Techniques in non-Abelian additive combinatorics

Professor Tim Gowers

It’s combinatorics where you look at subsets of things rather than substructures of things, and yet you can often say a lot about the structure of those subsets. The problems are purely combinatorial, but the ideas come from all sorts of areas of math. And here’s a motivating problem:

**Problem 0.1.** In a group of order \( n \), you want to find an as large as possible product-free subset of the group.

If it’s abelian, can get a constant proportion \( 2n/7 \) by decomposing into cyclic groups and taking the middle third of one of them and all of the others (the worst case is \( n = 7 \)). But if it’s not abelian, not clear how to do it.

There’s notes on Tim’s blog - there’s a column on the right and a thing called ‘Lecture notes’ - for a course similar to the current one.

1. Chapter 1: Fourier analysis on finite abelian groups

**Definition 1.1.** A character on a finite abelian group \( G \) is a homomorphism from \( G \) to \( \mathbb{C}^* \) (the multiplicative group of complex numbers).

It’s easy to prove that if \( \chi \) is a character, then \( |\chi(x)| = 1 \) for every \( x \).

**Notation 1.2.** We’ll write \( \mathcal{E}_x[f(x)] \) for \( \frac{1}{|G|} \sum_{x \in G} f(x) \)

**Lemma 1.3.** If \( \chi \) is a non-trivial character, then \( \mathcal{E}_x[\chi(x)] = 0 \).

*Proof. Since \( \chi \) is non-trivial, there exists \( y \) such that \( \chi(y) \neq 1 \). Then \( \mathcal{E}_x[\chi(x)] = \mathcal{E}_x[\chi(xy)] = \chi(y)\mathcal{E}_x[\chi(x)] \) and we’re done. □*

**Corollary 1.4.** If \( \chi_1, \chi_2 \) are characters, then

\[
\mathcal{E}_x[\chi_1(x)\overline{\chi_2(x)}] = \begin{cases} 
1, & \text{if } \chi_1 = \chi_2 \\
0, & \text{otherwise}
\end{cases}
\]

*Proof. If \( \chi_1 = \chi_2 \), then \( \chi_1(x)\overline{\chi_2(x)} = 1 \) for every \( x \), and if \( \chi_1 \neq \chi_2 \), then \( \chi_1\overline{\chi_2} \) is a nontrivial character, and we’re done by 1.3. □*

So if we define an inner product on \( \mathbb{C}^G \) by

\[
\langle f, g \rangle = \mathcal{E}_x[f(x)\overline{g(x)}]
\]

then this says that the characters form an orthonormal set.

**Lemma 1.5.** There are \( |G| \) distinct characters.
Proof (sketch). By the structure theorem for finite abelian groups, \( G \) is a product of cyclic groups. If \( G = C_{n_1} \times C_{n_2} \times \cdots \times C_{n_k} \), write a typical element of \( G \) as \((x_1, \ldots, x_k)\) where \( x_i \in \mathbb{Z}_{n_i} = \mathbb{Z}/n_i\mathbb{Z} \).

For each \((r_1, \ldots, r_k) \in \mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}\), can define a character by sending

\[
(x_1, \ldots, x_k) \mapsto \prod_{j=1}^{k} e^{2\pi i r_j x_j/n_j}
\]

Another way to say it that makes it more transparent is to write the terms as \((e^{2\pi i r_j/n_j})^{x_j}\) □

So in fact the characters form an orthonormal basis.

**Definition 1.6.** \( \widehat{G} \), the Pontryagin dual of \( G \) (or just the dual group) is the group of characters with pointwise multiplication.

This is isomorphic to \( G \) via the proof of Lemma 1.5. We use the different notation because we use a different measure on \( \widehat{G} \); the measure where each point has measure 1. Moreover, \( \widehat{G} \) and \( G \) are not naturally isomorphic.

**Definition 1.7.** The Fourier transform of a function \( f : G \to \mathbb{C} \) is the function \( \widehat{f} : \widehat{G} \to \mathbb{C} \) given by the formula

\[
\widehat{f}(\chi) = \mathbb{E}_x \left[ f(x)\overline{\chi(x)} \right] = \langle f, \chi \rangle
\]

So when you take functions on \( G \), there’s an obvious basis, but there’s another basis - the basis of characters, which is also very useful.

Also, why are we bothering with functions on \( G \) at all? We’ll soon be looking at subsets of groups, and to them we can associate their indicator functions, which the Fourier transform tells us a lot of nice things about.

Also, what does this have to do with the usual Fourier transform?

**Example 1.8.** Let \( G = \mathbb{Z}_n \). Then the characters are of the form \( \omega_r : x \mapsto \omega^{rx} \) for some \( r \in \mathbb{Z}_n \), where \( \omega = e^{2\pi i/n} \).

So writing \( \widehat{f}(r) \) for \( \widehat{f}(\omega_r) \), we get

\[
\widehat{f} = \mathbb{E}_x \left[ f(x)\omega^{-rx} \right] = \mathbb{E}_x \left[ f(x)e^{-2\pi irx/n} \right]
\]

Also, if \( G = \mathbb{F}_p^m \) (i.e. a vector space over the finite field \( \mathbb{F}_p \)), then the characters take the form

\[
\omega_{\underline{r}} : \underline{z} \mapsto \omega^{\underline{r} \cdot \underline{z}}
\]

where \( \omega = e^{2\pi i/p} \) and \( \underline{r} \cdot \underline{z} = \sum_i r_i z_i \). So then

\[
\widehat{f}(\underline{r}) = \mathbb{E}_x \left[ f(x)\omega^{-\underline{r} \cdot \underline{z}} \right]
\]

**Notation 1.9.** Let \( f : G \to \mathbb{C} \). Then

\[
\|f\|_p = \left( \mathbb{E}_x [ |f(x)|^p] \right)^{1/p}, \quad \|f\|_\infty = \max_x |f(x)|
\]

The convolution \( f \ast g \) of \( f, g : G \to \mathbb{C} \) is defined by \( f \ast g(x) = \mathbb{E}_{u+v=x} [f(u)g(v)] \).

If \( \widehat{f}, \widehat{g} : \widehat{G} \to \mathbb{C} \), then

\[
\|\widehat{f}\|_p = \left( \sum_{\chi} |\widehat{f}(\chi)|^p \right)^{1/p}, \quad \|\widehat{f}\|_\infty = \max_{\chi} |\widehat{f}(\chi)|
\]
and

\[ \hat{f} \ast \hat{g}(\chi) = \sum_{\chi_1 \chi_2 = \chi} \hat{f} (\chi_1) \hat{g} (\chi_2) \]

and if \( G = \mathbb{Z}_n \), then the last formula could be written as \( \hat{f} \ast \hat{g}(r) = \sum_{s+t=r} \hat{f}(s) \hat{g}(t) \).

**Properties of the Fourier transform.**

The following are easy to prove but surprisingly useful:

1. \( \langle \hat{f}, \hat{g} \rangle = \langle f, g \rangle \) (Parseval’s identity). This also implies \( \| \hat{f} \|_2 = \| f \|_2 \).
2. \( \hat{f} \ast g(\chi) = \hat{f}(\chi) \hat{g}(\chi) \) (convolution identity). This is useful because convolutions come up a lot in nature, e.g. coefficients of polynomials.
3. \( f(x) = \sum_{\chi} \hat{f}(\chi) \chi(x) \) (inversion formula). For the cyclic group, this translates as \( f(x) = \sum_{r} \hat{f}(r) e^{inx} \)

***

**Proof.**

(1) We have

\[ \langle \hat{f}, \hat{g} \rangle = \sum_{\chi} \hat{f}(\chi) \overline{\hat{g}(\chi)} = \sum_{\chi} \mathbb{E}_x \left[ f(x) \overline{\chi(x)} \right] \mathbb{E}_y \left[ \overline{g(y)} \chi(y) \right] \]

\[ = \mathbb{E}_{x,y} \left[ f(x) \overline{g}(y) \sum_{\chi} \chi(x^{-1} y) \right] \]

But

\[ \sum_{\chi} \chi(u) = \begin{cases} |G|, & \text{if } u = \text{id}_G \\ 0, & \text{otherwise} \end{cases} \]

So, this equals

\[ \mathbb{E}_{x,y} \left[ f(x) \overline{g}(y) |G| \delta_{xy} \right] = \mathbb{E}_x \left[ f(x) \overline{g(x)} \right] = \langle f, g \rangle \]

(2) We have

\[ \hat{f} \ast g(\chi) = \mathbb{E}_x \left[ f \ast g(x) \overline{\chi(x)} \right] = \mathbb{E}_x \left[ \mathbb{E}_{u+v=x} \left[ f(u) g(v) \overline{\chi(u+v)} \right] \right] \]

\[ = \mathbb{E}_x \left[ \mathbb{E}_{u+v=x} \left[ f(u) g(v) \chi(u) \overline{\chi(v)} \right] \right] = \mathbb{E}_x \left[ f(u) \overline{\chi(u)} \right] \mathbb{E}_u \left[ g(v) \overline{\chi(v)} \right] = \hat{f}(\chi) \hat{g}(\chi) \]

(3) We have

\[ \sum_{\chi} \hat{f}(\chi) g(\chi) = \sum_{\chi} \mathbb{E}_y \left[ f(y) \overline{\chi(y)} \chi(x) \right] = \mathbb{E}_y \left[ f(y) \sum_{\chi} \chi(xy^{-1}) \right] = \mathbb{E}_y \left[ f(y) |G| \delta_{xy} \right] = f(x) \]

\( \square \)

We often take Fourier transforms of characteristic functions of subsets of \( G \). If \( A \) is such a subset, write

\[ A(x) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{otherwise} \end{cases} \]

Then \( \hat{A}(\chi_0) = \mathbb{E}_x \left[ A(x) \right] = |A|/|G| \) where \( \chi_0 \) is the trivial character, and

\[ \sum_{\chi} |\hat{A}(\chi)|^2 = \mathbb{E}_x \left[ |A(x)|^2 \right] = \mathbb{E}_x \left[ A(x) \right] = |A|/|G| \]
We’re all set now to prove something.

**Theorem 1.10.** Let $A \subseteq \mathbb{F}_3^n$ be a subset of density $\geq 8/n$. Then there exist distinct $x, y, z \in A$ such that $x + y + z = 0$.

**Remark 1.11.** Because $y = -2y$, this is like saying that we have an arithmetic progression. Another way to say it is that $x, y, z$ form an affine line. Also, observe the density $8/n \to 0$ as $n \to \infty$, and that $n = \Theta(\log |\mathbb{F}_3^n|)$.

**Proof.** We would like to show that

$$E_{x+y+z=0} [A(x)A(y)A(z)] > 3^{-n}$$

where the contribution of the trivial things is at most $3^{-n}$. Notice that we’re not trying to find a single triple $(x, y, z)$, but rather how many there are (because there will usually not be a single canonical one). This idea is fundamental throughout the subject.

Now notice that the left hand side is

$$A * A * A(0) = \sum_{\chi} \hat{A}(\chi)^3$$

by the inversion formula and convolution identity. Going on,

$$= \hat{A}(\chi_0)^3 + \sum_{\chi \neq \chi_0} \hat{A}(\chi)^3$$

But $\hat{A}(\chi_0)^3 = \alpha^3$ where $\alpha = \text{density of } A$. The intuition is that the first term is the contribution from the ‘random’ part of $A$, and the second from the ‘structured’ part.

Also,

$$\left| \sum_{\chi \neq \chi_0} \hat{A}(\chi)^3 \right| \leq \max_{\chi \neq \chi_0} \left| \hat{A}(\chi) \right| \sum_{\chi} \left| \hat{A}(\chi) \right|^2 = \alpha \max_{\chi \neq \chi_0} \left| \hat{A}(\chi) \right|$$

So we’re done if

$$\max_{\chi \neq \chi_0} \left| \hat{A}(\chi) \right| < \frac{\alpha^2}{2}$$

and $\alpha^3/2 > 3^{-n}$. But this might very well be false, in which case however we have another piece of information: that $A$ has a large Fourier coefficient.

If $\chi$ is a non-trivial character with $|\hat{A}(\chi)| \geq \alpha^2/2$, then it is given by a formula $x \mapsto \omega^{r \cdot x}$ for some $r \in \mathbb{F}_3^n$ for $r \neq 0$, where $\omega = e^{2\pi i/3}$. For $i = 0, 1, 2$, let

$$X_i = \{ x \in \mathbb{F}_3^n \mid r \cdot x = i \}$$

Then $X_0, X_1, X_2$ are affine subspaces of codimension 1. Also let $f(x) = A(x) - \alpha$. Observe that $\hat{f}(\chi) = \hat{A}(\chi)$ except if $\chi = \chi_0$ when $\hat{f}(\chi)$. So the remaining case is

$$\left| \hat{f}(r) \right| = \left| \mathbf{E}_x [f(x)\omega^{-r \cdot x}] \right| \geq \alpha^2/2$$

But

$$\left| \mathbf{E}_x [f(x)\omega^{-r \cdot x}] \right| = \frac{1}{3} \left| \mathbf{E}_{x \in X_0} [f(x)] + \omega^{-1} \mathbf{E}_{x \in X_1} [f(x)] + \omega^{-2} \sum_{x \in X_2} f(x) \right|$$

By the triangle inequality there exists $i$ such that

$$\left| \mathbf{E}_{x \in X_i} [f(x)] \right| > \alpha^2/2$$
Now if $E_{x \in X_i} [f(x)] \leq -\alpha^2/2$, there exists $j$ such that

$$E_{x \in X_j} [f(x)] \geq \alpha^2/4,$$

because $\sum_i E_{x \in X_i} [f(x)] = 0$. If $E_{x \in X_j} [f(x)] \geq \alpha^2/4$, then

$$E_{x \in X_j} [A(x)] \geq \alpha + \alpha^2/4 \implies |A \cap X_j| \geq (\alpha + \alpha^2/4 |X_j|$$

So now we switched to a subspace of dimension one less and increased the density. If we keep doing it, it can happen only so many times.

So it remains to discuss how many times we can iterate the function $\theta \mapsto \theta + \theta^2/4$ starting at $\alpha$ before one reaches 1. Observe that it takes at most $4/\alpha$ iterations to get from $\theta$ to $2\theta$. Hence the maximum number of iterations is at most

$$\frac{4}{\alpha} + \frac{4}{4\alpha} + \frac{4}{4\alpha} + \cdots \leq \frac{8}{\alpha}$$

and therefore if $n \geq 8/\alpha$, we have enough dimensions to keep iterating. This shows that a density of $8/n$ suffices.

**Remark 1.12.** Is this result best possible? Is there a lower bound? Observe that the set of sequences with all coordinates 0 or 1 contains no triple $x + y + z = 0$ of distinct elements, and has density $(2/3)^n$

***

**Remark 1.13.** Observe that if $\alpha \geq 8/n$ then $\alpha + \alpha^2/4 \geq 8/(n-1)$. Also,

$$\frac{1}{4} \left( \frac{8}{n} \right)^3 = \frac{128}{n^3} \geq 3^{-n}$$

for all $n \in \mathbb{N}$.

Recall results on approximating algebraic numbers: for every algebraic $\alpha$ and $\varepsilon > 0$, there exists $c_{\alpha,\varepsilon}$ such that

$$\left| \alpha - \frac{p}{q} \right| \geq \frac{c_{\alpha,\varepsilon}}{q^{2+\varepsilon}}$$

for all rational $p/q$. Roth managed to prove that!

**Theorem 1.14** (Roth’s theorem). We shall show that there exists $C$ such that for all $n$ and all subsets $A \subseteq \{1, \ldots, n\}$ of density at least $\frac{C}{\log \log n}$, $A$ contains an arithmetic progression (ap) of length 3.

First, some preliminaries:

**Dilation rule:**

**Lemma 1.15.** If $f : \mathbb{Z}_n \to \mathbb{C}$ and $a$ is invertible modulo $n$, and $g(x) = f(ax)$ for all $x$, then $\hat{g}(r) = \hat{f}(a^{-1}r)$ for all $r$.

**Proof.** We have

$$\hat{g}(r) = E_x [g(x)\omega^{-rx}] = E_x [f(ax)\omega^{-rx}] = E_x [f(x)\omega^{-a^{-1}rx}] = \hat{f}(a^{-1}r)$$

Note also that if $f$ is real-valued, then $\hat{f}(-r) = \overline{\hat{f}(r)}$ for all $r$. We’d like to use Fourier analysis on $\mathbb{Z}_n$ to count APs on $\{1, \ldots, n\}$, so we can pretend it wraps around but restrict the second and third elements to the middle third of the circle so that we get genuine APs.
So let $A, B, C$ be three subsets of $\mathbb{Z}_n$. We shall look at APs $(x, y, z)$ with $x \in A, y \in B, z \in C$. Let’s look at

$$E_{x+z=2y} [A(x)B(y)C(z)] = E_{x+z=2y} [A(x)B(y/2)C(z)]$$

(assume $n$ is odd)

Write $B_2(y)$ for $B(y/2)$. Note that $\hat{B}_2(r) = \hat{B}(2r)$ by the dilation rule. So the RHS is

$$E_\alpha [E_{x+z=u} [A(x)C(z)] B_2(u)] = \langle A \ast C, B_2 \rangle$$

$$= \langle \hat{A}C, \hat{B}_2 \rangle$$

(convolution + Parseval)

$$= \sum_r \hat{A}(r)\hat{C}(r)\hat{B}(-2r) \quad (B_2(r) = \hat{B}(2r))$$

Let $A, B, C$ have densities $\alpha, \beta, \gamma$. The last expression equals

$$\alpha\beta\gamma + \sum_{r \neq 0} \hat{A}(r)\hat{C}(r)\hat{B}(-2r)$$

We’d like the second term not to cancel the first term. But

$$\left| \sum_{r \neq 0} \hat{A}(r)\hat{C}(r)\hat{B}(-2r) \right| \leq \max_{r \neq 0} |\hat{A}(r)| \sum_r |\hat{C}(r)| \left| \hat{B}(-2r) \right|$$

$$\leq \max_{r \neq 0} |\hat{A}(r)| \left( \sum_r |\hat{C}(r)|^2 \right)^{1/2} \left( \sum_r |\hat{B}(-2r)|^2 \right)^{1/2}$$

(Cauchy-Schwarz)

$$= \max_{r \neq 0} |\hat{A}(r)| \beta^{1/2}\gamma^{1/2}$$

(by Parseval)

So, we will be done if $\max_{r \neq 0} |\hat{A}(r)| < \alpha\beta^{1/2}\gamma^{1/2}/2$ and $\alpha\beta\gamma/2 > 1/n$. Now let’s take $B = C = A \cap [n/3, 2n/3]$. Let’s not worry about integer parts, because they hardly make any difference.

If $\beta$ (and therefore $\gamma$) is less than $\alpha/5$ then either $A \cap [0, n/3] > 2\alpha n/5$ or $A \cap [2n/3, n] > 2\alpha n/5$. Either way, we can pass to a subprogression of length $\geq n/4$ in which $A$ has density $\geq 6\alpha/5$. That’s very good!

Otherwise, we have some $r \neq 0$ such that $|\hat{A}(r)| \geq \frac{\alpha^2}{10}$. As before, let $f(x) = A(x) - \alpha$. Then $|\hat{f}(r)| \geq \frac{\alpha^2}{10}$ and note that $\hat{f}(r) = E_x [f(x)\omega^{-rx}] = \frac{1}{\sqrt{\pi}} \sum_x f(x)\omega^{-rx}$.

Now we need a

**Lemma 1.16.** For every $r, \varepsilon$ we can split $\{1, \ldots, n\}$ up into APs of length at least $c(\varepsilon)\sqrt{n}$, on each of which $\omega^{-rx}$ varies by at most $\varepsilon$.

**Proof.** Let $m$ be an integer (around $\sqrt{n}$) to be chosen later. By the pigeonhole principle, we can find $1 \leq u < v \leq m$ such that

$$|\omega^{-rv} - \omega^{-ru}| \leq \frac{2\pi}{m}$$

Let $d = v - u$. Then $|\omega^{-rd} - 1| \leq \frac{2\pi}{m}$, and therefore by the triangle inequality $|\omega^{-jrd} - 1| \leq \frac{2\pi j}{m}$ for all $j$. So $\omega^{-rx}$ does not vary by more than $\varepsilon$ for any AP of the form $x, x + d, \ldots, x + rd$ as long as $t \leq \frac{\varepsilon m}{2\pi}$.

Now partition $\{1, \ldots, n\}$ into residue classes modulo $d$. Each residue class has size $\geq \frac{n}{2m}$. So we can partition it into APs of the above type all of length between $\frac{cm}{4\pi}$ and $\frac{cm}{2\pi}$, provided $\frac{cm}{2\pi} \leq \frac{n}{2m}$. So take $m = \sqrt{\frac{2m}{\varepsilon}}$, so all our progressions can be taken to have length at least

$$\frac{cm}{\sqrt{\pi}} \sqrt{\frac{2m}{\varepsilon}} = \frac{1}{4\sqrt{\pi}} \sqrt{n} \quad \square$$
Let $P_1, \ldots, P_k$ be progressions given by the lemma with $\varepsilon = \alpha^2/20$. Then
\[
\left| \frac{1}{n} \sum_{x} f(x) \omega^{-rx} \right| \leq \frac{1}{n} \sum_{i=1}^{k} \left| \sum_{x \in P_i} f(x) \omega^{-rx} \right|
\]
Let $x_i$ be some element of $P_i$. Then the right hand side is
\[
\leq \frac{1}{n} \sum_{i=1}^{k} \left( \left| \sum_{x \in P_i} f(x) \omega^{-rx_i} \right| + \sum_{x \in P_i} f(x) (\omega^{-rx} - \omega^{-rx_i}) \right)
\]
\[
\leq \frac{1}{n} \sum_{i=1}^{k} \left| \sum_{x \in P_i} f(x) \right| + \frac{\alpha^2}{20}
\]
Therefore, $\sum_{i=1}^{k} |\sum_{x \in P} f(x)| \geq \frac{\alpha^2 n}{20}$.

Also, $\sum_{i=1}^{k} \sum_{x \in P_i} f(x) = 0$ and so there exists $i$ such that
\[
\left| \sum_{x \in P_i} f(x) \right| + \sum_{x \in P_i} f(x) \geq \frac{\alpha^2}{20} |P_i|
\]
hence
\[
\sum_{x \in P_i} f(x) \geq \frac{\alpha^2}{40} |P_i|
\]
which means $|A \cap P_i| \geq \left( \alpha + \frac{\alpha^2}{40} \right) |P_i|$.

***

We showed last time that if $A$ contains no 3AP and has density $\alpha$, you can find an AP $P$ such that $|A \cap P| \geq (\alpha + \alpha^2/40) |P|$ and $|P| \geq \frac{\alpha \sqrt{n}}{10}$ or something like that.

So we iterate. At each iteration, $\frac{100n}{\alpha^2} \rightarrow \sqrt{\frac{100n}{\alpha^2}}$; so at most $80/\alpha$ iterations are possible. So the final $n$ is at least
\[
\frac{\alpha}{10} \left( \frac{100n}{\alpha^2} \right)^{\frac{80}{\alpha}}
\]
which we need to be bigger than $10\alpha^{-3}$ (take these calculations with a slight pinch of salt). Take logs: we want
\[
\log \frac{\alpha}{10} + \left( \frac{1}{2} \right)^{\frac{80}{\alpha}} \left( \log 100n + 2 \log \frac{1}{\alpha} \right) \geq \log 10 + 3 \log \frac{1}{\alpha}
\]
It will be sufficient if we show that
\[
\left( \frac{1}{2} \right)^{\frac{80}{\alpha}} \log n \geq 3 \log \frac{5}{\alpha}
\]
and then take logs again:
\[
\log \log n - \frac{80}{\alpha} \log 2 \geq \log 3 + \log \log \frac{5}{\alpha}
\]
so it will be sufficient if $\log \log n \geq \frac{160}{\alpha}$, i.e. $\alpha \geq \frac{160}{\log \log n}$.

**Theorem 1.17.** Let $A$ be a subset of $\mathbb{F}_p^n$ of density $\alpha$. Then
\[
2A - 2A = \{ x + y - z - w \mid x, y, z, w \in A \}
\]
contains a subspace of codimension at most $\alpha^{-2}$. 


Proof. We have $x \in 2A - 2A$ iff $A \ast A \ast (-A) \ast (-A)(x) \neq 0$. By the convolution identity and the inversion formula,

$$A \ast A \ast (-A) \ast (-A)(x) = \sum_r |\hat{A}(r)|^4 \omega^{r \cdot x}$$

where $\omega = e^{2\pi i/p}$. Let $K = \{ r \mid |\hat{A}(r)| \geq \alpha^{3/2} \}$. Let $x$ be such that $r \cdot x = 0$ for every $r \in K$. Then

$$\sum_{r \in K} |\hat{A}(r)|^4 \omega^{r \cdot x} = \sum_{r \in K} |\hat{A}(r)|^4 \geq |\hat{A}(0)|^4 = \alpha^4$$

Also,

$$\left| \sum_{r \not\in K} |\hat{A}(r)|^4 \omega^{r \cdot x} \right| \leq \sum_{r \not\in K} |\hat{A}(r)|^4 \leq \max_{r \not\in K} |\hat{A}(r)|^2 \sum_r |\hat{A}(r)|^2 < \alpha^3 \alpha$$

Therefore, $x \in 2A - 2A$. It remains to check how large $K$ can be. But

$$\alpha \geq \sum_{r \in K} |\hat{A}(r)|^2 \geq |K| \alpha^3$$

and hence $|K| \leq \alpha^{-2}$ and we’re done. □

Why is this an interesting result? It turns out there’s a nice theory of sets with small sum-sets (i.e. $A + A$ not much bigger than $A$). An extreme case in $\mathbb{F}_p^n$ is a subgroup.

Also, this result gets worse as $p$ grows.

What would an analogue of this over $\mathbb{Z}_p$ look like? It’s not clear, as it doesn’t have subgroups (as opposed to $\mathbb{F}_p^n$, which had plenty).

Definition 1.18. Let $K \subseteq \hat{Z}_N$ and let $\varepsilon > 0$. The Bohr set $B(K, \varepsilon)$ is

$$\{ x \in \mathbb{Z}_N \mid |1 - \omega^{r \cdot x}| < \varepsilon \text{ for each } r \in K \}$$

Theorem 1.19. Let $A \subseteq \mathbb{Z}_N$ have density $\alpha$. Then $2A - 2A$ contains a Bohr set $B(K, \sqrt{2})$ with $|K| \leq \alpha^{-2}$.

Proof. Similarly to previous result, let $K = \{ r \mid |\hat{A}(r)| \geq \alpha^{3/2} \}$. Then as before $|K| \leq \alpha^{-2}$. If $x \in B(K, \sqrt{2})$, then

$$\Re \left( \sum_{r \in K} |\hat{A}(r)|^4 \omega^{r \cdot x} \right) \geq \alpha^4$$

by the choice of $K$ (since $|\hat{A}(0)|^4 = \alpha^4$ and $\Re(\omega^{r \cdot x}) \geq 0$ for $r \in K$). As before,

$$\left| \sum_{r \not\in K} |\hat{A}(r)|^4 \omega^{r \cdot x} \right| < \alpha^4$$

and we’re done. This is important because Bohr sets have good ‘subspace-like’ structure. □

If you start thinking about improving the bounds in Roth’s theorem, you’ll see that passing to a new AP every time is expensive; but you can use Bohr-sets instead.