6 Forbidding 3-Term Arithmetic Progressions

Chapter Highlights

- Fourier analytic proof of Roth's theorem
- Finite field model in additive combinatorics: $\mathbb{F}_p^n$ as a model for the integers
- Basics of discrete Fourier analysis
- Density increment argument in the proof of Roth's theorem
- The polynomial method proof of Roth's theorem in $\mathbb{F}_3^n$
- Arithmetic analogue of the regularity lemma, and application to Roth's theorem with popular difference

In this chapter, we study Roth's theorem, which says that every 3-AP-free subset of $[N]$ has size $o(N)$.

Previously, in Section 2.4, we gave a proof of Roth's theorem using the graph regularity lemma. The main goal of this chapter is to give a Fourier analytic proof of Roth's theorem. This is also Roth's original proof (1953).

We begin by proving Roth's theorem in the **finite field model**. That is, we first prove an analogue of Roth's theorem in $\mathbb{F}_3^n$. The finite field vector space serves as a fruitful playground for many additive combinatorics problems. Techniques such as Fourier analysis are often simpler to carry out in the finite field model. After we develop the techniques in the finite field model, we then prove Roth's theorem in the integers. It can be a good idea to first try out ideas in the finite field model before bringing them to the integers, as there maybe additional technical difficulties in the integers.

Later in Section 6.5, we will see a completely different proof of Roth's theorem in $\mathbb{F}_3^n$ using the **polynomial method**, which gives significantly better quantitative bounds. This proof surprised many people at the time of its discovery. However, unless Fourier analysis, this polynomial method technique only applies to the finite field setting, and it is unknown how to apply it to the integers.

6.1 Fourier Analysis in Finite Field Vector Spaces

We review some basic facts about Fourier analysis in $\mathbb{F}_p^n$ for a prime p. Everything here can be extended to arbitrary abelian groups. As we saw in Section 3.3, eigenvalues of

Cayley graphs on an abelian group and the Fourier transform are intimately related.

Throughout this section, we fix a prime p and let

$$\omega = \exp(2\pi i/p).$$

Definition 6.1.1 (Fourier transform in $\mathbb{F}_p^n$)

The Fourier transform of $f\colon \mathbb{F}_p^n \to \mathbb{C}$ is a function $\widehat{f}\colon \mathbb{F}_p^n \to \mathbb{C}$ defined by setting, for each $r \in \mathbb{F}_p^n$,

$$\widehat{f}(r) := \mathbb{E}_{x\in\mathbb{F}_p^n} f(x)\omega^{-r\cdot x} = \frac{1}{p^n}\sum_{x\in\mathbb{F}_p^n} f(x)\omega^{-r\cdot x}$$

where $r \cdot x = r_1x_1 + \cdots + r_nx_n$.

In particular, $\widehat{f}(0) = \mathbb{E}f$ is the average of f. This value often plays a special role compared to other values $\widehat{f}(r)$.

To simplify notation, it is generally understood that the variables being averaged or summed over are varying uniformly in the domain $\mathbb{F}_p^n$.

Let us now state several important properties of the Fourier transform. We will see that all these properties are consequences of the orthogonality of the Fourier basis.

The next result allows us to write f in terms of $\widehat{f}$.

Theorem 6.1.2 (Fourier inversion formula)

Let $f\colon \mathbb{F}_p^n \to \mathbb{C}$. For every $x \in \mathbb{F}_p^n$,

$$f(x) = \sum_{r\in\mathbb{F}_p^n} \widehat{f}(r)\omega^{r\cdot x}.$$

The next result tells us that the Fourier transform preserves inner products.

Theorem 6.1.3 (Parseval's identity)

Given $f, g\colon \mathbb{F}_p^n \to \mathbb{C}$, we have

$$\mathbb{E}_{x\in\mathbb{F}_p^n} f(x)\overline{g(x)} = \sum_{r\in\mathbb{F}_p^n} \widehat{f}(r)\overline{\widehat{g}(r)}.$$

In particular, as a special case ($f = g$),

$$\mathbb{E}_{x\in\mathbb{F}_p^n} |f(x)|^2 = \sum_{r\in\mathbb{F}_p^n} |\widehat{f}(r)|^2.$$

As is nowadays the standard in additive combinatorics, we adopt the following convention for the Fourier transform in finite abelian groups:

$$\begin{aligned} &\textit{average in physical space} \quad \mathbb{E} f \\ &\textit{and sum in frequency (Fourier) space} \quad \sum \widehat{f}. \end{aligned}$$

For example, following this convention, we define an "averaging" inner product for functions $f, g\colon \mathbb{F}_p^n \to \mathbb{C}$ by

$$\langle f, g \rangle := \mathbb{E}_{x \in \mathbb{F}_p^n} \overline{f(x)} g(x) \qquad \text{and} \qquad \|f\|_2 := \langle f, f \rangle^{1/2}.$$

In the frequency/Fourier domain, we define the "summing" inner product for functions $\alpha, \beta\colon \mathbb{F}_p^n \to \mathbb{C}$ by

$$\langle \alpha, \beta \rangle_{\ell^2} := \sum_{x \in \mathbb{F}_p^n} \overline{\alpha(x)} \beta(x). \qquad \text{and} \qquad \|\alpha\|_{\ell^2} := \langle \alpha, \alpha \rangle_{\ell^2}^{1/2}.$$

Writing $\gamma_r : \mathbb{F}_p^n \to \mathbb{C}$ for the function defined by

$$\gamma_r(x) := \omega^{r \cdot x}$$

(this is a **character** of the group $\mathbb{F}_p^n$), the Fourier transform can be written as

$$\widehat{f}(r) = \mathbb{E}_x \overline{\gamma_r(x)} f(x) = \langle \gamma_r, f \rangle. \tag{6.1.1}$$

Parseval's identity can be stated as

$$\langle f, g \rangle = \langle \widehat{f}, \widehat{g} \rangle_{\ell^2} \qquad \text{and} \qquad \|f\|_2 = \|\widehat{f}\|_{\ell^2}.$$

With these conventions, we often do not need to keep track of normalization factors.

The above identities can be proved via direct verification, by plugging in the formula for the Fourier transform. We give a more conceptual proof below.

Proof of the Fourier inversion formula (Theorem 6.1.2). Let $\gamma_r(x) = \omega^{r \cdot x}$. Then the set of functions

$$\{\gamma_r : r \in \mathbb{F}_p^n\}$$

forms an orthonormal basis for the space of functions $\mathbb{F}_p^n \to \mathbb{C}$ with respect to the averaging inner product $\langle \cdot, \cdot \rangle$. Indeed,

$$\langle \gamma_r, \gamma_s \rangle = \mathbb{E}_x \omega^{(s-r) \cdot x} = \begin{cases} 1 & \text{if } r = s, \\ 0 & \text{if } r \neq s \end{cases}$$

Furthermore, there are p^n functions γ_r (as r ranges over $\mathbb{F}_p^n$). So they form a basis of the p^n-dimensional vector space of all functions $f\colon \mathbb{F}_p^n \to \mathbb{C}$. We will call this basis the **Fourier basis**.

Now, given an arbitrary $f\colon \mathbb{F}_p^n \to \mathbb{C}$, the "coordinate" of f with respect to the basis vector γ_r of the Fourier basis is $\langle \gamma_r, f\rangle = \widehat{f}(r)$ by (6.1.1). So

$$f = \sum_r \widehat{f}(r)\gamma_r.$$

This is precisely the Fourier inversion formula. □

Proof of Parseval's identity (Theorem 6.1.3). Continuing from the previous proof, since the Fourier basis is orthonormal, we can evaluate $\langle f, g\rangle$ with respects to coordinates in this basis, thereby by yielding

$$\langle f, g\rangle = \sum_{r\in\mathbb{F}_p^n} \overline{\langle f, \gamma_r\rangle}\,\langle g, \gamma_r\rangle = \sum_{r\in\mathbb{F}_p^n} \overline{\widehat{f}(r)}\widehat{g}(r).$$ □

Remark 6.1.4. Parseval's identity is sometimes also referred to by the name **Plancheral**. Parseval derived the identity for the Fourier series of a periodic function on $\mathbb{R}$, whereas Plancheral derived it for the Fourier transform on $\mathbb{R}$.

The convolution is an important operation.

Definition 6.1.5 (Convolution)

Given $f, g\colon \mathbb{F}_p^n \to \mathbb{C}$, define $f * g\colon \mathbb{F}_p^n \to \mathbb{C}$ by

$$(f * g)(x) := \mathbb{E}_{y\in\mathbb{F}_p^n} f(y)g(x-y).$$

In other words, $(f*g)(x)$ is the average of $f(y)g(z)$ over all pairs (y, z) with $y+z = x$.

Example 6.1.6. (a) If f is supported on $A \subset \mathbb{F}_p^n$ and g is supported on $B \subset \mathbb{F}_p^n$, then $f * g$ is supported on the sum set $A + B = \{a + b : a \in A, b \in B\}$.

(b) Let W be a subspace of $\mathbb{F}_p^n$. Let $\mu_W = (p^n/|W|)1_W$ be the indicator function on W normalized so that $\mathbb{E}\mu_W = 1$. Then for any $f\colon \mathbb{F}_p^n \to \mathbb{C}$, the function $f * \mu_W$ is obtained from f by replacing its value at x by its average value on the coset $x + W$.

The second example suggests that convolution can be thought of as smoothing a function, damping its potentially rough perturbations.

The Fourier transform conveniently converts convolutions to multiplication.

Theorem 6.1.7 (Convolution identity)

For any $f, g\colon \mathbb{F}_p^n \to \mathbb{C}$ and any $r \in \mathbb{F}_p^n$,

$$\widehat{f * g}(r) = \widehat{f}(r)\widehat{g}(r).$$

Proof. We have

$$\begin{aligned}\widehat{f * g}(r) &= \mathbb{E}_x (f * g)(x)\omega^{-r\cdot x} = \mathbb{E}_x \mathbb{E}_{y,z:y+z=x} f(y)g(z)\omega^{-r\cdot(y+z)} \\ &= \mathbb{E}_{y,z} f(y)g(z)\omega^{-r\cdot(y+z)} = \left(\mathbb{E}_y f(y)\omega^{-r\cdot y}\right)\left(\mathbb{E}_z g(z)\omega^{-r\cdot z}\right) = \widehat{f}(r)\widehat{g}(r). \qquad \square\end{aligned}$$

By repeated applications of the convolution identity, we have

$$(f_1 * \cdots * f_k)^\wedge = \widehat{f_1}\widehat{f_2}\cdots\widehat{f_k}$$

(here we write $f^\wedge$ for $\widehat{f}$ for typographical reasons).

Now we introduce a quantity relevant to Roth's theorem on 3-APs.

Definition 6.1.8 (3-AP density)

Given $f, g, h\colon \mathbb{F}_p^n \to \mathbb{C}$, we write

$$\Lambda(f, g, h) := \mathbb{E}_{x,y} f(x)g(x+y)h(x+2y), \tag{6.1.2}$$

and

$$\Lambda_3(f) := \Lambda(f, f, f), \tag{6.1.3}$$

Note that for any $A \subset \mathbb{F}_p^n$,

$$\Lambda(1_A) = p^{-2n} \left|\{(x, y) : x, x+y, x+2y \in A\}\right| = \text{“3-AP density of } A\text{.”}.$$

Here we include “trivial” 3-APs (i.e., those with with $y = 0$).

The following identity, relating the Fourier transform and 3-APs, plays a central role in the Fourier analytic proof of Roth's theorem.

Proposition 6.1.9 (Fourier and 3-AP)

Let p be an odd prime. If $f, g, h : \mathbb{F}_p^n \to \mathbb{C}$, then

$$\Lambda(f, g, h) = \sum_r \widehat{f}(r)\widehat{g}(-2r)\widehat{h}(r).$$

We will give two proofs of this proposition. The first proof is more mechanically straightforward. It is similar to the proof of the convolution identity earlier. The second proof directly applies the convolution identity, and may be a bit more abstract/conceptual.

First proof. We expand the left-hand side using the formula for Fourier inversion.

$$\begin{aligned}&\mathbb{E}_{x,y} f(x)g(x+y)h(x+2y)\\ &\quad= \mathbb{E}_{x,y}\left(\sum_{r_1}\widehat{f}(r_1)\omega^{r_1\cdot x}\right)\left(\sum_{r_2}\widehat{g}(r_2)\omega^{r_2\cdot(x+y)}\right)\left(\sum_{r_3}\widehat{h}(r_3)\omega^{r_3\cdot(x+2y)}\right)\\ &\quad= \sum_{r_1,r_2,r_3}\widehat{f}(r_1)\widehat{g}(r_2)\widehat{h}(r_3)\mathbb{E}_x\omega^{x\cdot(r_1+r_2+r_3)}\mathbb{E}_y\omega^{y\cdot(r_2+2r_3)}\\ &\quad= \sum_{r_1,r_2,r_3}\widehat{f}(r_1)\widehat{g}(r_2)\widehat{h}(r_3)1_{r_1+r_2+r_3=0}1_{r_2+2r_3=0}\\ &\quad= \sum_{r}\widehat{f}(r)\widehat{g}(-2r)\widehat{h}(r).\end{aligned}$$

In the last step, we use that $r_1+r_2+r_3=0$ and $r_2+2r_3=0$ together imply $r_1=r_2=r_3$. □

Second proof. Write $g_1(y) = g(-y/2)$. So $\widehat{g_1}(r) = \widehat{g}(-2r)$. Applying the convolution identity,

$$\begin{aligned}\mathbb{E}_{x,y} f(x)g(x+y)h(x+2y) &= \mathbb{E}_{x,y,z:x-2y+z=0} f(x)g(y)h(z)\\ &= \mathbb{E}_{x,y,z:x+y+z=0} f(x)g_1(y)h(z)\\ &= (f*g_1*h)(0)\\ &= \sum_r \widehat{f*g_1*h}(r) && \text{[Fourier inversion]}\\ &= \sum_r \widehat{f}(r)\widehat{g_1}(r)\widehat{h}(r) && \text{[Convolution identity]}\\ &= \sum_r \widehat{f}(r)\widehat{g}(-2r)\widehat{h}(r).\end{aligned}$$

□

Remark 6.1.10. In the following section, we will work in $\mathbb{F}_3^n$. Since $-2 = 1$ in $\mathbb{F}_3$ (and so $g_1 = g$ above), the proof looks even simpler. In particular, by Fourier inversion and the convolution identity,

$$\begin{aligned}\Lambda_3(1_A) &= 3^{-2n}\left|\{(x,y,z)\in A^3 : x+y+z=0\}\right|\\ &= (1_A*1_A*1_A)(0) = \sum_r (1_A*1_A*1_A)^\wedge(r) = \sum_r \widehat{1_A}(r)^3.\end{aligned} \tag{6.1.4}$$

When $A=-A$, the eigenvalues of the adjacency matrix of the Cayley graph $\mathrm{Cay}(\mathbb{F}_3^n, A)$ are $3^n\widehat{1_A}(r)$, $r\in\mathbb{F}_3^n$ (c.f. Section 3.3). The quantity $3^{2n}\Lambda_3(1_A)$ is the number of closed walks of length 3 in the Cayley graph $\mathrm{Cay}(\mathbb{F}_p^n, A)$. So the above identity is saying that the number of closed walks of length 3 in $\mathrm{Cay}(\mathbb{F}_3^n, A)$ equals to the third moment of the eigenvalues of the adjacency matrix, which is a general fact for every graph. (When $A \neq -A$, we can consider the directed or bipartite version of this argument.)

The following exercise generalizes the above identity.

Exercise 6.1.11. Let $a_1, \dots, a_k$ be nonzero integers, none divisible by the prime p. Let $f_1, \dots, f_k : \mathbb{F}_p^n \to \mathbb{C}$. Show that

$$\mathbb{E}_{x_1,\dots,x_k \in \mathbb{F}_p^n : a_1x_1+\cdots+a_kx_k=0} f_1(x_1)\cdots f_k(x_k) = \sum_{r\in\mathbb{F}_p^n} \widehat{f_1}(a_1r)\cdots\widehat{f_k}(a_kr).$$

6.2 Roth's Theorem in the Finite Field Model

In this section, we use Fourier analysis to prove the following finite field analogue of Roth's theorem (Meshulam 1995). Later in the chapter, we will convert this proof to the integer setting.

In an abelian group, a set A is said to be **3-AP-free** if A does not have three distinct elements of the form $x, x+y, x+2y$. A 3-AP-free subset of $\mathbb{F}_3^n$ is also called a **cap set**. The **cap set problem** asks to determine the size of the largest cap set in $\mathbb{F}_3^n$.

Theorem 6.2.1 (Roth's theorem in $\mathbb{F}_3^n$)
Every 3-AP-free subset of $\mathbb{F}_3^n$ has size $O(3^n/n)$.

Remark 6.2.2 (General finite fields). We work in $\mathbb{F}_3^n$ mainly for convenience. The argument presented in this section also shows that for every odd prime p, there is some constant C_p so that every 3-AP-free subset of $\mathbb{F}_p^n$ has size $\le C_p p^n/n$.

In $\mathbb{F}_3^n$, there are several equivalent interpretations of $x, y, z \in \mathbb{F}_3^n$ forming a 3-AP (allowing the possibility for a trivial 3-AP with $x = y = z$):

- $(x, y, z) = (x, x+d, x+2d)$ for some d;
- $x - 2y + z = 0$;
- $x + y + z = 0$;
- x, y, z are three distinct points of a line in $\mathbb{F}_3^n$ or are all equal;
- for each i, the i-th coordinates of x, y, z are all distinct or all equal.

Remark 6.2.3 (SET card game). The card same SET comes with a deck of 81 cards (see Figure 6.2.1). Each card one of three possibilities in each of the following four features:

- Number: 1, 2, 3;
- Symbol: diamond, squiggle, oval;
- Shading: solid, striped, open;
- Color: red, green, purple.

Each of the $3^4 = 81$ combinations appears exactly once as a card.

In this game, a combination of three cards is called a "set" if each of the four features shows up as all identical or all distinct among the three cards. For the example, the three

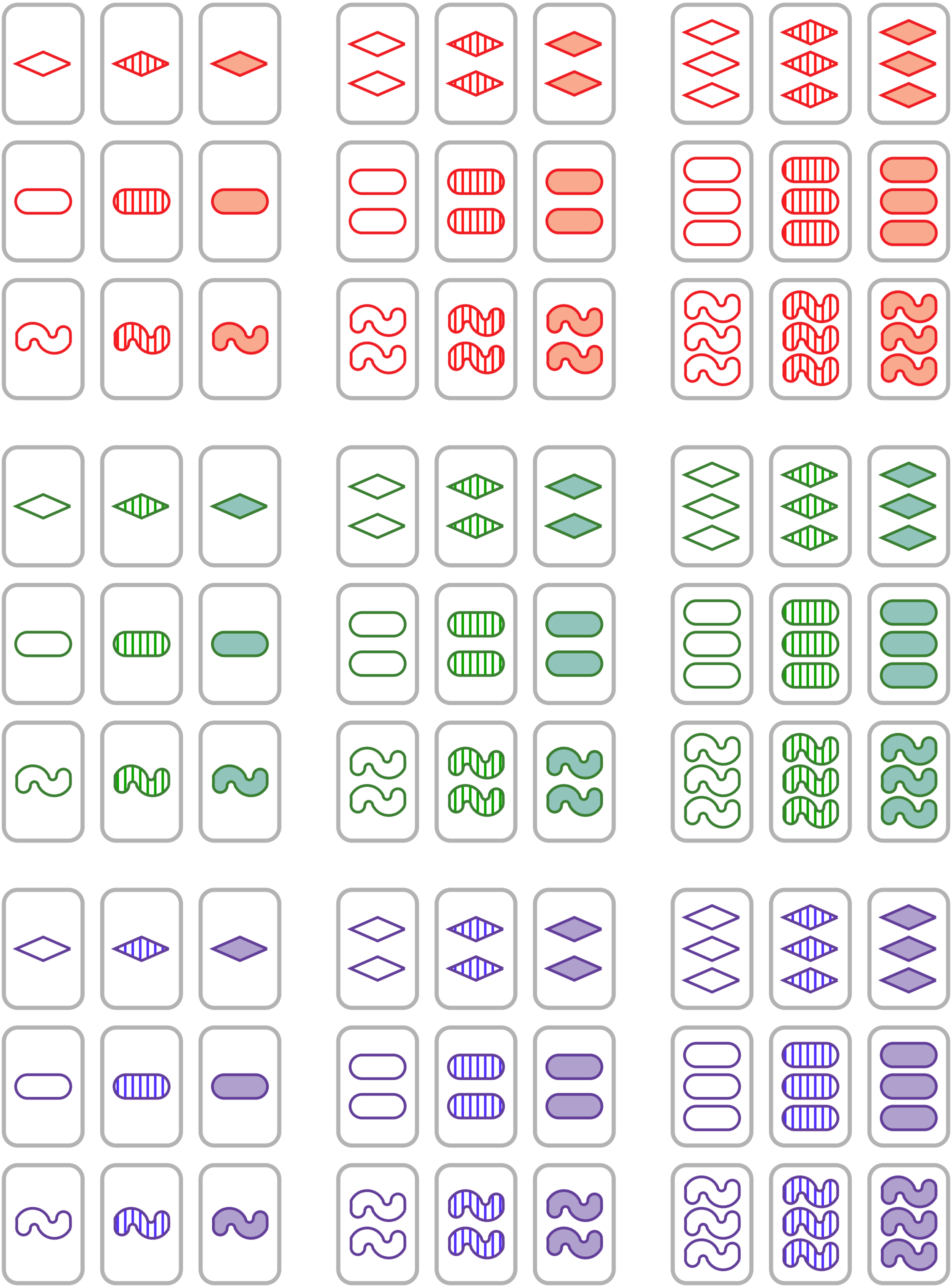
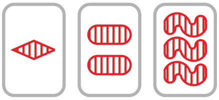


Figure 6.2.1: The complete deck of 81 cards in the game SET.

cards shown below form a “set”: number (all distinct), symbol (all distinct), shading (all striped), color (all red).



In a standard play of the game, the dealer lays down twelve cards on the table until some player finds a “set”, in which case the player keeps the three cards of the “set” as their score, then dealer replenishes the table by laying down more cards. If no set is found, then the dealer continues to lay down more cards until a set is found.

The cards of the game correspond to points of $\mathbb{F}_3^4$. A “set” is precisely a 3-AP. The cap set problem in $\mathbb{F}_3^4$ asks for the number of cards without a “set.” The size of the maximum cap set in $\mathbb{F}_3^4$ is 20 (Pellegrino 1970).

Here is the proof strategy of Roth’s theorem in $\mathbb{F}_3^n$:

(1) A 3-AP-free set has a large Fourier coefficient.

(2) A large Fourier coefficient implies density increment on some hyperplane.

(3) Iterate.

As in the proof of the graph regularity lemma (where we refined partitions to obtain an *energy increment*), the above process must terminate in a bounded number of steps since the density of a subset is always between 0 and 1.

Similar to what we saw in Chapter 3 on pseudorandom graphs, a set $A \subset \mathbb{F}_3^n$ has pseudorandom properties if and only if all its Fourier coefficients $\widehat{1_A}(r)$, for $r \neq 0$, are small in absolute value. When A is pseudorandom in this Fourier-uniform sense, the 3-AP-density of A is similar to that of a random set with the same density. On the flip side, a large Fourier coefficient in A points to non-uniformity along the direction of the Fourier character. Then we can restrict A to some hyperplane and extract a density increment.

The following counting lemma shows that a Fourier-uniform subset of $\mathbb{F}_3^n$ has 3-AP density similar to that of a random set. It has a similar flavor as the proof that **EIG** implies $\mathbf{C_4}$ in Theorem 3.1.1. It is also related to the counting lemma for graphons (Theorem 4.5.1). Recall the 3-AP-density Λ_3 from Definition 6.1.8.

Lemma 6.2.4 (3-AP counting lemma)

Let $f : \mathbb{F}_3^n \to [0, 1]$. Then

$$\left|\Lambda_3(f) - (\mathbb{E}f)^3\right| \leq \max_{r \neq 0} |\widehat{f}(r)| \, \|f\|_2^2 .$$

Proof. By Proposition 6.1.9 (also see (6.1.4)),

$$\Lambda_3(f) = \sum_r \widehat{f}(r)^3 = \widehat{f}(0)^3 + \sum_{r \neq 0} \widehat{f}(r)^3.$$

Since $\mathbb{E}f = \widehat{f}(0)$, we have

$$\left|\Lambda_3(f) - (\mathbb{E}f)^3\right| \le \sum_{r\neq 0} |\widehat{f}(r)|^3 \le \max_{r\neq 0} |\widehat{f}(r)| \cdot \sum_r |\widehat{f}(r)|^2 = \max_{r\neq 0} |\widehat{f}(r)| \, \|f\|_2^2 \, .$$

The final step is by Parseval. □

Remark 6.2.5. It would be insufficient to bound each term $|\widehat{f}(r)|^3$ by $\|\widehat{f}\|_\infty^3$. Instead, Parseval comes for the rescue. See Remark 3.1.19 for a similar issue.

Step 1. A 3-AP-free set has a large Fourier coefficient

Lemma 6.2.6 (3-AP-free implies large Fourier coefficient)
Let $A \subset \mathbb{F}_3^n$ and $\alpha = |A|/3^n$. If A is 3-AP-free and $3^n \ge 2\alpha^{-2}$, then there is $r \neq 0$ such that $|\widehat{1_A}(r)| \ge \alpha^2/2$.

Proof. Since A is 3-AP-free, $\Lambda_3(A) = |A|/3^{2n} = \alpha/3^n$, as all 3-APs are trivial, i.e., with common difference zero. By the counting lemma, Lemma 6.2.4,

$$\alpha^3 - \frac{\alpha}{3^n} = \alpha^3 - \Lambda_3(1_A) \le \max_{r\neq 0} |\widehat{1_A}(r)| \, \|1_A\|_2^2 = \max_{r\neq 0} |\widehat{1_A}(r)| \alpha .$$

By the hypothesis $3^n \ge 2\alpha^{-2}$, the left-hand side above is $\ge \alpha^3/2$. So there is some $r \neq 0$ with $|\widehat{1_A}(r)| \ge \alpha^2/2$. □

Step 2. A large Fourier coefficient implies density increment on some hyperplane

Lemma 6.2.7 (Large Fourier coefficient implies density increment)
Let $A \subset \mathbb{F}_3^n$ with $\alpha = |A|/3^n$. Suppose $|\widehat{1_A}(r)| \ge \delta > 0$ for some $r \neq 0$. Then A has density at least $\alpha + \delta/2$ when restricted to some hyperplane.

Proof. We have

$$\widehat{1_A}(r) = \mathbb{E}_x 1_A(x) \omega^{-r\cdot x} = \frac{\alpha_0 + \alpha_1\omega + \alpha_2\omega^2}{3}$$

where $\alpha_0, \alpha_1, \alpha_2$ are densities of A on the cosets of $r^\perp$. We want to show that one of $\alpha_0, \alpha_1, \alpha_2$ is significantly larger than α. This is easy to check directly, but let us introduce a trick that we will also use later in the integer setting.

We have $\alpha = (\alpha_0 + \alpha_1 + \alpha_2)/3$. By the triangle inequality,

$$\begin{aligned} 3\delta &\le \left|\alpha_0 + \alpha_1\omega + \alpha_2\omega^2\right| \\ &= \left|(\alpha_0 - \alpha) + (\alpha_1 - \alpha)\omega + (\alpha_2 - \alpha)\omega^2\right| \\ &\le |\alpha_0 - \alpha| + |\alpha_1 - \alpha| + |\alpha_2 - \alpha| \\ &= \sum_{j=0}^{2} \left(|\alpha_j - \alpha| + (\alpha_j - \alpha)\right). \end{aligned}$$

Consequently, there exists j such that $|\alpha_j - \alpha| + (\alpha_j - \alpha) \ge \delta$. Note that $|t| + t$ equals $2t$ if $t > 0$ and 0 if $t \le 0$. So $\alpha_j - \alpha \ge \delta/2$, as desired. □

Combining the previous two lemmas, here is what we have proved so far.

Lemma 6.2.8 (3-AP-free implies density increment)
Let $A \subset \mathbb{F}_3^n$ and $\alpha = |A|/3^n$. If A is 3-AP-free and $3^n \ge 2\alpha^{-2}$, then A has density at least $\alpha + \alpha^2/4$ when restricted to some hyperplane. □

We now view this hyperplane H as $\mathbb{F}_3^{n-1}$ (we may need to select a new origin for H if $0 \notin H$). The restriction of A to H, i.e., $A \cap H$, is now a 3-AP-free subset of H. The density increased from α to $\alpha + \alpha^2/4$. Next we iterate this density increment.

Remark 6.2.9 (Translation invariance). It is important that the pattern we are forbidding (3-AP) is translation-invariant. What is wrong with the argument if instead we forbid the pattern $x + y = z$? Note that $\{x \in \mathbb{F}_3^n : x_1 = 2\}$ avoids solutions to $x + y = z$, and this set has density $1/3$.

Step 3. Iterate the density increment

We start with a 3-AP-free $A \subset \mathbb{F}_3^n$. Let $V_0 := \mathbb{F}_3^n$ with density $\alpha_0 := \alpha = |A|/3^n$. Repeatedly apply Lemma 6.2.8. After i rounds, we restrict A to a codimension i affine subspace V_i (with $V_0 \supset V_1 \supset \cdots$). Let $\alpha_i = |A \cap V_i|/|V_i|$ be the density of A in V_i. As long as $2\alpha_i^{-2} \le |V_i| = 3^{n-i}$, we can apply Lemma 6.2.8 to obtain a V_{i+1} with density increment

$$\alpha_{i+1} \ge \alpha_i + \alpha_i^2/4.$$

Since $\alpha = \alpha_0 \le \alpha_1 \le \cdots \le 1$, and α_i increases by $\ge \alpha_i^2/4 \ge \alpha^2/4$ at each step, the process terminates after $m \le 4/\alpha^2$ rounds, at which point we must have $3^{n-m} < 2\alpha_m^{-2} \le 2\alpha^{-2}$ (or else we can continue via Lemma 6.2.8). So $n < m + \log_3(2\alpha^{-2}) = O(1/\alpha^2)$, i.e., $\alpha \le 1/\sqrt{n}$. This is just shy of the bound $\alpha = O(1/n)$ that we aim to prove. So let us re-do the density increment analysis more carefully to analyze how quickly α_i grows.

Each round, α_i increases by at least $\alpha^2/4$. So it takes $\leq \lceil 4/\alpha \rceil$ initial rounds for α_i to double. Once $\alpha_i \geq 2\alpha$, it then increases by at least $\alpha_i^2/4$ each round afterwards, so it takes $\leq \lceil 1/\alpha_i \rceil \leq \lceil 1/\alpha \rceil$ additional round for the density to double again. And so on: the k-th doubling time is at most $\lceil 4^{2-k}/\alpha \rceil$. Since the density is always at most α, the density can double at most $\log_2(1/\alpha)$ times. So the total number of rounds is at most

$$\sum_{j \leq \log_2(1/\alpha)} \left\lceil \frac{4^{2-j}}{\alpha} \right\rceil = O\left(\frac{1}{\alpha}\right).$$

Suppose the process terminates after m steps with density α_m. Then, examining the hypothesis of Lemma 6.2.8, we find that the size of the final subspace $|V_m| = 3^{n-m}$ is less than $\alpha_m^{-2} \leq \alpha^{-2}$. So $n \leq m + O(\log(1/\alpha)) \leq O(1/\alpha)$. Thus $\alpha = |A|/N = O(1/n)$. This completes the proof of Roth's theorem in $\mathbb{F}_3^n$ (Theorem 6.2.1).

Remark 6.2.10 (Quantitative bounds)**.** The best published lower bound on the size of a cap set is $\geq 2.21^n$ (Edel 2004). This is obtained by constructing a cap set in $\mathbb{F}_3^{480}$ of size $m = 2^{327}(2^{73} + 3^7 7^{76}) \geq 2.21^{480}$, which then implies, by a product construction, a cap set in $\mathbb{F}_3^{480k}$ of size m^k for each positive integer k.

It was an open problem of great interest whether an upper bound of the form c^n, with constant $c < 3$, was possible on the size of cap sets in $\mathbb{F}_3^n$. With significant effort, the Fourier analytic strategy above was extended to prove an upper bound of the form $3^n/n^{1+c}$ (Bateman and Katz 2012). So it came as quite a shock to the community when a very short polynomial method proof was discovered, giving an upper bound $O(2.76^n)$ (Croot, Lev, and Pach 2017; Ellenberg and Gijswijt 2017). We will discuss this proof in Section 6.5. However, the polynomial method proof appears to be specific to the finite field model, and it is not known how to extend it to the integers.

The following exercise shows why the above strategy does not generalize to 4-APs at least in a straightforward manner.

Exercise 6.2.11 (Fourier uniformity does not control 4-AP counts)**.** Let

$$A = \{x \in \mathbb{F}_5^n : x \cdot x = 0\}.$$

Prove that:

(a) $|A| = (5^{-1} + o(1))5^n$ and $|\widehat{1_A}(r)| = o(1)$ for all $r \neq 0$;
(b) $|\{(x, y) \in \mathbb{F}_5^n : x, x + y, x + 2y, x + 3y \in A\}| \neq (5^{-4} + o(1))5^{2n}$.

Hint: First write 1_A as an exponential sum. Compare with the Gauss sum from Theorem 3.3.14.

Exercise 6.2.12 (Linearity testing)**.** Show that for every prime p there is some $C_p > 0$ such that if $f\colon \mathbb{F}_p^n \to \mathbb{F}_p$ satisfies

$$\mathbb{P}_{x,y\in\mathbb{F}_p^n}(f(x)+f(y)=f(x+y)) = 1-\epsilon$$

then there exists some $a \in \mathbb{F}_p^n$ such that

$$\mathbb{P}_{x\in\mathbb{F}_p^n}(f(x)=a\cdot x) \geq 1-C_p\epsilon.$$

In the above $\mathbb{P}$ expressions x and y are chosen i.i.d. uniform from $\mathbb{F}_p^n$.

The following exercises introduce Gowers uniformity norms. Gowers (2001) used them to prove Szemerédi's theorem by extending the Fourier analytic proof strategy of Roth's theorem to what is now called **higher order Fourier analysis**.

The U^2 norm in the following exercise plays a role similar to Fourier analysis.

Exercise 6.2.13 (Gowers U^2 uniformity norm)**.** Let $f\colon \mathbb{F}_p^n \to \mathbb{C}$, define

$$\|f\|_{U^2} := \left(\mathbb{E}_{x,y,y'\in\mathbb{F}_p^n} f(x)\overline{f(x+y)f(x+y')}f(x+y+y')\right)^{1/4}.$$

(a) Show that the expectation above is always a nonnegative real number, so that the above expression is well defined. Also, show that $\|f\|_{U^2} \geq |\mathbb{E}f|$.

(b) (Gowers Cauchy–Schwarz) For $f_1, f_2, f_3, f_4\colon \mathbb{F}_p^n \to \mathbb{C}$, let

$$\langle f_1, f_2, f_3, f_4\rangle = \mathbb{E}_{x,y,y'\in\mathbb{F}_p^n} f_1(x)\overline{f_2(x+y)f_3(x+y')}f_4(x+y+y').$$

Prove that

$$|\langle f_1, f_2, f_3, f_4\rangle| \leq \|f_1\|_{U^2}\,\|f_2\|_{U^2}\,\|f_3\|_{U^2}\,\|f_4\|_{U^2}$$

(c) (Triangle inequality) Show that

$$\|f+g\|_{U^2} \leq \|f\|_{U^2} + \|g\|_{U^2}.$$

Conclude that $\|\ \|_{U^2}$ is a norm.

Hint: Note that $\langle f_1, f_2, f_3, f_4\rangle$ is multilinear. Apply (b).

(d) (Relation with Fourier) Show that

$$\|f\|_{U^2} = \|\widehat{f}\|_{\ell^4}.$$

Furthermore, deduce that if $\|f\|_\infty \leq 1$, then

$$\|\widehat{f}\|_\infty \leq \|f\|_{U^2} \leq \|\widehat{f}\|_\infty^{1/2}.$$

(The second inequality gives a so-called "inverse theorem" for the U^2 norm: if $\|f\|_{U^2} \geq \delta$ then $|\widehat{f}(r)| \geq \delta^2$ for some $r \in \mathbb{F}_p^n$, i.e., if f is not U^2-uniform, then it must correlate with some function of the form $x \mapsto \omega^{r\cdot x}$.)

The inadequacy of Fourier analysis towards understanding 4-APs is remedied by the U^3 norm, which is significantly more mysterious than the U^2 norm. Some easier properties of the U^3 norm are given in the exercise below. Understanding properties of functions with large U^3 norm (known as the inverse problem) lies at the heart of **quadratic Fourier analysis**, which we do not discuss in this book (see Further Reading). The structure of set addition, which is the topic of the next chapter, plays a central role in this theory.

Exercise 6.2.14 (Gowers U^3 uniformity norm)**.** Let $f\colon \mathbb{F}_p^n \to \mathbb{C}$. Define

$$\|f\|_{U^3} := \Big(\mathbb{E}_{x,y_1,y_2,y_3} f(x)\overline{f(x+y_1)f(x+y_2)f(x+y_3)}\cdots$$
$$\cdot f(x+y_1+y_2)f(x+y_1+y_3)f(x+y_2+y_3)\overline{f(x+y_1+y_2+y_3)}\Big)^{1/8}.$$

Alternatively, for each $y \in \mathbb{F}_p^n$, define the multiplicative finite difference $\Delta_y f\colon \mathbb{F}_p^n \to \mathbb{C}$ by $\Delta_y f(x) := f(x)\overline{f(x+y)}$, we can rewrite the above expression in terms of the U^2 uniformity norm from Exercise 6.2.13 as

$$\|f\|_{U^3}^8 = \mathbb{E}_{y\in\mathbb{F}_p^n} \left\|\Delta_y f\right\|_{U^2}^4.$$

(a) (Monotonicity of U^k norms) Verify that the above two definitions for $\|f\|_{U^3}$ coincides and give well defined nonnegative real numbers. Also, show that

$$\|f\|_{U^2} \le \|f\|_{U^3}.$$

(b) (Separation of norms) Let p be odd and $f\colon \mathbb{F}_p^n \to \mathbb{C}$ be defined by $f(x) = e^{2\pi i x\cdot x/p}$. Prove that $\|f\|_{U^3} = 1$ and $\|f\|_{U^2} = p^{-n/4}$.

(c) (Triangle inequality) Prove that

$$\|f+g\|_{U^3} \le \|f\|_{U^3} + \|g\|_{U^3}.$$

Conclude that $\|\ \|_{U^3}$ is a norm.

(d) (U^3 norm controls 4-APs) Let $p \ge 5$ be a prime, and $f_1, f_2, f_3, f_4\colon \mathbb{F}_p^n \to \mathbb{C}$ all taking values in the unit disk. We write

$$\Lambda(f_1,f_2,f_3,f_4) := \mathbb{E}_{x,y\in\mathbb{F}_p^n} f_1(x)f_2(x+y)f_3(x+2y)f_4(x+3y).$$

Prove that

$$|\Lambda(f_1,f_2,f_3,f_4)| \le \min_s \|f_s\|_{U^3}.$$

Furthermore, deduce that if $f, g\colon \mathbb{F}_p^n \to [0,1]$, then

$$|\Lambda(f,f,f,f) - \Lambda(g,g,g,g)| \le 4\,\|f-g\|_{U^3}.$$

Hint: Re-parameterize as in Section 2.10 and then repeatedly apply Cauchy–Schwarz.

6.3 Fourier Analysis in the Integers

Now we review the basic notions of Fourier analysis on the integers. In the next section, we adapt the proof of Roth's theorem from $\mathbb{F}_3^n$ to $\mathbb{Z}$. The notions that we introduce below are better known as **Fourier series**.

Here $\mathbb{R}/\mathbb{Z}$ is the set of reals mod 1. A function $f\colon \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ is the same as a function $f\colon \mathbb{R} \to \mathbb{C}$ that is periodic mod 1, i.e., $f(x+1) = f(x)$ for all $x \in \mathbb{R}$.

Definition 6.3.1 (Fourier transform in $\mathbb{Z}$)
Given a finitely supported $f\colon \mathbb{Z} \to \mathbb{C}$, define $\widehat{f}\colon \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ by setting, for all $\theta \in \mathbb{R}$,

$$\widehat{f}(\theta) := \sum_{x\in\mathbb{Z}} f(x)e(-x\theta),$$

where

$$e(t) := \exp(2\pi i t), \qquad t \in \mathbb{R}.$$

Note that $\widehat{f}(\theta) = \widehat{f}(\theta + n)$ for all integers n. So $\widehat{f}\colon \mathbb{R}/\mathbb{Z} \to \mathbb{C}$ is well defined.

The various identities in Section 6.1 have counterparts stated below. We leave the proofs as exercises for the reader.

Theorem 6.3.2 (Fourier inversion formula)
Given a finitely supported $f\colon \mathbb{Z} \to \mathbb{C}$, for any $x \in \mathbb{Z}$,

$$f(x) = \int_0^1 \widehat{f}(\theta)e(x\theta)\,d\theta.$$

Theorem 6.3.3 (Parseval's identity)
Given finitely supported $f, g\colon \mathbb{Z} \to \mathbb{C}$,

$$\sum_{x\in\mathbb{Z}} \overline{f(x)}g(x) = \int_0^1 \overline{\widehat{f}(\theta)}\widehat{g}(\theta)\,d\theta$$

In particular, as a special case ($f = g$),

$$\sum_{x\in\mathbb{Z}} |f(x)|^2 = \int_0^1 |\widehat{f}(\theta)|^2\,d\theta$$

Note the normalization conventions: we sum in the physical space $\mathbb{Z}$ (there is no sensible way to average in $\mathbb{Z}$) and average in the frequency space $\mathbb{R}/\mathbb{Z}$.

Definition 6.3.4 (Convolution)

Given finitely supported $f, g\colon \mathbb{Z} \to \mathbb{C}$, define $f * g\colon \mathbb{Z} \to \mathbb{C}$ by

$$(f * g)(x) := \sum_{y \in \mathbb{Z}} f(y) g(x - y).$$

Theorem 6.3.5 (Convolution identity)

Given finitely supported $f, g\colon \mathbb{Z} \to \mathbb{C}$, for any $\theta \in \mathbb{R}/\mathbb{Z}$,

$$\widehat{f * g}(\theta) = \widehat{f}(\theta)\widehat{g}(\theta).$$

Given finitely supported $f, g, h\colon \mathbb{Z} \to \mathbb{C}$, define

$$\Lambda(f, g, h) := \sum_{x, y \in \mathbb{Z}} f(x) g(x + y) h(x + 2y)$$

and

$$\Lambda_3(f) := \Lambda(f, f, f).$$

Then for any finite set A of integers,

$$\Lambda_3(A) = |\{(x, y) : x, x + y, x + 2y \in A\}|$$

counts the number of 3-APs in A, where each non-trivial 3-AP is counted twice, forward and backward, and each trivial 3-AP is counted once.

Proposition 6.3.6 (Fourier and 3-AP)

Given finitely supported $f, g, h\colon \mathbb{Z} \to \mathbb{C}$,

$$\Lambda(f, g, h) = \int_0^1 \widehat{f}(\theta)\widehat{g}(-2\theta)\widehat{h}(\theta)\, d\theta.$$

Exercise 6.3.7. Prove all the identities above.

Exercise 6.3.8 (Counting solutions to a single linear equation)**.** Let $c_1, \dots, c_k \in \mathbb{Z}$. Let $A \subset \mathbb{Z}$ be a finite set. Show that

$$|\{(a_1, \dots, a_k) \in A^k : c_1 a_1 + \cdots + c_k a_k = 0\}| = \int_0^1 \widehat{1_A}(c_1 t)\widehat{1_A}(c_2 t) \cdots \widehat{1_A}(c_k t)\, dt.$$

Exercise 6.3.9. Show that if a finite set A of integers contains $\beta |A|^2$ solutions $(a, b, c) \in A^3$ to $a + 2b = 3c$, then it contains at least $\beta^2 |A|^3$ solutions $(a, b, c, d) \in A^4$ to $a + b = c + d$.

6.4 Roth's Theorem in the Integers

In Section 6.2 we saw a Fourier analytic proof of Roth's theorem in $\mathbb{F}_3^n$. In this section, we adapt the proof to the integers and obtain the following result. This is Roth's original proof (1953).

Theorem 6.4.1 (Roth's theorem)
Every 3-AP-free subset of $[N] = \{1, \dots, N\}$ has size $O(N/\log\log N)$.

The proof of Roth's theorem in $\mathbb{F}_3^n$ proceeded by density increment when restricting to subspaces. An important difference between $\mathbb{F}_3^n$ and $\mathbb{Z}$ is that $\mathbb{Z}$ has no subspaces (more on this later). Instead, we will proceed in $\mathbb{Z}$ by restricting to *subprogressions*. In this section, by a **progression** we mean an arithmetic progression.

We have the following analog of Lemma 6.2.4. It says that if f and g are "Fourier-close,", then they have similar 3-AP counts. We write

$$\|\widehat{f}\|_\infty := \sup_\theta |\widehat{f}(\theta)| \qquad \text{and} \qquad \|f\|_{\ell^2} := \left(\sum_{x\in\mathbb{Z}} |f(x)|^2\right)^{1/2}.$$

Proposition 6.4.2 (3-AP counting lemma)
Let $f, g : \mathbb{Z} \to \mathbb{C}$ be finitely supported functions. Then

$$|\Lambda_3(f) - \Lambda_3(g)| \le 3\|\widehat{f-g}\|_\infty \max\left\{\|f\|_{\ell^2}^2, \|g\|_{\ell^2}^2\right\}.$$

Proof. We have

$$\Lambda_3(f) - \Lambda_3(g) = \Lambda(f-g, f, f) + \Lambda(g, f-g, f) + \Lambda(g, g, f-g).$$

Let us bound the first term on the right-hand side. We have

$$\begin{aligned}
&|\Lambda(f-g, f, f)| \\
&\quad = \left|\int_0^1 \widehat{(f-g)}(\theta)\widehat{f}(-2\theta)\widehat{f}(\theta)\, d\theta\right| && \text{[Prop. 6.3.6]} \\
&\quad \le \|\widehat{f-g}\|_\infty \left|\int_0^1 \widehat{f}(-2\theta)\widehat{f}(\theta)\, d\theta\right| && \text{[Triangle ineq.]} \\
&\quad \le \|\widehat{f-g}\|_\infty \left(\int_0^1 \left|\widehat{f}(-2\theta)\right|^2 d\theta\right)^{1/2} \left(\int_0^1 \left|\widehat{f}(\theta)\right|^2 d\theta\right)^{1/2} && \text{[Cauchy-Schwarz]} \\
&\quad \le \|\widehat{f-g}\|_\infty \|f\|_{\ell^2}^2. && \text{[Parseval]}
\end{aligned}$$

By similar arguments, we have

$$|\Lambda(g, f-g, f)| \le \|\widehat{f-g}\|_\infty \, \|f\|_{\ell^2} \, \|g\|_{\ell^2}$$

and

$$|\Lambda(g, g, f-g)| \le \|\widehat{f-g}\|_\infty \, \|g\|_{\ell^2}^2 \, .$$

Combining with the first sum gives the result. □

Now we prove Roth's theorem by following the same steps as in Section 6.2 for the finite field setting.

Step 1. A 3-AP-free set has a large Fourier coefficient

Instead of directly studying the Fourier coefficients of 1_A (which is not a good idea since $\widehat{1_A}(\theta) \approx |A|$ is always large whenever $\theta \approx 0$), we apply a useful and standard trick and study the Fourier coefficients of the de-meaned function

$$1_A - \alpha 1_{[N]}.$$

This function has sum zero, and so its Fourier transform is zero at zero, which allows us to focus on the interesting values away from zero. Subtracting by $\alpha 1_{[N]}$ here has the same effect as considering $\widehat{1_A}(r)$ only for nonzero r in the finite field model.

Lemma 6.4.3 (3-AP-free implies large Fourier)

Let $A \subset [N]$ be a 3-AP free set with $|A| = \alpha N$. If $N \ge 5\alpha^{-2}$, then there exists $\theta \in \mathbb{R}/\mathbb{Z}$ satisfying

$$\left| \sum_{x=1}^{N} (1_A - \alpha)(x) e(\theta x) \right| \ge \frac{\alpha^2}{10} N.$$

Proof. Since A is 3-AP-free, the quantity $1_A(x)1_A(x+y)1_A(x+2y)$ is nonzero only for trivial APs, i.e. when $y = 0$. Thus

$$\Lambda_3(1_A) = |A| = \alpha N.$$

On the other hand, a 3-AP in $[N]$ can be counted by counting pairs of integers with the same parity to form the first and third element of the 3-AP, yielding,

$$\Lambda_3(1_{[N]}) = \lfloor N/2 \rfloor^2 + \lceil N/2 \rceil^2 \ge N^2/2.$$

Now apply the counting lemma (Proposition 6.4.2) to $f = 1_A$ and $g = \alpha 1_{[N]}$. We have $\|1_A\|_{\ell^2}^2 = |A| = \alpha N$ and $\|\alpha 1_{[N]}\|_{\ell^2}^2 = \alpha^2 N$. So

$$\frac{\alpha^3 N^2}{2} - \alpha N \le \alpha^3 \Lambda_3(1_{[N]}) - \Lambda_3(1_A) \le 3\alpha N \left\| (1_A - \alpha 1_{[N]})^\wedge \right\|_\infty .$$

Thus, using $N \geq 5/\alpha^2$, we have (the final step uses $N \geq 5\alpha^{-2}$)

$$\left\|(1_A - \alpha 1_{[N]})^\wedge\right\|_\infty \geq \frac{\frac{1}{2}\alpha^3 N^2 - \alpha N}{3\alpha N} = \frac{1}{6}\alpha^2 N - \frac{1}{3} \geq \frac{1}{10}\alpha^2 N.$$

Therefore there exists some $\theta \in \mathbb{R}$ with

$$\left|\sum_{x=1}^{N}(1_A - \alpha)(x)e(\theta x)\right| = (1_A - \alpha 1_{[N]})^\wedge(\theta) \geq \frac{1}{10}\alpha^2 N. \qquad \square$$

Step 2. A large Fourier coefficient implies density increment on a subprogression

In the finite field model, if $\widehat{1_A}(r)$ is large for some $r \in \mathbb{F}_3^n \setminus \{0\}$, then we obtained a density increment by restricting A to some coset of the hyperplane $r^\perp$.

How can we adapt this argument in the integers?

In the finite field model, we used that the Fourier character $\gamma_r(x) = \omega^{r \cdot x}$ is constant on each coset of the hyperplane $r^\perp \subset \mathbb{F}_3^n$. In the integer setting, we want to partition $[N]$ into subprogressions such that the character $\mathbb{Z} \to \mathbb{C} : x \mapsto e(x\theta)$ is roughly constant on each subprogression. As a simple example, assume that θ is a rational a/b for some fairly small b. Then $x \mapsto e(x\theta)$ is constant on arithmetic progressions with common difference b. Thus we could partition $[N]$ into arithmetic progressions with common difference b. This is useful as long as b is not too large. On the other hand, if b is too large, or if θ is irrational, then we would want to approximate θ be a rational number with small denominator.

We write

$$\|\theta\|_{\mathbb{R}/\mathbb{Z}} := \text{distance from } \theta \text{ to the nearest integer.}$$

Lemma 6.4.4 (Dirichlet's lemma)

Let $\theta \in \mathbb{R}$ and $0 < \delta < 1$. Then there exists a positive integer $d \leq 1/\delta$ such that $\|d\theta\|_{\mathbb{R}/\mathbb{Z}} \leq \delta$.

Proof. Let $m = \lfloor 1/\delta \rfloor$. By the pigeonhole principle, among the $m+1$ numbers $0, \theta, \cdots, m\theta$, we can find $0 \leq i < j \leq m$ such that the fractional parts of $i\theta$ and $j\theta$ differ by at most δ. Set $d = |i - j|$. Then $\|d\theta\|_{\mathbb{R}/\mathbb{Z}} \leq \delta$, as desired. □

Given θ, we now partition $[N]$ into subprogressions with roughly constant $e(x\theta)$ inside each progression. The constants appearing in rest of this argument are mostly unimportant.

Lemma 6.4.5 (Partition into progression level sets)

Let $0 < \eta < 1$ and $\theta \in \mathbb{R}$. Suppose $N \geq (4\pi/\eta)^6$. Then one can partition $[N]$ into subprogressions P_i, each with length

$$N^{1/3} \leq |P_i| \leq 2N^{1/3},$$

such that

$$\sup_{x,y \in P_i} |e(x\theta) - e(y\theta)| < \eta, \quad \text{for each } i.$$

Proof. By Lemma 6.4.4, there is a positive integer $d < \sqrt{N}$ such that $\|d\theta\|_{\mathbb{R}/\mathbb{Z}} \leq 1/\sqrt{N}$. Partition $[N]$ greedily into progressions with common difference d of lengths between $N^{1/3}$ and $2N^{1/3}$. Then, for two elements x, y within the same progression P_i, we have

$$|e(x\theta) - e(y\theta)| \leq |P_i|\, |e(d\theta) - 1| \leq 2N^{1/3} \cdot 2\pi \cdot N^{-1/2} \leq \eta.$$

Here we use the inequality $|e(d\theta) - 1| \leq 2\pi \, \|d\theta\|_{\mathbb{R}/\mathbb{Z}}$ from the fact that the length of a chord on a circle is at most the length of the corresponding arc. □

We can now apply this lemma to obtain a density increment.

Lemma 6.4.6 (3-AP-free implies density increment)

Let $A \subset [N]$ be 3-AP-free, with $|A| = \alpha N$ and $N \geq (16/\alpha)^{12}$. Then there exists a subprogression $P \subset [N]$ with $|P| \geq N^{1/3}$ and $|A \cap P| \geq (\alpha + \alpha^2/40)\,|P|$.

Proof. By Lemma 6.4.3, there exists θ satisfying

$$\left| \sum_{x=1}^{N} (1_A - \alpha)(x) e(x\theta) \right| \geq \frac{\alpha^2}{10} N.$$

Next, apply Lemma 6.4.5 with $\eta = \alpha^2/20$ (the hypothesis $N \geq (4\pi/\eta)^6$ is satisfied since $(16/\alpha)^{12} \geq (80\pi/\alpha^2)^6 = (4\pi/\eta)^6$) to obtain a partition $P_1, \ldots, P_k$ of $[N]$ satisfying $N^{1/3} \leq |P_i| \leq 2N^{1/3}$ and

$$|e(x\theta) - e(y\theta)| \leq \frac{\alpha^2}{20} \quad \text{for all } i \text{ and } x, y \in P_i.$$

So on each P_i,

$$\left| \sum_{x \in P_i} (1_A - \alpha)(x) e(x\theta) \right| \leq \left| \sum_{x \in P_i} (1_A - \alpha)(x) \right| + \frac{\alpha^2}{20} |P_i|.$$

Thus

$$\begin{aligned}\frac{\alpha^2}{10}N &\le \left|\sum_{x=1}^{N}(1_A-\alpha)(x)e(x\theta)\right| \\ &\le \sum_{i=1}^{k}\left|\sum_{x\in P_i}(1_A-\alpha)(x)e(x\theta)\right|. \\ &\le \sum_{i=1}^{k}\left(\left|\sum_{x\in P_i}(1_A-\alpha)(x)\right| + \frac{\alpha^2}{20}|P_i|\right) \\ &= \sum_{i=1}^{k}\left|\sum_{x\in P_i}(1_A-\alpha)(x)\right| + \frac{\alpha^2}{20}N\end{aligned}$$

Thus

$$\frac{\alpha^2}{20}N \le \sum_{i=1}^{k}\left|\sum_{x\in P_i}(1_A-\alpha)(x)\right|$$

and hence

$$\frac{\alpha^2}{20}\sum_{i=1}^{k}|P_i| \le \sum_{i=1}^{k}\big||A\cap P_i|-\alpha|P_i|\big|.$$

We want to show that there exists some P_i such that A has a density increment when restricted to P_i. The following trick is convenient. Note that

$$\begin{aligned}\frac{\alpha^2}{20}\sum_{i=1}^{k}|P_i| &\le \sum_{i=1}^{k}\big||A\cap P_i|-\alpha|P_i|\big| \\ &= \sum_{i=1}^{k}\left(\big||A\cap P_i|-\alpha|P_i|\big| + (|A\cap P_i|-\alpha|P_i|)\right),\end{aligned}$$

as the newly added terms in the final step sum to zero. Thus there exists an i such that

$$\frac{\alpha^2}{20}|P_i| \le \big||A\cap P_i|-\alpha|P_i|\big| + (|A\cap P_i|-\alpha|P_i|).$$

Since $|t|+t$ is $2t$ for $t>0$ and 0 for $t\le 0$, we deduce

$$\frac{\alpha^2}{20}|P_i| \le 2(|A\cap P_i|-\alpha|P_i|),$$

which yields

$$|A\cap P_i| \ge \left(\alpha+\frac{\alpha^2}{40}\right)|P_i|. \qquad \square$$

By translation and rescaling, we can identify P with $[N']$ with $N' = |P|$. Then $A \cap P$ becomes a subset $A' \subset [N']$. Note that A' is 3-AP-free (here we are invoking the important fact that 3-APs are translation and dilation invariant). We can now iterate the argument. (Think about where the argument goes wrong for patterns such as $\{x, y, x+y\}$ and $\{x, x+y, x+y^2\}$.)

Step 3. Iterate the density increment

This step is nearly identical to the proof in the finite field model. Start with $\alpha_0 = \alpha$ and $N_0 = N$. After i iterations, we arrive at a subprogression of length N_i where A has density α_i. As long as $N_i \ge (16/\alpha_i)^{12}$, we can apply apply Lemma 6.4.6 to pass down to a subprogression with

$$N_{i+1} \ge N_i^{1/3} \quad \text{and} \quad \alpha_{i+1} \ge \alpha_i + \alpha_i^2/40.$$

We double α_i from α_0 after $\le \lceil 40/\alpha \rceil$ iterations. Once the density reaches at least 2α, the next doubling takes $\le \lceil 20/\alpha \rceil$ iterations, and so on. In general, the k-th doubling requires $\le \lceil 40 \cdot 2^{-k}/\alpha \rceil$ iterations. There are at most $\log_2(1/\alpha)$ doublings since the density is always at most 1. Summing up, the total number of iterations is

$$m \le \sum_{i=1}^{\log_2(1/\alpha)} \left\lceil 40 \cdot 2^{-k}/\alpha \right\rceil = O(1/\alpha).$$

When the process terminates, we must have $N^{1/2^m} \le N_m$ by Lemma 6.4.6. So

$$N^{1/2^m} \le N_m < (16/\alpha_i)^{12} \le (16/\alpha)^{12}.$$

So

$$N \le (16/\alpha)^{12 \cdot 2^m} \le (16/\alpha)^{2^{O(1/\alpha)}}.$$

Therefore

$$\frac{|A|}{N} = \alpha = O\left(\frac{1}{\log\log N}\right).$$

This completes the proof of Roth's theorem (Theorem 6.4.1). □

We saw that the proofs in $\mathbb{F}_3^n$ and $\mathbb{Z}$ have largely the same set of ideas, but the proof in $\mathbb{Z}$ is somewhat more technically involved. The finite field model is often a good sandbox to try out Fourier analytic ideas.

Remark 6.4.7 (Bohr sets)**.** Let us compare the results in $\mathbb{F}_3^n$ and $[N]$. Write $N = 3^n$ for the size of the ambient space in both cases, for comparison. We obtained an upper bound of $O(N/\log N)$ for 3-AP-free sets in $\mathbb{F}_3^n$ and $O(N/\log\log N)$ in $[N] \subset \mathbb{Z}$. Where does the difference in quantitative bounds stem from?

In the density increment step for $\mathbb{F}_3^n$, at each step, we pass down to a subset which had size a constant factor (namely $1/3$) of the original one. However, in $[N]$, each iteration gives us a subprogression which has size equal to the cube root of the previous subprogression. The extra log for Roth's theorem in the integers comes from this rapid reduction in the sizes of the subprogressions.

Can we do better? Perhaps by passing down to subsets of $[N]$ that look more like subspaces? Indeed, this possible. Bourgain (1999) used **Bohr sets** to prove an improved bound of $N/(\log N)^{1/2+o(1)}$ on Roth's theorem. Given $\theta_1, \dots, \theta_k$, and some $\epsilon > 0$, a Bohr set has the form

$$\left\{x \in [N] : \|x\theta_j\|_{\mathbb{R}/\mathbb{Z}} \le \epsilon \text{ for each } j = 1, \dots, k\right\}.$$

To see why this is analogous to subspaces, note that we can define a subspace of $\mathbb{F}_3^n$ as a set of the following form

$$\left\{x \in \mathbb{F}_3^n : r_j \cdot x = 0 \text{ for each } j = 1, \dots, k\right\}.$$

where $r_1, \dots, r_k \in \mathbb{F}_3^n \setminus \{0\}$. Bohr sets are used widely in additive combinatorics, and in nearly all subsequent work on Roth's theorem in the integers, including the proof of the current best bound $N/(\log N)^{1+c}$ for some constant $c > 0$ (Bloom and Sisask 2020).

We will see Bohr sets again in the proof of Freiman's theorem in Chapter 7.

The next exercise is analogous to Exercise 6.2.11, which was in $\mathbb{F}_5^n$.

Exercise 6.4.8* (Fourier uniformity does not control 4-AP counts)**.** Fix $0 < \alpha < 1$. Let N be a prime. Let

$$A = \left\{x \in [N] : x^2 \bmod N < \alpha N\right\}.$$

Viewing $A \subset \mathbb{Z}/N\mathbb{Z}$, prove that, as $N \to \infty$ with fixed α,

(a) $|A| = (\alpha + o(1))N$ and $\max_{r \neq 0} |\widehat{1_A}(r)| = o(1)$;

(b) $|(x, y) \in \mathbb{Z}/N\mathbb{Z} : x, x + y, x + 2y, x + 3y \in A| \neq (\alpha^4 + o(1))N^2$.

6.5 Polynomial Method

An important breakthrough of Croot, Lev, and Pach (2017) showed how to apply the **polynomial method** to Roth-type problems in the finite field model. Their method quickly found many applications. Less than a week after the Croot, Lev, and Pach paper was made public, Ellenberg and Gijswijt (2017) adapted their argument to prove the following bound on the cap set problem. The discovery came as quite a shock to the community, especially as the proof is so short.

Theorem 6.5.1 (Cap set upper bound)
Every 3-AP-free subset of $\mathbb{F}_3^n$ has size $O(2.76^n)$.

The presentation of the proof below is due to **?**.

Recall from linear algebra the usual **rank** of a matrix. Here we can view an $|A| \times |A|$ matrix over the field $\mathbb{F}$ as a function $F\colon A \times A \to \mathbb{F}$. A function F is said to have rank 1 if $F(x, y) = f(x)g(y)$ for some nonzero functions $f, g\colon A \to \mathbb{F}$. More generally, the rank of F is the minimum k so that F can be written as a sum of k rank 1 functions.

More generally, for other notions of rank, we can first define the set of rank 1 functions, and then define the rank of F to be the minimum k so that F can be written as a sum of k rank 1 functions.

Whereas a function $A \times A \to \mathbb{F}$ corresponds to a matrix, a function $A \times A \times A \to \mathbb{F}$ correspond to a 3-tensor. There is a notion of **tensor rank**, where the rank 1 functions are those of the form $F(x, y, z) = f(x)g(y)h(z)$. This is a standard and important notion (which comes with a lot of mystery), but it is not the one that we shall use.

Definition 6.5.2 (Slice rank)
A function $F\colon A \times A \times A \to \mathbb{F}$ is said to have **slice rank 1** if it can be written as

$$f(x)g(y,z), \quad f(y)g(x,z), \quad \text{or} \quad f(z)g(x,y),$$

for some nonzero functions $f\colon A \to \mathbb{F}$ and $g\colon A \times A \to \mathbb{F}$.

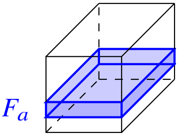
The **slice rank** of a function $F\colon A \times A \times A \to \mathbb{F}$ is the minimum k so that F can be written as a sum of k slice rank 1 functions.

Here is an easy fact about the slice rank.

Lemma 6.5.3 (Trivial upper bound for slice rank)
Every function $F\colon A \times A \times A \to \mathbb{F}$ has slice rank at most $|A|$.

Proof. Let F_a be the restriction of F to the “slice” $\{(x, y, z) \in A \times A \times A : x = a\}$, i.e.,

$$F_a(x,y,z) = \begin{cases} F(x,y,z) & \text{if } x = a, \\ 0 & \text{if} x \neq a. \end{cases}$$



Then F_a has slice rank ≤ 1 since $F_a(x, y, z) = \delta_a(x)F(a, y, z)$, where δ_a denotes the function taking value 1 at a and 0 elsewhere. Thus $F = \sum_{a \in A} F_a$ has slice rank at most $|A|$. □

For the next lemma, we need the following fact from linear algebra.

Lemma 6.5.4 (Vector with large support)

Every k-dimensional subspace of an n-dimensional vector space (over any field) contains a point with at least k nonzero coordinates.

Proof. Form a $k \times n$ matrix M whose rows form a basis of this k-dimensional subspace W. Then M has rank k. So it has some invertible $k \times k$ submatrix with columns $S \subset [n]$ with $|S| = k$. Then for every $z \in \mathbb{F}^S$, there is some linear combination of the rows whose coordinates on S are identical to those of z. In particular, there is some vector in the k-dimensional subspace W whose S-coordinates are all nonzero. □

A diagonal matrix with nonzero diagonal entries has full rank. We show that a similar statement holds true for the slice rank.

Lemma 6.5.5 (Slice rank of a diagonal)

Suppose $F\colon A \times A \times A \to \mathbb{F}$ satisfies $F(x, y, z) \neq 0$ if and only if $x = y = z$. Then F has slice rank $|A|$.

Proof. From Lemma 6.5.3, we already know that the slice rank of F is $\leq |A|$. It remains to prove that the slice rank of F is is $\geq |A|$.

Suppose $F(x, y, z)$ can be written as a sum of functions of the form

$$f(x)g(y,z), \quad f(y)g(x,z), \quad \text{and} \quad f(z)g(x,y),$$

with m_1 summands of the first type, m_2 of the second type, and m_3 of the third type. By Lemma 6.5.4, there is some function $h\colon A \to \mathbb{F}$ that is orthogonal to all the f's from the third type of summands (i.e., $\sum_{x\in A} f(x)h(x) = 0$), and such that $|\operatorname{supp} h| \geq |A| - m_3$. Let

$$G(x,y) = \sum_{z\in A} F(x,y,z)h(z).$$

Only summands of the first two types remain. Each summand of the first type turns into a rank 1 function (in the matrix sense of the rank)

$$(x,y) \mapsto \sum_z f(x)g(y,z)h(z) = f(x)\widetilde{g}(y)$$

for some new function $\widetilde{g}\colon A \to \mathbb{F}$. Similarly with functions of the second type. So G (viewed as an $|A| \times |A|$ matrix) has rank $\leq m_1 + m_2$. On the other hand,

$$G(x,y) = \begin{cases} h(x) & \text{if } x = y, \\ 0 & \text{if } x \neq y. \end{cases}$$

This G has rank $|\operatorname{supp} h| \geq |A| - m_3$. Combining, we get

$$|A| - m_3 \leq \operatorname{rank} G \leq m_1 + m_2.$$

So $m_1 + m_2 + m_3 \geq |A|$. This shows that the slice rank of F is $\geq |A|$. □

Now we prove an upper bound on the slice rank by invoking magical powers of polynomials.

Lemma 6.5.6 (Upper bound on the slice rank of $1_{x+y+z=0}$)
Define $F\colon A \times A \times A \to \mathbb{F}_3$ by

$$F(x,y,z) = \begin{cases} 1 & \text{if } x+y+z=0, \\ 0 & \text{otherwise.} \end{cases}$$

Then the slice rank of F is at most

$$3 \sum_{\substack{a,b,c \geq 0 \\ a+b+c=n \\ b+2c \leq 2n/3}} \frac{n!}{a!b!c!}.$$

Proof. In $\mathbb{F}_3$, one has

$$1 - x^2 = \begin{cases} 1 & \text{if } x = 0, \\ 0 & \text{if } x \neq 0. \end{cases}$$

So, writing $x = (x_1, \dots, x_n)$, $y = (y_1, \dots, y_n)$, and $z = (z_1, \dots, z_n)$, we have

$$F(x,y,z) = \prod_{i=1}^{n} (1 - (x_i + y_i + z_i)^2). \tag{6.5.1}$$

If we expand the right-hand side, we obtain a polynomial in $3n$ variables with degree $2n$. This is a sum of monomials, each of the form

$$x_1^{i_1} \cdots x_n^{i_n} y_1^{j_1} \cdots y_n^{j_n} z_1^{k_1} \cdots z_n^{k_n},$$

where $i_1, i_2, \dots, i_n, j_1, \dots, j_n, k_1, \dots, k_n \in \{0,1,2\}$. For each term, by the pigeonhole principle, at least one of $i_1 + \cdots + i_n$, $j_1 + \cdots + j_n$, $k_1 + \cdots + k_n$ is at most $2n/3$. So we can split these summands into three sets:

$$\begin{aligned}\prod_{i=1}^{n}(1-(x_i+y_i+z_i)^2) = &\sum_{i_1+\cdots+i_n\le\frac{2n}{3}} x_1^{i_1}\cdots x_n^{i_n} f_{i_1,\ldots,i_n}(y,z)\\ &+\sum_{j_1+\cdots+j_n\le\frac{2n}{3}} y_1^{j_1}\cdots y_n^{j_n} g_{j_1,\ldots,j_n}(x,z)\\ &+\sum_{k_1+\cdots+k_n\le\frac{2n}{3}} z_1^{k_1}\cdots z_n^{k_n} h_{k_1,\ldots,k_n}(x,y).\end{aligned}$$

Each summand has slice rank at most 1. The number of summands in the first sum is precisely the number of triples of nonnegative integers a, b, c with $a + b + c = n$ and $b + 2c \le 2n/3$ (a, b, c correspond to the numbers of i_*'s that are equal to $0, 1, 2$ respectively) . The lemma then follows. □

Here is a standard estimate. The proof is similar to that of the Chernoff bound.

Lemma 6.5.7 (A trinomial coefficient estimate)

For every positive integer n,

$$\sum_{\substack{a,b,c\ge 0\\ a+b+c=n\\ b+2c\le 2n/3}} \frac{n!}{a!b!c!} \le 2.76^n.$$

Proof. Let $x \in [0, 1]$. The sum equals to the coefficients of all the monomials x^k with $k \le 2n/3$ in the expansion of $(1+x+x^2)^n$. By deleting contributions x^k with $k > 2n/3$ and using $x^{2n/3} \le x^k$ whenever $k \le 2n/3$, we have

$$\sum_{\substack{a,b,c\ge 0\\ a+b+c=n\\ b+2c\le 2n/3}} \frac{n!}{a!b!c!} \le \frac{(1+x+x^2)^n}{x^{2n/3}}.$$

Setting $x = 0.6$ shows that the left-hand side sum is $\le (2.76)^n$. □

Remark 6.5.8. Taking the optimal value $x = (\sqrt{33}-1)/8 = 0.59307\ldots$ in the final step, we obtain $\le (2.75510\ldots)^n$. This is the true exponential asymptotics of the sum in Lemma 6.5.7. See, e.g., Sanov's theorem from large deviation theory. We have no idea how close this is to the optimal bound for the cap set problem. However, quite surprisingly, such bound is tight for a variant of the cap sets known as the tri-colored sum-free sets (Blasiak et al. 2017; Kleinberg et al. 2018).

Proof of Theorem 6.5.1. Let $A \subset \mathbb{F}_3^n$ be 3-AP-free. Define $F\colon A \times A \times A \to \mathbb{F}_3$ by

$$F(x,y,z) = \begin{cases} 1 & \text{if } x+y+z=0, \\ 0 & \text{otherwise.} \end{cases}$$

Since A is 3-AP-free, one has $F(x,y,z) = 1$ if and only if $x = y = z \in A$. By Lemma 6.5.5, F has slice rank $|A|$. On the other hand, by Lemmas 6.5.6 and 6.5.7, F has slice rank $\le 3(2.76)^n$. So $|A| \le 3(2.76)^n$. □

It is straightforward to extend the above proof from $\mathbb{F}_3$ to any other fixed $\mathbb{F}_p$, resulting:

Theorem 6.5.9 (Roth's theorem in the finite field model)

For every odd prime p, there is some $c_p < p$ so that every 3-AP-free subset of $\mathbb{F}_p^n$ has size at most $3c_p^n$.

It remains an intriguing open problem to extend the techniques to other settings.

Open problem 6.5.10 (Szemerédi's theorem in the finite field model)

Is there a constant $c < 5$ such that every 4-AP-free subset of $\mathbb{F}_5^n$ has size $O(c^n)$?

Open problem 6.5.11 (Corner-free theorem in the finite field model)

Is there a constant $c < 2$ such that every corner-free subset of $\mathbb{F}_2^n \times \mathbb{F}_2^n$ has size $O(c^{2n})$? Here a corner is a configuration of the form $\{(x,y),(x+d,y),(x,y+d)\}$.

Finally, the proof technique in this section seems specific to the finite field model. It is an intriguing open problem to apply the polynomial method for Roth's theorem in the integers. Due to the Behrend example (Section 2.5), we cannot expect power-saving bounds in the integers.

Exercise 6.5.12 (Tricolor sum-free set). Let $a_1, \dots, a_m, b_1, \dots, b_m, c_1, \dots, c_m \in \mathbb{F}_2^n$. Suppose that the equation $a_i + b_j + c_k = 0$ holds if and only if $i = j = k$. Show that there is some constant $c > 0$ such that $m \le (2-c)^n$ for all sufficiently large n.

Exercise 6.5.13 (Sunflower-free set). Three sets A, B, C form a **sunflower** if $A \cap B = B \cap C = A \cap B = A \cap B \cap C$. Prove that there exists some constant $c > 0$ such that if $\mathcal{F}$ is a collection of subsets of $[n]$ without a sunflower, then $|\mathcal{F}| \le (2-c)^n$ provided that n is sufficiently large.

6.6 Arithmetic Regularity

Here we develop an arithmetic analogue of Szemerédi's graph regularity lemma from Chapter 2. Just as the graph regularity method has powerful applications, so too does the arithmetic regularity lemma as well as the general strategy behind it.

First, we need a notion of what it means for a subset of $\mathbb{F}_p^n$ to be uniform, in a sense analogous to ϵ-regular pairs from the graph regularity lemma. We also saw the following notion in the Fourier analytic proof of Roth's theorem.

Definition 6.6.1 (Fourier uniformity)

We say that $A \subset \mathbb{F}_p^n$ is **ϵ-uniform** if $|\widehat{1_A}(r)| \le \epsilon$ for all $r \in \mathbb{F}_p^n \setminus \{0\}$.

The following exercises explains how Fourier uniformity is analogous to the discrepancy-type condition for ϵ-regular pairs in the graph regularity lemma.

Exercise 6.6.2 (Uniformity and discrepancy). Let $A \subset \mathbb{F}_p^n$ with $|A| = \alpha p^n$. Let **HyperplaneDISC**(η) denote the property that for every hyperplane W of $\mathbb{F}_p^n$,

$$\left| \frac{|A \cap W|}{|W|} - \alpha \right| \le \eta.$$

(a) Prove that if A satisfies **HyperplaneDISC**(ϵ), then A is ϵ-uniform.
(b) Prove that if A is ϵ-uniform, then it satisfies **HyperplaneDISC**$((p-1)\epsilon)$.

Definition 6.6.3 (Fourier uniformity on affine subspaces)

For an affine subspace W of $\mathbb{F}_p^n$ (i.e., the coset of a subspace), we say that A is **ϵ-uniform on W** if $A \cap W$ is ϵ-uniform when viewed as a subset of W.

Here is an arithmetic analogue of Szemerédi's graph regularity lemma that we saw in Chapter 2. It is due to Green (2005a).

Theorem 6.6.4 (Arithmetic regularity lemma)

For every $\epsilon > 0$ and prime p, there exists M so that for every $A \subset \mathbb{F}_p^n$, there is some subspace W of $\mathbb{F}_p^n$ with codimension at most M such that A is ϵ-uniform on all but at most ϵ-fraction of cosets of W.

The proof is very similar to the proof of the graph regularity lemma in Chapter 2. Each subspace W induces a partition of of the whole space $\mathbb{F}_p^n$ into W-cosets, and we keep track the energy (mean-squared density) of the partition. We show that if the conclusion of Theorem 6.6.4 does not hold for the current W, then we can replace W by a smaller subspace so that the energy increases significantly. Since the energy is always bounded between 0 and 1, there are at most a bounded number of iterations.

Definition 6.6.5 (Energy)

Given $A \subset \mathbb{F}_p^n$, and W a subspace of $\mathbb{F}_p^n$, we define the **energy** of W with respect to a fixed A to be

$$q_A(W) := \mathbb{E}_{x \in \mathbb{F}_p^n} \left[\frac{|A \cap (W+x)|^2}{|W|^2} \right].$$

Given a subspace W of $\mathbb{F}_p^n$. Define $\mu_W \colon \mathbb{F}_p^n \to \mathbb{R}$ by

$$\mu_W := \frac{p^n}{|W|} 1_W.$$

(One can regard μ_W as the uniform probability distribution on W; it is normalized so that $\mathbb{E}\mu_W = 1$.) Then,

$$(1_A * \mu_W)(x) = \frac{|A \cap (W + x)|}{|W|} \quad \text{for every } x \in \mathbb{F}_p^n.$$

We have (check!)

$$\widehat{\mu_W}(r) = \begin{cases} 1 & \text{if } r \in W^\perp, \\ 0 & \text{if } r \notin W^\perp. \end{cases}$$

So by the convolution identity (Theorem 6.1.7).

$$\widehat{1_A * \mu_W}(r) = \widehat{1_A}(r)\widehat{\mu_W}(r) = \begin{cases} \widehat{1_A}(r) & \text{if } r \in W^\perp, \\ 0 & \text{if } r \notin W^\perp. \end{cases} \tag{6.6.1}$$

To summarize, convolving by μ_W averages 1_A along cosets of W in the physical space, and filters $W^\perp$ in the Fourier space.

Energy interacts nicely with the Fourier transform. By Parseval's identity (Theorem 6.1.3), we have

$$q_A(W) = \|1_A * \mu_W\|_2^2 = \sum_{r \in \mathbb{F}_p^n} |\widehat{1_A * \mu_W}(r)|^2 = \sum_{r \in W^\perp} |\widehat{1_A}(r)|^2. \tag{6.6.2}$$

The next lemma is analogous to Lemma 2.1.9. It is an easy consequence of convexity. It also directly follows from (6.6.2).

Lemma 6.6.6 (Energy never decreases under refinement)
Let $A \subset \mathbb{F}_p^n$. For subspaces $U \le W \le \mathbb{F}_p^2$, we have $q_A(U) \ge q_A(W)$. □

The next lemma is analogous to the energy boost lemma for irregular pairs in the proof of graph regularity (Lemma 2.1.10).

Lemma 6.6.7 (Local energy increment)
If $A \subset \mathbb{F}_p^n$ is not ϵ-uniform, then there is some codimension-1 subspace W with $q_A(W) > (|A|/p^n)^2 + \epsilon^2$.

Proof. Suppose A is not ϵ-uniform. Then there is some $r \neq 0$ such that $|\widehat{1_A}(r)| > \epsilon$. Let $W = r^\perp$. Then by (6.6.2),

$$\begin{aligned} q_A(W) &= |\widehat{1_A}(0)|^2 + |\widehat{1_A}(r)|^2 + |\widehat{1_A}(2r)|^2 + \cdots + |\widehat{1_A}((p-1)r)|^2 \\ &\ge |\widehat{1_A}(0)|^2 + |\widehat{1_A}(r)|^2 > (|A|/p^n)^2 + \epsilon^2. \end{aligned}$$ □

By applying the above lemmas locally to each W-coset, we obtain the following global increment, analogous to Lemma 2.1.11

Lemma 6.6.8 (Global energy increment)
Let $A \subset \mathbb{F}_p^n$. Let W be a subspace of $\mathbb{F}_p^n$. Suppose that f is not ϵ-uniform on $> \epsilon$-fraction of W-cosets. Then there is some subspace U of W with $\operatorname{codim} U - \operatorname{codim} W \le p^{\operatorname{codim} W}$ such that

$$q_A(U) > q_A(W) + \epsilon^3.$$

Proof. By Lemma 6.6.7, for each coset W' of W on which f is not ϵ-uniform, we can find some $r \in \mathbb{F}_p^n \setminus W^\perp$ so that replacing W by its intersection with $r^\perp$ increases its energy on W' by more than ϵ^2. In other words,

$$q_{A\cap W'}(W' \cap r^\perp) > \frac{|A \cap W'|^2}{|W'|^2} + \epsilon^2.$$

Let R be a set of such r's, one for each W-coset on which f is not ϵ-uniform (allowing some r's to be chosen repeatedly).

Let $U = W \cap R^\perp$. Then $\operatorname{codim} U - \operatorname{codim} W \le |R| \le |\mathbb{F}^p/W| = p^{\operatorname{codim} W}$.

Applying the monotonicity of energy (Lemma 6.6.6) on each W-coset and using the observation in the first paragraph in this proof, we see the "local" energy of U is more than that of W on by $> \epsilon^2$ on each of the $> \epsilon$-fraction of W-cosets on which f is not ϵ-uniform, and is at least as great as that of W on each of the remaining W-cosets. There the energy increases by $> \epsilon^2$ when refining from W to U. □

Proof of the arithmetic regularity lemma (Theorem 6.6.4). Starting with $W_0 = \mathbb{F}_p^n$, we construct a sequence of subspaces $W_0 \ge W_1 \ge W_2 \ge \cdots$ where each at step, unless A is ϵ-uniform on all but $\le \epsilon$-fraction of W-cosets, then we apply Lemma 6.6.8 to find $W_{i+1} \le W_i$. The energy increases by $> \epsilon^3$ at each iteration, so there are $< \epsilon^{-3}$ iterations. We have $\operatorname{codim} W_{i+1} \le \operatorname{codim} W_i + p^{\operatorname{codim} W_i}$ at each i, so the final $W = W_m$ has codimension at most some function of p and ϵ (one can check that it is an exponential tower of p's of height $O(\epsilon^{-3})$). This W satisfies the desired properties. □

Remark 6.6.9 (Lower bound). Recall that Gowers (1997) showed that there exist graphs whose ϵ-regular partition requires at least $\operatorname{tower}(\Omega(\epsilon^{-c}))$ parts (Theorem 2.1.14). There is a similar tower-type lower bound for the arithmetic regularity lemma (Green 2005a; Hosseini, Lovett, Moshkovitz, and Shapira 2016).

Remark 6.6.10 (Abelian groups). Green (2005a) also established an arithmetic regularity lemma over arbitrary finite abelian groups. Instead of subspaces, one uses Bohr sets (see Remark 6.4.7).

Arithmetic regularity decomposition

Now let us give another arithmetic regularity result. It has the same spirit as the above regularity lemma, but phrased in terms of a decomposition rather than a partition. This perspective of regularity as decompositions, popularized by Tao, allows one to adapt the ideas of regularity to more general settings where we cannot neatly partition the underlying space into easily describable pieces. It is very useful and has many applications in additive combinatorics.

Theorem 6.6.11 (Arithmetic regularity decomposition)

For every sequence $\epsilon_0 \geq \epsilon_1 \geq \epsilon_2 \geq \cdots > 0$, there exists M so that every $f \colon \mathbb{F}_p^n \to [0,1]$ can be written as

$$f = f_{\mathrm{str}} + f_{\mathrm{psr}} + f_{\mathrm{sml}}$$

where

- (structured piece) $f_{\mathrm{str}} = f_W$ for some subspace W of codimension at most M;
- (pseudorandom piece) $\|\widehat{f_{\mathrm{psr}}}\|_\infty \leq \epsilon_{\operatorname{codim} W}$;
- (small piece) $\|f_{\mathrm{sml}}\|_2 \leq \epsilon_0$.

Remark 6.6.12. It is worth comparing Theorem 6.6.11 to the strong graph regularity lemma (Theorem 2.8.3). It is important that the uniformity requirement on the pseudorandom piece depends on the codim W.

In other more advanced applications, we would like f_{str} to come from some structured class of functions. For example, in higher order Fourier analysis, f_{str} is a nilsequence.

Proof. Let $k_0 = 0$ and $k_{i+1} = \max\{k_i, \lceil \epsilon_{k_i}^{-2} \rceil\}$ for each $i \geq 0$. Note that $k_0 \leq k_1 \leq \cdots$.

Let us label the elements $r_1, r_2, \ldots, r_{p^n}$ of $\mathbb{F}_p^n$ so that

$$|\widehat{f}(r_1)| \geq |\widehat{f}(r_2)| \geq \cdots .$$

By Parseval (Theorem 6.1.3), we have

$$\sum_{j=1}^{p^n} |\widehat{f}(r_j)|^2 = \mathbb{E} f^2 \leq 1.$$

There is some positive integer $m \leq \lceil \epsilon_0^{-2} \rceil$ so that

$$\sum_{k_m < j \leq k_{m+1}} |\widehat{f}(r_j)|^2 \leq \epsilon_0^2, \tag{6.6.3}$$

since otherwise adding up the sum over all $m \leq \lceil \epsilon_0^{-2} \rceil$ would contradict $\sum_r |\widehat{f}(r)|^2 \leq 1$. Also, we have

$$|\widehat{f}(r_k)| \leq \frac{1}{\sqrt{k}} \quad \text{for every } k. \tag{6.6.4}$$

The idea now is to split

$$f(x) = \sum_{j=1}^{p^n} \widehat{f}(r_j)\omega^{r_j \cdot x}$$

into

$$f = f_{\mathrm{str}} + f_{\mathrm{sml}} + f_{\mathrm{psr}}$$

according to the sizes of the Fourier coefficients. Roughly speaking, the large spectrum will go into the structured piece f_{str}, the very small spectrum will go into pseudorandom piece f_{psr}, and the remaining middle terms will form the small piece f_{sml} (which has small L^2 norm by (6.6.3)).

Let $W = \{r_1, \ldots, r_{k_m}\}^{\perp}$ and set

$$f_{\mathrm{str}} = f_W .$$

Then, by (6.6.1),

$$\widehat{f_{\mathrm{str}}}(r) = \begin{cases} \widehat{f}(r) & \text{if } r \in W^{\perp}, \\ 0 & \text{if } r \in W^{\perp}. \end{cases}$$

Let us define f_{psr} and f_{sml} via their Fourier transform (and we can recover the functions via the inverse Fourier transform). For each $j = 1, 2, \ldots, p^n$, set

$$\widehat{f_{\mathrm{psr}}}(r_j) = \begin{cases} \widehat{f}(r_j) & \text{if } j > k_{m+1} \text{ and } r_j \notin W^{\perp}, \\ 0 & \text{otherwise.} \end{cases}$$

Finally, let $f_{\mathrm{sml}} = f - f_{\mathrm{psr}} - f_{\mathrm{sml}}$, so that

$$\widehat{f_{\mathrm{sml}}}(r_j) = \begin{cases} \widehat{f}(r_j) & \text{if } k_m < j \le k_{m+1} \text{ and } r_j \notin W^{\perp}, \\ 0 & \text{otherwise.} \end{cases}$$

Now we check that all the conditions are satisfied.

Structured piece. We have $f_{\mathrm{str}} = f_W$ where $\operatorname{codim} W \le k_m \le k_{\lceil \epsilon_0^{-2} \rceil}$, which is bounded as a function of the sequence $\epsilon_0 \ge \epsilon_1 \ge \ldots$.

Pseudorandom piece. For every $j > k_{m+1}$, we have $|\widehat{f}(r_j)| \le 1/\sqrt{k_{m+1}}$ by (6.6.4), which is in turn $\le \epsilon_{k_m} \le \epsilon_{\operatorname{codim} W}$ by the definition of k_m. It follows that $\|\widehat{f_{\mathrm{psr}}}\| \le \epsilon_{\operatorname{codim} W}$.

Small piece. By (6.6.3),

$$\|\widehat{f_{\mathrm{sml}}}\|_2^2 \le \sum_{k_m < j \le k_{m+1}} |\widehat{f}(r_j)|^2 \le \epsilon_0^2 . \qquad \square$$

Exercise 6.6.13. Deduce Theorem 6.6.4 from Theorem 6.6.11 by using an appropriate sequence ϵ_i and using the same W guaranteed by Theorem 6.6.11.

Remark 6.6.14 (Spectral proof of the graph regularity lemma)**.** The proof technique of Theorem 6.6.11 can be adapted to give an alternate proof of the graph regularity lemma (along with certain weak and strong variants). Instead of iteratively refining partitions and tracking energy increments as we did in Chapter 2, we can first take a spectral decomposition of the adjacency matrix A of a graph:

$$A = \sum_{i=1}^{n} \lambda_i v_i v_i^{\intercal},$$

where $v_1, \dots, v_n$ is an orthonormal system of eigenvectors with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. Then, as in the proof of Theorem 6.6.11, we can decompose A as

$$A = A_{\mathrm{str}} + A_{\mathrm{psr}} + A_{\mathrm{sml}}$$

with

$$A_{\mathrm{str}} = \sum_{i \leq k} \lambda_i v_i v'_i \quad A_{\mathrm{psr}} = \sum_{i > k'} \lambda_i v_i v'_i \quad \text{and} \quad A_{\mathrm{sml}} = \sum_{k < i \leq k'} \lambda_i v_i v'_i$$

for some appropriately chosen k and k' similar to the proof of Theorem 6.6.11.

We have

$$\sum_{i=1}^{n} \lambda_i^2 = \operatorname{tr} A^2 \leq n^2.$$

So $\lambda_i \leq n/\sqrt{i}$ for each i. We can guarantee that the spectral norm of A_{psr} is small enough as a function of k and ϵ. Furthermore, we can guarantee that $\operatorname{tr} A_{\mathrm{sml}}^2 = \sum_{k<i\leq k'} \lambda_i^2 \leq \epsilon$.

To turn A_{str} into a vertex partition, we can use the approximate level sets of the top k eigenvectors $v_1, \dots, v_k$. Some bookkeeping calculations then shows that this is a regularity partition. Intuitively, A_{psr} provides us with regular pairs. Some of these regular pairs may not stay regular after adding A_{sml}, but since A_{sml} has $\leq \epsilon$ mass (in terms of L^2 norm), it destroys at most a negligible fraction of regular pairs.

See Tao (2007a, Lemma 2.11) or Tao's blog post *The Spectral Proof of the Szemerédi Regularity Lemma* (2012) for more details of the proof.

6.7 Popular Common Difference

Roth's theorem has the following qualitative strengthening. Given $A \subset \mathbb{F}_3^n$ with density α, there is some "popular common difference" $y \neq 0$ so that the number of 3-APs in A with common difference y is $\geq \alpha^3 - o(1)$, i.e., at least approximately as much as one should expect if A were a random subset of density α. This was proved by Green (2005a) as an application of his arithmetic regularity lemma (from the previous section).

Theorem 6.7.1 (Roth's theorem with popular common difference in $\mathbb{F}_3^n$)
For all $\epsilon > 0$, there exists $n_0 = n_0(\epsilon)$ such that for $n \geq n_0$ and every $A \subset \mathbb{F}_3^n$ with $|A| = \alpha 3^n$, there exists $y \neq 0$ such that

$$\left|\{x \in \mathbb{F}_3^n : x, x+y, x+2y \in A\}\right| \geq (\alpha^3 - \epsilon)3^n.$$

In particular, Theorem 6.7.1 implies that every 3-AP-free subset of $\mathbb{F}_3^n$ has size $o(3^n)$.

Exercise 6.7.2. Show that it is *false* that every $A \subset \mathbb{F}_3^n$ with $|A| = \alpha 3^n$, the number of pairs $(x, y) \in \mathbb{F}_3^n$ with $x, x+y, x+2y \in A$ is $\geq (\alpha^3 - o(1))3^{2n}$, where $o(1) \to 0$ as $n \to 0$.

We will prove Theorem 6.7.1 via the next result, which concerns the number of 3-APs with common difference coming from some subspace of bounded codimension, which is picked via the arithmetic regularity lemma.

Theorem 6.7.3 (Roth's theorem with common difference in some subspace)
For every $\epsilon > 0$, there exists M so that for every $A \subset \mathbb{F}_3^n$, there exists a subspace W with codimension at most M, so that

$$\left|\{(x, y) \in \mathbb{F}_3^n \times W : x, x+y, x+2y \in A\}\right| \geq (\alpha^3 - \epsilon)3^n \left|W\right|.$$

Proof. By the arithmetic regularity lemma (Theorem 6.6.4), there is some M depending only on ϵ and a subspace W of $\mathbb{F}_p^n$ of codimension $\leq M$ so that A is ϵ-uniform on all but at most ϵ-fraction of W-cosets.

Let $u + W$ be a W-coset on which A is ϵ-uniform. Denote the density of A in $u + W$ by

$$\alpha_u = \frac{|A \cap (u + W)|}{|W|}.$$

Restricting ourselves inside $u+W$ for a moment, by the 3-AP counting lemma Lemma 6.2.4, the number of 3-APs of A (including trivial ones) that are contained in $u + W$ is

$$|\{(x, y) \in (u + W) \times W : x, x+y, x+2y \in A\}| \geq (\alpha_u^3 - \epsilon) \left|W\right|^2.$$

Since A is ϵ-uniform on all but at most ϵ-fraction of W-cosets, by varying $u + W$ over all such cosets, we find that the total number of 3-APs in A with common difference in W is

$$\left|\{(x, y) \in \mathbb{F}_3^n \times W : x, x+y, x+2y \in A\}\right| \geq (1-\epsilon)(\alpha^3 - \epsilon)3^n \left|W\right| \geq (\alpha^3 - 2\epsilon)3^n \left|W\right|.$$

This proves the theorem (with ϵ replaced by 2ϵ). □

Exercise 6.7.4. Give another proof of Theorem 6.7.3 using Theorem 6.6.11 (arithmetic regularity decomposition $f = f_{\text{str}} + f_{\text{psr}} + f_{\text{sml}}$).

Proof of Theorem 6.7.1. First apply Theorem 6.7.3 with find a subspace W of codimension $\leq M = M(\epsilon)$. Choose $n_0 = M + \log_3(1/\epsilon)$. So $n \geq n_0$ guarantees $|W| \geq 1/\epsilon$.

We need to exclude 3-APs with common difference zero. We have

$$\begin{aligned}(\alpha^3 - \epsilon)3^n\,|W| &\leq \left|\{(x,y) \in \mathbb{F}_3^n \times W : x, x+y, x+2y \in A\}\right| \\ &= \left|\{(x,y) \in \mathbb{F}_3^n \times (W \setminus \{0\}) : x, x+y, x+2y \in A\}\right| + |A|\,.\end{aligned}$$

We have $|A| \leq 3^n \leq \epsilon 3^n\,|W|$, so

$$(\alpha^3 - 2\epsilon)3^n\,|W| \leq \left|\{(x,y) \in \mathbb{F}_3^n \times (W \setminus \{0\}) : x, x+y, x+2y \in A\}\right|.$$

By averaging, there exists $y \in W \setminus \{0\}$ satisfying

$$\left|\{x \in \mathbb{F}_3^n : x, x+y, x+2y \in A\}\right| \geq (\alpha^3 - 2\epsilon)3^n.$$

This proves the theorem (with ϵ replaced by 2ϵ). □

By adapting the above proof strategy with Bohr sets, Green (2005a) proved that a Roth's theorem with popular differences in finite abelian groups of odd order, as well as in the integers.

Theorem 6.7.5 (Roth's theorem with popular difference in finite abelian groups)
For all $\epsilon > 0$, there exists $N_0 = N_0(\epsilon)$ such that for all finite abelian groups Γ of odd order $|\Gamma| \geq N_0$, and every $A \subset \Gamma$ with $|A| = \alpha\,|\Gamma|$, there exists $y \in \Gamma \setminus \{0\}$ such that

$$|\{x \in \Gamma : x, x+y, x+2y \in A\}| \geq (\alpha^3 - \epsilon)\,|\Gamma|\,.$$

Theorem 6.7.6 (Roth's theorem with popular difference in the integers)
For all $\epsilon > 0$, there exists $N_0 = N_0(\epsilon)$ such that for every $N \geq N_0$, and every $A \subset [N]$ with $|A| = \alpha N$, there exists $y \neq 0$ such that

$$|\{x \in [N] : x, x+y, x+2y \in A\}| \geq (\alpha^3 - \epsilon)N.$$

See Tao's blog post *A Proof of Roth's Theorem* (2014) for a proof of Theorem 6.7.6 using Bohr sets, following an arithmetic regularity decomposition in the spirit of Theorem 6.6.11.

Remark 6.7.7 (Bounds)**.** The above proof of Theorem 6.7.1 gives $n_0 = \text{tower}(\epsilon^{-O(1)})$. The bounds Theorems 6.7.5 and 6.7.6 are also tower-type. What is the smallest $n_0(\epsilon)$ for which Theorem 6.7.1 holds? It turns out to be $\text{tower}(\Theta(\log(1/\epsilon)))$, as proved by Fox and Pham (2019) over finite fields and Fox, Pham, and Zhao (2022) over the integers. Although it had been known since Gowers (1997) that tower-type bounds are necessary for the regularity lemmas themselves, Roth's theorem with popular differences is the first regularity application where a tower-type bound is shown to be indeed necessary.

Using quadratic Fourier analysis, Green and Tao (2010c) extended the popular difference result over to 4-APs.

Theorem 6.7.8 (Popular difference for 4-APs)
For all $\epsilon > 0$, there exists $N_0 = N_0(\epsilon)$ such that for every $N \geq N_0$ and $A \subset [N]$ with $|A| = \alpha N$, there exists $y \neq 0$ such that

$$|\{x : x, x+y, x+2y, x+3y \in A\}| \geq (\alpha^4 - \epsilon)N.$$

It may be a surprising that such a statement is false for APs of length 5 or longer. This was shown by Bergelson, Host, and Kra (2005) with an appendix by Ruzsa giving a construction that is a clever modification of the Behrend construction (Section 2.5).

Theorem 6.7.9 (Popular difference fails for 5-APs)
Let $0 < \alpha < 1/2$. For all sufficiently large N, there exists $A \subset [N]$ with $|A| \geq \alpha N$ such that for all $y \neq 0$,

$$|\{x : x, x+y, x+2y, x+3y, x+4y \in A\}| \leq \alpha^{c \log(1/\alpha)} N.$$

Here $c > 0$ is some absolute constant.

For more on results of this type, as well as for popular difference for high dimensional patterns, see Sah, Sawhney, and Zhao (2021).

Chapter Summary

- Basic tools of discrete Fourier analysis: Fourier transform, Fourier inversion formula, Parsevel's identity, convolution identity (Fourier transform converts convolutions to multiplication).
- The **finite field model** (e.g., $\mathbb{F}_3^n$) offers a convenient playground for Fourier analysis in additive combinatorics. Many techniques can then be adapted to the integer setting, although often with additional technicalities.
- **Roth's theorem.** Using Fourier analysis, We proved that every 3-AP-free subset has size at most
  - $O(3^n/n)$ in $\mathbb{F}^n$, and
  - $O(N/\log\log N)$ in $[N] \subset \mathbb{Z}$.
- The Fourier analytic proof of Roth's theorem (both in $\mathbb{F}_3^n$ and in $\mathbb{Z}$) proceeds via a **density increment argument**:
  (1) A 3-AP-free set has a large Fourier coefficient;
  (2) A large Fourier coefficient implies density increment on some hyper-

plane/subprogression;

(3) Iterate the density increment.

- Using the the **polynomial method**, we showed that every 3-AP-free subset of $\mathbb{F}_3^n$ has size $O(2.76^n)$.
- **Arithmetic regularity lemma.** Given $A \subset \mathbb{F}_p^n$, we can find a bounded codimension subspace so that A is Fourier-uniform on almost all cosets.
  - An application: **Roth's theorem with popular difference.** For every $A \subset \mathbb{F}_3^n$, there is some "popular 3-AP common difference" with frequency at least nearly as much as if A were random.

Further Reading

Green has several surveys and lecture notes on the topics covered in this and subsequent chapters.

- *Finite Field Models in Additive Combinatorics* (2005c) — Green argues that one should begin the study of many additive combinatorics problems in the finite field setting (also see the the follow up by Wolf (2015)).
- *Montreal Lecture Notes on Quadratic Fourier Analysis* — introduces quadratic Fourier analysis and explains how to prove the popular common difference theorem for 4-APs in $\mathbb{F}_5^n$.
- Lecture notes from his Cambridge course *Additive Combinatorics* (2009b) — an excellent introduction to the subject.

Tao's FOCS 2007 tutorial *Structure and Randomness in Combinatorics* (2007a) explains many facets of arithmetic regularity and applications.

For more on algebraic methods in combinatorics (pre-dating methods in Section 6.5), see the books:

- *Thirty-three Miniatures* by Matoušek (2010);
- *Linear Algebra Methods in Combinatorics* by Babai and Frankl;
- *Polynomial Methods in Combinatorics* by Guth (2016).

See any undergraduate textbook on Fourier analysis for an introduction from an analysis point of view. In particular, the book *Fourier Analysis* by Stein and Shakarchi (2003) is highly recommended. The analysis viewpoint usually has a very different emphasis compared to the topic of this chapter, though many standard tools (e.g., Parseval) are common to both. It is helpful to be familiar with certain general principles of Fourier analysis, such as the relationship between smoothness and decay.