ for some $0 \leq h < n$ and $1 \leq i \leq m$. It is easy to check that

$$\lambda : x^h y_i \mapsto e(h/n)$$

defines a character on G, and clearly $\lambda(x) = e(1/n) \neq 1$.

We can now proceed as above: if $S = \mathbb{E}_\gamma\,\gamma(x)$ then

$$\lambda(x)S = \mathop{\mathbb{E}}_{\gamma}(\lambda+\gamma)(x) = \mathop{\mathbb{E}}_{\mu}\mu(x) = S,$$

hence $S = 0$.

We can now prove the inversion formula; fix $f : G \to \mathbb{C}$ and $x \in G$. We have

$$\mathop{\mathbb{E}}_{\gamma}\widehat{f}(\gamma)\gamma(x) = \sum_y f(y)\mathop{\mathbb{E}}_{\gamma}\gamma(x-y) = f(x)$$

by the second orthogonality relationship.

□

Lemma 3 (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 analytic 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 4 (Diagonalising convolution)**.** *For any $f, g : G \to \mathbb{C}$,*

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

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:

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

□

We will now demonstrate how Fourier analysis can be used to study linear problems by two classical results in additive combinatorics: the first bounding the size of sets without three-term arithmetic progressions, the second finding subspaces inside sum sets.

The theory we have discussed thus far is valid in any finite abelian group (and indeed beyond). For simplicity's sake, we will now specialise to study just $\mathbb{F}_p^n$. These are much simpler than arbitrary finite abelian groups because of the abundance of large structured subsets – namely, subspaces. In this way $\mathbb{F}_p^n$ serves as a very useful 'model case' for arguments in Fourier analysis.

Note however that for applications to number theory (and in particular sets of integers) one must work over groups like $\mathbb{Z}/N\mathbb{Z}$. While the fundamentals of Fourier analysis as above remain the same, the details of applications are much more delicate, largely because the lack of subspaces mean one must work with sets that are only 'approximately' structured (namely, Bohr sets).

1. Application: Meshulam's theorem

Theorem 1 (Meshulam)**.** *If p is an odd prime and $A \subseteq \mathbb{F}_p^n$ has no non-trivial three-term arithmetic progressions then*

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

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

Proof. It suffices to show that if $A \subseteq \mathbb{F}_p^n$ of size αp^n has no non-trivial three-term arithmetic progressions then either $|A| \ll p^{n/2}$ or there is a subspace $V \leq \mathbb{F}_p^n$ of codimension 1 and some x such that

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

for some constant $c > 0$. (We leave as an exercise to show that this implies the claimed bound.)

The key observation is that the number of three-term progressions (including the trivial ones!) in A is counted by

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

By Parseval's identity this is equal to

$$\langle \widehat{1_A}^2, \widehat{1_{2\cdot A}}\rangle = \mathop{\mathbb{E}}_{\gamma} \widehat{1_A}(\gamma)^2 \widehat{1_A}(-2\gamma).$$

(Here we have written -2γ for the character $x \mapsto \gamma(-2x)$. Note that if p is odd then if $\gamma \not\equiv 1$ then $-2\gamma \not\equiv 1$.)

Since there are only trivial three-term progressions in A this count is exactly equal to $|A|$. The contribution from the trivial character is exactly $|A|^3/p^n$. In particular either $|A| \leq p^{n/2}/2$ or

$$\alpha |A|^2 \ll \mathop{\mathbb{E}}_{\gamma\neq 0} |\widehat{1_A}(\gamma)^2 \widehat{1_A}(-2\gamma)| \leq \sup_{\gamma\neq 0} |\widehat{1_A}(\gamma)| \mathop{\mathbb{E}}_{\gamma} |\widehat{1_A}(\gamma)|^2.$$

By Parseval's identity

$$\mathop{\mathbb{E}}_{\gamma} |\widehat{1_A}(\gamma)|^2 = \sum_x 1_A(x)^2 = |A|,$$

and hence there is some non-trivial character γ such that

$$|\widehat{1_A}(\gamma)| \gg \alpha |A|.$$

(Compare this to the trivial upper bound of $|A|$.)

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

$$\langle 1_A - \alpha, c\gamma + 1\rangle = \overline{c}\widehat{1_A}(\gamma) \gg \alpha |A|.$$

(In particular, this inner product is a non-negative real number.) Now let $V \leq \mathbb{F}_p^n$ be the subspace on which $\gamma \equiv 1$ – this is imposing a single linear constraint, so V has codimension 1.

Let $w_1, \dots, w_p$ be coset representatives of V. Dividing the sum into cosets of V we have

$$\langle 1_A - \alpha, c\gamma + 1\rangle = \sum_{i=1}^{p} (c\gamma(w_i) + 1)(|A \cap (V + w_i)| - \alpha |V|).$$

Since the left-hand side is a non-negative real value we deduce

$$\sum_i \mathrm{Re}(c\gamma(w_i)+1)(|A\cap(V+w_i)|-\alpha|V|)\geq c\alpha|A|$$

where $c>0$ is some absolute positive constant. Since $0\leq \Re(c\gamma(w_i)+1)\leq 2$ we deduce that there exists some w_i such that

$$|A\cap(V+w_i)|-\alpha|V|\geq \frac{c}{2p}\alpha|A|\gg \alpha^2|V|,$$

as required. □

2. Application: Bogolyubov-Ruzsa

Theorem 2 (Bogolyubov)**.** *If $A\subseteq \mathbb{F}_p^n$ has density $\alpha=|A|/p^n$ then $A+A-A-A$ contains a subspace of codimension $\leq 2\alpha^{-2}$.*

There are much better bounds available here, which we will discuss in the next lecture – for now the important thing is the *qualitative* fact that when A has density $\gg 1$ the sumset $A+A-A-A$ contains a subspace of density $\gg 1$. (This is not true for A itself, or even for $A+A$.)

Proof. Let $\eta=(\alpha/2)^{1/2}$ and

$$\Delta=\{\gamma:|\widehat{1_A}(\gamma)|\geq\eta|A|\}.$$

Note that by Parseval's theorem we can bound the size of Δ from above:

$$\eta^2|A|^2|\Delta|p^{-n}\leq \mathop{\mathbb{E}}_{\gamma}|\widehat{1_A}(\gamma)|^2=|A|,$$

whence $|\Delta|\leq\eta^{-2}\alpha^{-1}$. Let V be the subspace defined by $\gamma(x)\equiv 1$ for all $\gamma\in\Delta$. These are at most $|\Delta|$ linear constraints and so V has codimension $\leq\eta^{-2}\alpha^{-1}$. We claim that $V\subseteq A+A-A-A$.

Indeed, let $v\in V$. It suffices to show that $1_A*1_A*1_{-A}*1_{-A}(v)>0$. By the inverse Fourier transform and diagonalisation

$$1_A*1_A*1_{-A}*1_{-A}(v)=\mathop{\mathbb{E}}_{\gamma}|\widehat{1_A}(\gamma)|^4\gamma(v)\geq\alpha|A|^3-\mathop{\mathbb{E}}_{\gamma}1_{\gamma\notin\Delta}|\widehat{1_A}(\gamma)|^4,$$

the first summand coming from the trivial character $\gamma\equiv 1$. The second summand is, by definition of Δ, at most

$$\eta^2|A|^2\mathop{\mathbb{E}}_{\gamma}|\widehat{1_A}(\gamma)|^2=\eta^2|A|^3.$$

Since $\eta=(\alpha/2)^{1/2}$ we have shown that

$$1_A*1_A*1_{-A}*1_{-A}(v)\geq\tfrac{1}{2}\alpha|A|^3>0$$

as required. □

CHAPTER 2

Almost periodicity

Informally we say that t is an 'almost period' of a function f if $\tau_t f \approx f$. It is a useful heuristic to view this a kind of 'continuity' statement: saying that a function f is continuous at a point x is essentially saying that $\tau_t f(x) \approx f(x)$ for all t in a small ball around the origin. The general goal of almost-periodicity is to find similar continuity statements for functions of combinatorial interest (for example $f = 1_A * 1_B$), i.e. that there exists some large 'structured' set of almost-periods.

The usual way to make this precise is to fix some L^p norm and ask for $\|\tau_t f - f\|_p$ to be small. Since finding a large set of almost periods is finding a large set under which f is approximately invariant, there are obvious applications to additive problems.

Many applications of Fourier analysis to linear problems can be viewed as going via some kind of almost-periodicity result. For example, using essentially the same argument as for Bogolyubov's lemma, we can prove the following.

Theorem 3. *If $A, B \subseteq \mathbb{F}_p^n$ with densities α, β and $C \subseteq \mathbb{F}_p^n$ then there is a subspace $V \leq \mathbb{F}_p^n$ of codimension*

$$\ll \epsilon^{-2}\alpha^{-1}\beta^{-1}$$

such that

$$\|\tau_t(1_A * 1_B * 1_C) - 1_A * 1_B * 1_C\|_\infty \leq \epsilon \left|A\right| \left|B\right|.$$

Proof. For any $x \in \mathbb{F}_p^n$ we have

$$1_A * 1_B * 1_C(x) = \mathop{\mathbb{E}}_{\gamma} \widehat{1_A}(\gamma)\widehat{1_B}(\gamma)\widehat{1_C}(\gamma)\gamma(x).$$

Therefore

$$|1_A * 1_B * 1_C(x+t) - 1_A * 1_B * 1_C(x)| \leq \mathop{\mathbb{E}}_{\gamma} |\widehat{1_A}(\gamma)\widehat{1_B}(\gamma)\widehat{1_C}(\gamma)||\gamma(t) - 1|.$$

Let V be the subspace which annihilates

$$\Delta = \{\gamma : |\widehat{1_A}(\gamma)| \geq \eta \left|A\right|\}.$$

Then the right-hand side is, for $t \in V$,

$$\leq \eta \left|A\right| \left|B\right|^{1/2} \left|C\right|^{1/2} \leq \epsilon \left|A\right| \left|B\right|$$

with $\eta \leq \epsilon(\left|B\right| / \left|C\right|)^{1/2}$, so can take $\eta = \epsilon\beta^{1/2}$, say. □

Using some more tools from Fourier analysis this can be improved a little; for example, Chang's lemma (see below) improves the codimension bound to

$$\ll \epsilon^{-2}\mathcal{L}(\alpha)\beta^{-1}.$$

Importantly, however, this remains 'polynomial' in the densities.

A very important advance in linear methods was the use of probabilistic methods to study 'almost-periodicity' by Croot and Sisask, which allowed for much more effective bounds, polynomial in the *logarithm* of the densities.

Theorem 4. *If $A, B \subseteq \mathbb{F}_p^n$ with densities α, β and $C \subseteq \mathbb{F}_p^n$ then there is a subspace $V \leq \mathbb{F}_p^n$ of codimension*

$$\ll \epsilon^{-2}\mathcal{L}(\epsilon\alpha)^2\mathcal{L}(\alpha)\mathcal{L}(\beta)$$

such that

$$\|\tau_t(1_A * 1_B * 1_C) - 1_A * 1_B * 1_C\|_\infty \leq \epsilon |A| |B|.$$

We will discuss the proof of this later. For now we will give two sample applications.

3. Application: Bogolyubov-Ruzsa

Theorem 5 (Sanders)**.** *If $A \subseteq \mathbb{F}_p^n$ has density α then $A + A - A - A$ contains a subspace of codimension*

$$\ll \mathcal{L}(\alpha)^4.$$

(For comparison, recall that the earlier Fourier argument gave $\ll \alpha^{-2}$, and if we additionally use Chang's lemma (see below) we get $\ll \alpha^{-1}\mathcal{L}(\alpha)$. In the exercises I suggest a different Fourier analytic argument that gives $\ll \alpha^{-1/2}$. There is no known route to a bound polynomial in $\mathcal{L}(\alpha)$ without using almost-periodicity.)

Proof. We apply Theorem 4 with $A = B$ and $C = -(A + A)$ and $\epsilon = 1/2$, say. This produces some V with the claimed codimension such that, for all $t \in V$ and $x \in \mathbb{F}_p^n$,

$$|1_A * 1_A * 1_{-(A+A)}(x + t) - 1_A * 1_A * 1_{-(A+A)}(x)| \leq \tfrac{1}{2} |A|^2.$$

We apply this with $x = 0$, and note that

$$1_A * 1_A * 1_{-(A+A)}(0) = \sum_{a,b\in A} \sum_{c\in A+A} 1_{a+b=c} = |A|^2.$$

Therefore $1_A * 1_A * 1_{-(A+A)}(t) \geq |A|^2/2$ and so cannot be 0, hence for all $t \in V$ we have $t \in A + A - A - A$ as claimed. □

4. Application: Translation invariant equations

Previously we saw how Fourier analysis can be used (with the density increment strategy) to prove a bound like $\alpha \ll 1/n$ for the density of $A \subseteq \mathbb{F}_p^n$ with a non-trivial solution to $x + y = 2z$.

Almost-periodicity doesn't apply directly to this problem, because there aren't enough variables. Schoen and Sisask observed that with just a single additional variable however we can say very strong things, replacing the polynomial bound $\alpha \ll 1/n$ with a quasi-exponential bound.

Theorem 6 (Schoen-Sisask)**.** *If $A \subseteq \mathbb{F}_p^n$ has density α and contains no non-trivial solutions to $x + y + z = 3w$ then*

$$\alpha \ll \exp(-n^{1/5}).$$

Proof. As before, using the density increment strategy it suffices to show that there exists some subspace $V \leq \mathbb{F}_p^n$ of codimension

$$\ll \mathcal{L}(\alpha)^4$$

such that $|(A-x)\cap V| \geq \frac{3}{2}\alpha |V|$ for some x.

We apply Theorem 4 with $B = -3A$ and $C = A+A$ and $\epsilon = 1/8$. As in the previous proof, this produces some V of the right codimension such that, for any $t \in V$,

$$1_A * 1_{-3A} * 1_{A+A}(t) \leq 1_A * 1_{-3A} * 1_{A+A}(0) + \tfrac{1}{8}|A|^2.$$

Now $1_A * 1_{-3A} * 1_{A+A}(0)$ is counting the number of solutions to $x+y=3w$ where $x,w \in A$ and $y \in A+A$. There are only trivial solutions to this, and hence $1_A * 1_{-3A} * 1_{A+A}(0) = |A|$, and hence for any $t \in V$

$$1_A * 1_{-3A} * 1_{A+A}(t) \leq \tfrac{1}{4}|A|^2.$$

But we want a lower bound for an increment! For this we note that $1_A * 1_{-3A} * 1 \equiv |A|^2$ and hence

$$1_A * 1_{-3A} * 1_{(A+A)^c}(t) \geq \tfrac{3}{4}|A|^2.$$

Averaging over all $t \in V$ and rearranging this implies

$$\sum_{x \in -3A} \sum_{y \in (A+A)^c} |(A-x-y)\cap V| \geq \tfrac{3}{4}|A|^2 |V|.$$

In particular there must exist some $x \in \mathbb{F}_p^n$ such that

$$|(A-x)\cap V| \geq \frac{p^n}{p^n - |A+A|}\tfrac{3}{4}\alpha |V|.$$

The first factor is $\geq 1/2$ if $|A+A| \geq \frac{1}{2}p^n$, and we are done.

But what if $A+A$ is small? Then almost-periodicity saves the day again, but in a slightly different way – applying it as in our proof of Bogolyubov-Ruzsa we can find some V of the right codimension such that for all $t \in V$

$$1_A * 1_A * 1_{-(A+A)}(t) \geq \tfrac{3}{4}|A|^2.$$

Averaging as above, there exists some $x \in \mathbb{F}_p^n$ such that

$$|(A-x)\cap V| \geq \frac{p^n}{|A+A|}\tfrac{3}{4}\alpha |V|.$$

So if $|A+A| \leq \frac{1}{2}p^n$ we are done. □

5. Proving Almost-Periodicity

The quantitative strength of almost-periodicity comes from the fact that, at heart, it is a 'physical' argument rather than a Fourier one (although Fourier analysis will play a role towards the end). The key result is the following; this is the 'primal form' of almost-periodicity from which everything else will follow.

Theorem 7. *If $A, B \subseteq G$ with densities α, β respectively then there exists some set T of size*

$$|T| \geq \exp(-O(p\epsilon^{-2}\mathcal{L}(\alpha)))|G|$$

such that, for all $t \in T - T$,

$$\|\tau_t(1_A * 1_B) - 1_A * 1_B\|_p \leq \epsilon |A| |B|^{1/p}.$$

Proof. For convenience let $\mu_A = |A|^{-1} 1_A$. Let $k \geq 1$ be some integer to be chosen later. Let $a_1, \dots, a_k \in A$ be chosen independently at random and write

$$\mu_{\mathbf{a}} * 1_B(x) = \frac{1}{k} \sum_i 1_B(x - a_i).$$

The first step is to show that, with a suitable choice of k,

$$\|\mu_{\mathbf{a}} * 1_B - \mu_A * 1_B\|_p \leq \frac{\epsilon}{2} |B|^{1/p}$$

with probability at least 0.99 (say). Indeed, taking pth powers and by Markov's inequality, it suffices to show that

$$\sum_x \mathbb{E} |\mu_{\mathbf{a}} * 1_B(x) - \mu_A * 1_B(x)|^p \ll (p/k)^{p/2} |B|,$$

and then choose $k \asymp p\epsilon^{-2}$. To see why this makes sense note that

$$\mathbb{E}\, \mu_{\mathbf{a}} * 1_B(x) = \frac{1}{k} \mathbb{E} \sum_i 1_B(x - a_i) = \mathbb{E}\, 1_B(x - a) = \frac{1}{|A|} \sum_{a \in A} 1_B(x - a) = \mu_A * 1_B(x).$$

So we want to bound the pth central moment of the random variable $\mu_{\mathbf{a}} * 1_B(x)$. Indeed, let $X_i = 1_B(x - a_i) - \mu_A * 1_B(x)$, which is a random variable with $\mathbb{E}\, X_i = 0$. We claim that

$$\mathbb{E} \left| \frac{1}{k} \sum_i X_i \right|^p \leq (Cp/k)^{p/2} \frac{1}{k} \sum_i \mathbb{E} |X_i|^p.$$

Once we have this, we're done, since

$$\mathbb{E} |X_i|^p \ll \mathbb{E}\, 1_B(x - a_i) + \mu_A * 1_B(x)^p,$$

and then summing over $x \in G$ gives

$$\sum_x \mathbb{E} |X_i|^p \ll |B| + \|\mu_A * 1_B\|_p^p \ll |B|.$$

So we just need to show that the central p moment of the average of independent random variables is controlled by the average of the individual moments. As you might expect, this is true, and is a standard result in probability theory, known as the Marcinkiewicz-Zygmund inequality. (Although often it is not stated/proved with the dependence on p and k made explicit in that way.)

We will not prove this probability fact here, but you might like to note that when $p = 2$ this is nothing more than the statement that the variance of the sum of independent random variables is at most the sum of the variances.

So we've shown that for 99% of $\mathbf{a} \in A^k$ we have

$$\|\mu_{\mathbf{a}} * 1_B - \mu_A * 1_B\|_p \leq \frac{\epsilon}{2} |B|^{1/p}$$

Call this set of $\mathbf{a} \in A^k$ the set of 'good' tuples, say L. Now we can divide all of G^k into equivalence classes by saying $\mathbf{a} \sim \mathbf{b}$ are equivalent if $\mathbf{a} - \mathbf{b} = (t, \dots, t)$ for some $t \in G$. The number of equivalence classes is $|G|^{k-1}$.

Therefore there must be some equivalence class which contains at least $0.99\alpha^k |G|$ many vectors from L – that is, there is some $\mathbf{a} \in G^k$ and $\geq 0.99\alpha^k |G|$ many $t \in G$ such that $\mathbf{a} - (t, \dots, t) \in L$.

We claim that this set of T suffices. Indeed, suppose $t_1, t_2 \in T$. Then by the triangle inequality

$$\begin{aligned}\|\tau_{t_1-t_2}(1_A * 1_B) - 1_A * 1_B\|_p &= \|\tau_{t_1}(1_A * 1_B) - \tau_{t_2}(1_A * 1_B)\|_p \\ &\leq \|\tau_{t_1}(1_A * 1_B) - \mu_{\mathbf{a}} * 1_B\|_p + \|\tau_{t_2}(1_A * 1_B) - \mu_{\mathbf{a}} * 1_B\|_p \\ &= \|1_A * 1_B - \tau_{-t_1}(\mu_{\mathbf{a}} * 1_B)\|_p + \|1_A * 1_B - \tau_{-t_2}(\mu_{\mathbf{a}} * 1_B)\|_p \\ &\leq \epsilon \left|A\right| \left|B\right|^{1/p}.\end{aligned}$$

□

To get to the form stated in Theorem 4 requires two further steps: we must pass to an L^∞ version rather than an L^p version, and then need to pass to a subspace rather than a set.

Corollary 1. *If $A, B \subseteq G$ with densities α, β respectively and $C \subseteq G$ then there exists some set T of size*

$$|T| \geq \exp(-O(\epsilon^{-2}\mathcal{L}(\alpha)\mathcal{L}(\beta))\left|G\right|$$

such that, for all $t \in T$,

$$\|\tau_t(1_A * 1_B * 1_C) - 1_A * 1_B * 1_C\|_\infty \leq \epsilon \left|A\right| \left|B\right|.$$

Proof. Let p be chosen later. By Theorem 7 there exists some set T of size

$$|T| \geq \exp(-O(p\epsilon^{-2}\mathcal{L}(\alpha))\left|G\right|$$

such that for all $t \in T$

$$\|\tau_t(1_A * 1_C) - 1_A * 1_C\|_p \leq \epsilon \left|A\right| \left|C\right|^{1/p}.$$

Let $x \in G$ be arbitrary, and write

$$|1_A * 1_B * 1_C(x+t) - 1_A * 1_B * 1_C(x)| = |\langle \tau_t(1_A * 1_C) - 1_A * 1_C, 1_{x-B}\rangle|.$$

By Hölder's inequality the right-hand side is at most

$$\epsilon \left|A\right| \left|C\right|^{1/p} \left|B\right|^{1-1/p} \leq \epsilon \left|A\right| \left|B\right| \beta^{-1/p}.$$

The claim follows choosing $p \asymp \mathcal{L}(\beta)$. □

Finally, to deduce Theorem 4 we need to upgrade this set T of almost-periods into a subspace. For this we will require a new tool: Chang's lemma.

Lemma 5 (Chang)**.** *If $A \subseteq G$ with density α and*

$$\Delta_\eta(A) = \{\gamma : |\widehat{1_A}(\gamma)| \geq \eta \left|A\right|\}$$

then

$$\dim\langle \Delta_\eta(A)\rangle \ll \eta^{-2}\mathcal{L}(\alpha).$$

For comparison, recall that by Parseval we know that $|\Delta_\eta(A)| \leq \eta^{-2}\alpha^{-1}$, so the 'trivial' dimension bound here would be $\ll \eta^{-2}\alpha^{-1}$. The power of Chang's lemma is that the polynomial loss in α is reduced to logarithmic.

Note that a logarithmic dependency on α is natural here – suppose that $\eta = 1$ and $A = V \leq \mathbb{F}_p^n$ is a subspace. Then $\Delta_\eta(A)$ is exactly $V^\perp$ which has dimension $= \mathrm{codim}(V) = \log_p(1/\alpha)$.

Proof of Theorem 4. Let $k \geq 1$ be chosen later. We apply Corollary 1 with ϵ replaced by ϵ/k to find some T of density

$$\tau \geq \exp(-O(\epsilon^{-2}k^2\mathcal{L}(\alpha)\mathcal{L}(\beta))$$

such that

$$\|\tau_t(1_A * 1_B * 1_C) - 1_A * 1_B * 1_C\|_\infty \leq \frac{\epsilon}{k} |A| |B|.$$

By the triangle inequality, for any $x \in kT$ we have

$$\|\tau_x(1_A * 1_B * 1_C) - 1_A * 1_B * 1_C\|_\infty \leq \epsilon |A| |B|.$$

In particular, since $\mu_T^{(k)}$ (the k-fold convolution) is supported on kT,

$$\|\mu_T^{(k)} * 1_A * 1_B * 1_C - 1_A * 1_B * 1_C\|_\infty \leq \epsilon |A| |B|.$$

It therefore suffices to find some subspace V such that for all $v \in V$

$$\|\tau_v(\mu_T^{(k)} * 1_A * 1_B * 1_C) - \mu_T^{(k)} * 1_A * 1_B * 1_C\|_\infty \leq \epsilon |A| |B|.$$

Writing this out in Fourier space and using the triangle inequality, it suffices to show that

$$\mathop{\mathbb{E}}_{\gamma} |\widehat{\mu_T}(\gamma)^k| |\widehat{1_A}(\gamma)\widehat{1_B}(\gamma)\widehat{1_C}(\gamma)| |\gamma(v) - 1| \leq \epsilon |A| |B|.$$

Let V be the subspace which annihilates $\Delta_{1/2}(T)$, so this left-hand side is

$$\leq 2^{-k} \mathop{\mathbb{E}}_{\gamma} |\widehat{1_A}(\gamma)\widehat{1_B}(\gamma)\widehat{1_C}(\gamma)| \leq 2^{-k}\alpha^{-1/2} |A| |B|.$$

Choosing $k \approx \mathcal{L}(\epsilon\alpha)$ finishes the proof – we just need to check the codimension of V. This is where Chang's lemma is vital:

$$\mathrm{codim}(V) \leq \dim\langle\Delta_{1/2}(T)\rangle \ll \mathcal{L}(\tau) \ll \epsilon^{-2}k^2\mathcal{L}(\alpha)\mathcal{L}(\beta).$$

□

CHAPTER 3

Kelley-Meka bounds for three-term arithmetic progressions

In 2023 there was a great leap forward in our understanding of sets without three-term arithmetic progressions, when Kelley and Meka proved the following bound.

Theorem 8 (Kelley-Meka)**.** *If $A \subseteq \{1, \dots, N\}$ has no non-trivial three-term arithmetic progressions then*

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

for some $c > 0$.

For comparison, the previous best bound was

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

for some (very tiny) constant $c > 0$. Furthermore this new upper bound of Kelley and Meka is not too far from the best lower bound we know: in 1946 Behrend constructed a set $A \subseteq \{1, \dots, N\}$ without non-trivial three-term arithmetic progressions such that

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

for some constant $c > 0$. This construction has very recently improved by Elsholtz, Hunter, Proske, and Sauermann, although only in the value of c. It is likely that the true density of the largest set without non-trivial three-term arithmetic progressions is $\exp(-O((\log N)^{1/2}))$.

Modifying the Kelley-Meka proof slightly Bloom and Sisask improved their exponent of $1/12$ to $1/9$. It would be of great interest to lower this exponent further, particularly to $1/3$, which seems to be a natural barrier for any kind of density increment method.

We will sketch the main ideas behind the Kelley-Meka approach. As before, we will consider only the simpler setting of $\mathbb{F}_p^n$ – all the main ideas are the same in the case of $\{1, \dots, N\}$ or $\mathbb{Z}/N\mathbb{Z}$, except that one must work with 'Bohr sets' throughout, so the technical details are scarier.

In this model setting Kelley-Meka prove the following.

Theorem 9 (Kelley-Meka)**.** *If p is an odd prime and $A \subseteq \mathbb{F}_p^n$ has no non-trivial three-term arithmetic progressions then*

$$|A| \ll C^{-n^{1/9}} p^n$$

for some $C > 1$.

(Again, compare this to the bound of $\ll p^n/n$ we proved using Fourier analysis, and to the bound $C^{-n}p^n$ achieved by the polynomial method.)

A key observation of Kelley and Meka is the following: we already know (via Schoen-Sisask and almost-periodicity) how to achieve quasi-polynomial bounds of this strength for a similar problem, if instead of ruling out solutions to $x+y=2z$ we rule out solutions to a four variable equation like $x+y+z=3w$. Can we somehow convert the three-term arithmetic progression case into something more of this form, which almost-periodicity can handle? The answer is yes, and via a surprisingly simple technique they call 'sifting'. Just as with almost-periodicity itself, random sampling comes to the rescue again.

Using the density increment method, it suffices to prove that if $A \subseteq \mathbb{F}_p^n$ has no non-trivial three-term arithmetic progressions then either $\alpha \ll p^{-n/2}$ or there is a subspace $V \leq \mathbb{F}_p^n$ of codimension $\ll \mathcal{L}(\alpha)^8$ and some x such that

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

for some constant $c>0$.

Here are the steps of their argument. We fix some set $A \subseteq \mathbb{F}_p^n$ with no non-trivial three-term arithmetic progressions with density $\alpha \gg p^{-n/2}$.

(1) Prove that, for some $q \ll \mathcal{L}(\alpha)$, we have

$$\|1_A \circ 1_A\|_q \geq (1+c)\alpha\,|A|\,|G|^{1/q}$$

for some $c>0$.

(2) Prove that if $S = \{x : 1_A \circ 1_A(x) \geq (1-\epsilon)\|1_A \circ 1_A\|_q\,|G|^{-1/q}\}$ (and q is large enough depending on ϵ) then there exists some $A' \subseteq A$ such that

$$|A'| \gg \alpha^q\,|G|$$

and

$$\langle 1_{A'} \circ 1_{A'}, 1_S\rangle \geq (1-\epsilon/4)\,|A'|^2\,.$$

(3) Deduce using almost-periodicity that there is a subspace $V \leq \mathbb{F}_p^n$ of codimension

$$\ll_\epsilon \mathcal{L}(\alpha')^4 \ll \mathcal{L}(\alpha)^8$$

such that

$$\langle \mu_V * 1_{A'} \circ 1_{A'}, 1_S\rangle \geq (1-\epsilon/2)\,|A'|^2\,.$$

Combining these steps and using the definition of S, and choosing ϵ sufficiently small, we have

$$\langle \mu_V * 1_{A'} \circ 1_{A'}, 1_A \circ 1_A\rangle \geq (1+c)\alpha\,|A|\,|A'|^2$$

for some constant $c>0$. By averaging we deduce that

$$\|\mu_V * 1_A\|_\infty \geq (1+c)\alpha$$

which is what we needed.

6. From few 3APs to large L^q norm

Suppose that A has no non-trivial three-term arithmetic progressions. Then we have already seen in our proof of Meshulam's bound that either $\alpha \ll p^{-n/2}$ or

$$\mathop{\mathbb{E}}_{\gamma} 1_{\gamma\neq 1}|\widehat{1_A}(\gamma)|^3 \gg \alpha\,|A|^2\,,$$

Since $\widehat{1_A}(1) = |A|$ we can add back in the trivial character to see that, for constant $c > 0$,

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

We can write the left-hand side as

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

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

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

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

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

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

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

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

$$\sum_{x_1,\ldots,x_{2m}} 1_A\circ 1_A(x_1)\cdots 1_A\circ 1_A(x_{2m})\mathop{\mathbb{E}}_{\gamma_1,\ldots,\gamma_{2m}} 1_{\gamma_1+\cdots-\gamma_{2m}=0}\gamma_1(x_1)\cdots\gamma_p(x_{2m}).$$

By orthogonality this is

$$\mathop{\mathbb{E}}_{x\in G} 1_A\circ 1_A(x)^{2m},$$

and hence we have shown that

$$\left(\mathop{\mathbb{E}}_{x\in G} 1_A\circ 1_A(x)^{2m}\right)^{1/2m} \geq (1+c)\alpha\,|A|\,.$$

This is all the information that we will use to get a density increment. (Kelley and Meka obtained it in an alternative way that uses less Fourier analysis.)

For comparison, note that by Hölder's inequality

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

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

7. Sifting

If $A \subseteq G$ and $\mathbf{t} \in G^{q-1}$ then we write

$$A(\mathbf{t}) = A \cap (A - t_1) \cdots (A - t_{q-1}).$$

Note that

$$\sum_{\mathbf{t}} |A(\mathbf{t})| = \sum_{a \in A} \sum_{t_1, \dots, t_{q-1}} 1_A(a + t_1) \cdots 1_A(a + t_{q-1}) = |A|^q .$$

By the Cauchy-Schwarz inequality we deduce that

$$\sum_{\mathbf{t}} |A(\mathbf{t})|^2 \geq |A|^{2q} |G|^{1-q} .$$

We can do better, however – in fact the left-hand side is equal to $\|1_A \circ 1_A\|_q^q$ (and then note that this inequality follows immediately from Hölder's inequality).

This is the fundamental lemma.

Lemma 6. *For any $x \in G$ and integer $q \geq 1$*

$$1_A \circ 1_A(x)^q = \sum_{\mathbf{t}} 1_{A(\mathbf{t})} \circ 1_{A(\mathbf{t})}(x).$$

Proof.

$$\begin{aligned} 1_A \circ 1_A(x)^q &= 1_A \circ 1_A(x) 1_A \circ 1_A(x)^{q-1} \\ &= \sum_{a,b \in A} 1_{a-b=x} 1_A \circ 1_A(a-b)^{q-1} \\ &= \sum_{a,b \in A} 1_{a-b=x} \left(\sum_t 1_A(a+t) 1_A(b+t) \right)^{q-1} \\ &= \sum_{t_1, \dots, t_{q-1}} \sum_{a,b \in A} 1_{a-b=x} 1_A(a+t_1) \cdots 1_A(b+t_{q-1}) \\ &= \sum_{\mathbf{t}} 1_{A(\mathbf{t})} \circ 1_{A(\mathbf{t})}(x). \end{aligned}$$

□

$$\|1_A \circ 1_A\|_{q+1}^{q+1} = \sum_{\mathbf{t}} \langle 1_{A(\mathbf{t})} \circ 1_{A(\mathbf{t})}, 1_A \circ 1_A \rangle.$$

By Hölder's inequality the left-hand side is at least

$$\|1_A \circ 1_A\|_q^{q+1} |G|^{-1/q} = \|1_A \circ 1_A\|_q |G|^{-1/q} \sum_{\mathbf{t}} |A(\mathbf{t})|^2 .$$

In particular, if $A' = A(\mathbf{t})$ for a uniformly randomly chosen $\mathbf{t}$ then

$$\mathbb{E} \langle 1_{A'} \circ 1_{A'}, 1_A \circ 1_A \rangle \geq \mathbb{E} \|1_A \circ 1_A\|_q |G|^{-1/q} |A'|^2 .$$

In particular if

$$\|1_A \circ 1_A\|_q \geq (1+c)\alpha |A| |G|^{1/q}$$

then

$$\mathbb{E} \langle 1_{A'} \circ 1_{A'}, 1_A \circ 1_A \rangle \geq \mathbb{E} (1+c)\alpha |A| |A'|^2 .$$

This is the magical step – note that $\alpha |A| |A'|^2$ is the 'expected' value of the inner product, so we have a constant discrepancy. But this inner product counts solutions to a 4 variable equation! And we paid a reasonable cost to do so, only decreasing the density of two copies of A by some mild amount. So the hope is that the almost-perodicity success of Schoen and Sisask also applies here. This is true, although it is convenient to actually prove a slightly different statement.

Lemma 7. *Let $S = \{x : 1_A \circ 1_A(x) \geq (1-\epsilon)\|1_A \circ 1_A\|_q |G|^{-1/q}\}$. Then there exists some $A' \subseteq A$ such that*

$$|A'| \gg \alpha^q |G|$$

and

$$\langle 1_{A'} \circ 1_{A'}, 1_S\rangle \geq (1-\epsilon/4)|A'|^2.$$

Proof.

$$\begin{aligned}\sum_{\mathbf{t}} \langle 1_{A(\mathbf{t})} \circ 1_{A(\mathbf{t})}, 1_{S^c}\rangle &= \langle (1_A \circ 1_A)^q, 1_{S^c}\rangle \\ &\leq (1-\epsilon)^q \|1_A \circ 1_A\|_q^q \\ &= (1-\epsilon)^q \sum_{\mathbf{t}} |A(\mathbf{t})|^2.\end{aligned}$$

Furthermore if T is the set of $\mathbf{t}$ where $|A(\mathbf{t})| \geq \frac{1}{2}\alpha^q |G|$ then

$$\sum_{\mathbf{t}\notin T} |A(\mathbf{t})|^2 < \tfrac{1}{2}\alpha^q |G| |A|^q \leq \tfrac{1}{2}\sum_{\mathbf{t}} |A(\mathbf{t})|^2.$$

Therefore

$$\sum_{\mathbf{t}} \langle 1_{A(\mathbf{t})} \circ 1_{A(\mathbf{t})}, 1_{S^c}\rangle \leq 2(1-\epsilon)^q \sum_{\mathbf{t}\in T} |A(\mathbf{t})|^2.$$

In particular there must exist some $\mathbf{t} \in T$ such that

$$\langle 1_{A(\mathbf{t})} \circ 1_{A(\mathbf{t})}, 1_{S^c}\rangle \leq 2(1-\epsilon)^q |A(\mathbf{t})|^2.$$

The claim follows since for any set A'

$$\langle 1_{A'} \circ 1_{A'}, 1\rangle = |A'|^2.$$

□