CS 6815: Lectures 6 and 7

Instructor: Eshan Chattopadhyay
Guest lecturer: Bobby Kleinberg
Scribes: Yuwen Wang and Matvey Soloviev

11th and 13th September 2018

1 Background and Notation.

Definition 1.1. *Let $a, b, c \in \mathbb{F}$. The elements are in* arithmetic progression *if they can be written as $a = x$, $b = x + y$, and $c = x + 2y$ for $x, y \in \mathbb{F}$. If y is non-zero, then they form a* non-degerate *arithmetic progession. We will additionally abbreviate 3-term arithmetic progressions with 3APs.*

The capset problem. If $A \subseteq \mathbb{F}_3^n$ has no non-degenerate 3APs, how large can $|A|$ be?

Immediately we can say two quick facts about 3APs in $\mathbb{F}_3^n$:

1. Using the arithmetic properties in $\mathbb{F}_3^n$, we have that

$$\begin{aligned} x_0, x_1, x_2 \text{ is a 3AP in } \mathbb{F}_3^n &\iff x_1 - x_0 = x_2 - x_1 \\ &\iff 0 = x_0 + x_2 - 2x_1 \\ &\iff 0 = x_0 + x_1 + x_2 \end{aligned}$$

2. Consider a matrix of "random" 3APs generated by the rule in 1.:

$$\begin{bmatrix} 1 & 1 & 2 & 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & 2 & 1 & 2 & 0 \\ 2 & 1 & 1 & 0 & 2 & 2 & 2 \end{bmatrix}$$

The columns of 3APs are either three distinct elements or three of the same element.

An adjacent problem and some history. If $A \subseteq [N]$ has no *distinct* elements in arithmetic progression, how large can $|A|$ be?

- Roth (1930's): $|A| \leq O(N \log\log N)$.
- Salem-Spencer-Gehrand-Elkins-Green-Wolff: There exists an A with $|A| \geq N \cdot 2^{-C\sqrt{\log N}}$, which can be made larger than $N^{1-\varepsilon}$ for any $\varepsilon > 0$.
- Sanders (2010): $|A| = O(N(\log\log N)^5/\log N)$.
- Related question: if $A \subseteq \mathbb{N}$ and $\sum_{n \in A} 1/n = \infty$ must A arbitraily long non-degenerate APs? Must it contain 3APs?

Mathematicians studying these problems (mainly using Fouier analysis) saw the capset problem as an easier spinoff problem in which if they made progress, they may make progress on these older problems. Thus, they were spurred on to use Fourier/Roth based techniques for the capset problem but a much more simple and efficient solution was found in 2016 using the polynomial method. Both methods will be useful for studying pseudorandomness.

2 Roth-Meshulam Theorem: upper bound using Fourier analysis.

Definition 2.1. *If G is a finite abelian group and $f, g : G \to \mathbb{C}$, their* convolution, *$f * g$, is the function*

$$(f * g)(x) = \mathop{\mathbb{E}}_{y \leftarrow G}[f(y)g(x-y)] = \mathop{\mathbb{E}}_{y+z=x}[f(y)g(z)]$$

Properties 2.2. *Let χ_v and χ_w be two characters on $G = \mathbb{F}_p^n$, and f, g be two function on $\mathbb{F}_p^n$. We can write f and g as $f = \sum_v \hat{f}(v)\overline{\chi_v}$ and $g = \sum_w \hat{g}(w)\overline{\chi_w}$.*

(a) $\chi_v * \chi_w = \begin{cases} \chi_v(x) & \text{if } v = w \\ 0 & \text{if } v \neq w \end{cases}$

(b) $f * (g + h) = f * g + f * h$

(c) $f * g = \sum_v \hat{f}(v)\hat{g}(v)\overline{\chi_v}$

(d) $\widehat{f * g} = \hat{f} \cdot \hat{g}$

Proof. (a)

$$\begin{aligned}(\chi_v * \chi_w)(x) &= \mathop{\mathbb{E}}_{y}[\chi_v(y)\chi_w(x-y)] \\ &= \mathop{\mathbb{E}}_{y}[\chi(\langle v, y\rangle)\chi(\langle w, x-y\rangle)] \\ &= \mathop{\mathbb{E}}_{y}[\chi(\langle v, y\rangle + \langle w, x-y\rangle)] \\ &= \mathop{\mathbb{E}}_{y}[\chi(\langle w, x\rangle)]\mathop{\mathbb{E}}_{y}[\chi(\langle v-w, y\rangle)] \\ &= \chi(\langle w, x\rangle)\mathbb{1}(v = w)\end{aligned}$$

(b) By linearity of expectation.

(c)

$$\begin{aligned}f * g &= \left(\sum_v \hat{f}(v)\overline{\chi_v}\right) * \left(\sum_w \hat{g}(w)\overline{\chi_w}\right) \\ &= \sum_v \sum_w \hat{f}(v)\hat{g}(w)(\overline{\chi_v} * \overline{\chi_w}) \qquad \text{(by (b))} \\ &= \sum_v \hat{f}(v)\hat{g}(v)\overline{\chi_v} \qquad \text{(by (a))}\end{aligned}$$

(d)

$$\begin{aligned}\widehat{f * g}(v) &= \mathop{\mathbb{E}}_{x}[(f * g)(x)\overline{\chi_v}(x)] \\ &= \mathop{\mathbb{E}}_{x}\left[\sum_w \hat{f}(w)\hat{g}(w)\overline{\chi_w}(x)\overline{\chi_v}(x)\right] \\ &= \sum_w \hat{f}(w)\hat{g}(w)\mathop{\mathbb{E}}_{x}[\overline{\chi_w}(x)\overline{\chi_v}(x)] \\ &= f(v)g(v) \qquad \text{(by orthonormality)}\end{aligned}$$

□

Given $A \subseteq \mathbb{F}_3^n$, let

$$f(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}.$$

And define μ to be $\mathbb{E}_x[f(x)] = \hat{f}(0)$.

Lemma 2.3. *If $\frac{1}{N} < \frac{1}{2}\mu^3$ and $|\hat{f}(v)| < \frac{1}{2}\mu^2$ for all $v \neq 0$, A must contain a non-degenerate 3AP.*

Proof. Let $N = 3^n$. Then

$$\frac{1}{N^2}\sum_{x,y,z} f(x)f(y)f(z) = \underbrace{\frac{\mu N}{N^2}}_{\text{\# of degenerate 3APs}} + \frac{1}{N^2}\#\{\text{non-degenerate 3APs}\}$$

The left hand side and also be written as

$$\begin{aligned} (f * f * f)(0) &= \sum_v (\hat{f}(v))^3 \\ &= (\hat{f}(0))^3 + \sum_{v\neq 0}(\hat{f}(v))^3 \\ &> \mu^3 - \sum_v |\hat{f}(v)|^3 \\ &> \mu^3 - \frac{\mu^2}{2}\sum_v |\hat{f}(v)|^2 \quad . \end{aligned}$$

Since f is an $\{0,1\}$-function,

$$\sum_v \hat{f}(v)\hat{f}(v) = \langle f, f\rangle = \mathbb{E}[f] = \mu.$$

Then,

$$\cdots = \frac{1}{2}\mu^3 > \frac{1}{N},$$

by assumptions, and

$$\frac{1}{N^2}\#\{\text{non-degenerate 3APs}\} > \frac{1-\mu}{N} > 0.$$

□

This lemma tells us that if $\mu \leq 2/N^{2/3}$, then $|A| = \mu N \leq 2N^{2/3}$ (then, A is small and we are done). Otherwise the $|\hat{f}(v)| \geq \frac{1}{2}\mu^2$ for some $v \neq 0$. The idea is that $\mathbb{F}_3$ will have a lower dimensional subspace T such that

$$\frac{|A \cap T|}{|T|} \geq \mu + \frac{1}{4}\mu^2.$$

Then, we can iterate this to get a contradiction.

To show this, define $T_i = \{x \mid \langle v, x\rangle = i\}$ for $i = 0,1,2$, where v is the vector assumed to exist above. Define also

$$\mu_i = \frac{|A \cap T_i|}{|T_i|} = \mu + \delta_i,$$

and $\omega = \overline{\chi(1)} = \exp(2\pi i/3)$. Then we have

$$\begin{aligned}\hat{f}(v) &= \frac{1}{3}\mu_0 + \frac{\omega}{3}\mu_1 + \frac{\omega^2}{3}\mu_2 \\ &= \left(\frac{1}{3} + \frac{\omega}{3} + \frac{\omega^2}{3}\right)\mu + \left(\frac{\delta_0}{3} + \frac{\delta_1\omega}{3} + \frac{\delta_2\omega^2}{3}\right) \\ &= \frac{\delta_0}{3} + \frac{\delta_1\omega}{3} + \frac{\delta_2\omega^2}{3}.\end{aligned}$$

In addition,

$$\frac{\mu_0}{3} + \frac{\mu_1}{3} + \frac{\mu_2}{3} = \mu \quad \implies \quad \frac{\delta_0}{3} + \frac{\delta_1}{3} + \frac{\delta_2}{3} = 0.$$

We then have

$$\begin{aligned}\left|\frac{\delta_0}{3} + \frac{\delta_1\omega}{3} + \frac{\delta_2\omega^2}{3}\right| &> \frac{1}{2}\mu^2 \\ \left|\frac{\delta_0}{3}\right| + \left|\frac{\delta_1}{3}\right| + \left|\frac{\delta_2}{3}\right| &> \frac{1}{2}\mu^2 \\ \sum_{i=0}^{2}\left(\frac{\delta_i}{3} + \left|\frac{\delta_i}{3}\right|\right) &> \frac{1}{2}\mu^2.\end{aligned}$$

Then there exists a i_0 such that

$$\frac{\delta_{i_0}}{3} + \left|\frac{\delta_{i_0}}{3}\right| > \frac{1}{6}\mu^2.$$

Then $\delta_{i_0} + |\delta_{i_0}| > \mu^2/2$ and $\delta_{i_0} > \mu^2/4$. This completes the proof for the following proposition:

Proposition 2.4. *If $N = 3^n$, A is a capset of $\mathbb{F}_3^n$, and $\mu = |A|/N > 2N^{-1/3}$, then $\mathbb{F}_3^{n-1}$ contains a capset of density of greater than $\mu + \mu^2/4$.*

What happens if we iterate this proposition?

$$\begin{array}{rccccccccc} \text{Density} & \mu & \to & \mu + \frac{1}{4}\mu^2 & \to & \mu + \frac{1}{4}\mu^2 + \frac{1}{4}\left(\mu + \frac{1}{4}\mu^2\right)^2 & \to & \cdots & & \\ \geq & \cdots & & \mu + \frac{\mu}{4}\mu & \to & \mu + \frac{2\mu}{4}\mu + o(\mu^2) & \to & \mu + \frac{3\mu}{4}\mu + o(\mu^2) & \to & \cdots \\ \text{Dimension} & n & \to & n-1 & \to & n-2 & \to & n-3 & \to & \cdots \end{array}$$

Generally, at dimension $n-k$, we have density $\geq \mu + \frac{k\mu}{4}\mu$. So if we start at density ρ, after subtracting $4/\rho$ dimensions, the density has doubled. Now iterate *this.*

$$\begin{array}{rccccccccccc} \text{Density} & \mu & \to & 2\mu & \to & 4\mu & \to & 8\mu & \to & \cdots & & \\ \text{Dimension} & n & \to & n - 4/\mu & \to & n - 4/\mu - 4/2\mu & \to & n - 4/\mu - 4/2\mu - 4/4\mu & \to & \cdots & \to & n - 8/\mu. \end{array}$$

Since the density cannot exceed 1, this process must stop at some density $\rho \leq 1$ and dimension d, where the premise of the proposition no longer holds: in other words, at that point, $\mu \leq \rho \leq 2D^{-1/3}$, where $D = 3^d$. (*)

Assume for the sake of contradition that $\mu > 16/n$, then

$$\begin{aligned} \frac{8}{\mu} < \frac{n}{2} \quad &\Longrightarrow \quad \frac{n}{2} < d \qquad \text{(dimension never drops below } n - 8/\mu \text{ in the chain)} \\ &\Longrightarrow \quad D^{1/3} = 3^{d/3} > 3^{n/6} \\ &\Longrightarrow \quad \frac{16}{n} < \mu \leq \rho \leq 2D^{-1/3} < \frac{2}{3^{n/6}} \qquad \text{(by (*))} \end{aligned}$$

The last line is always a contradiction for sufficiently large n, so we have proven the following theorem.

Theorem 2.5 (Roth-Meshulam). *Let A be a capset in $\mathbb{F}_3^n$. Then there exists n_0 such that for all $n \geq n_0$,*

$$\frac{|A|}{3^n} \leq \frac{16}{n}.$$

3 Upper bound using polynomial method.

Theorem 3.1 (2016 Ellenberg-Gijswijt using Croot-Lev-Pach). *Let A be a capset in $\mathbb{F}_3^n$, then*

$$|A| < 2.756^n \quad \textit{or} \quad \frac{|A|}{3^n} < 0.92^n.$$

Proof. The vector space of $\mathbb{F}_3$-valued function on $\mathbb{F}_3^n$ (as an $\mathbb{F}^3$-vector space) has two interesting bases:

1. The functions
$$\delta_{\boldsymbol{y}}(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \boldsymbol{x} = \boldsymbol{y} \\ 0 & \text{otherwise} \end{cases}.$$

2. The monomials of max degree
$$\begin{aligned} f(x_1, \ldots, x_n) &= x_1^{\alpha_1} \cdots x_n^{\alpha_n} \quad (\alpha_i \in \{0,1,2\}) \\ &= \boldsymbol{x}^{\boldsymbol{\alpha}}. \end{aligned}$$

   where $\boldsymbol{\alpha}$ is a tuple of the same length as $\boldsymbol{x}$.

The δ-basis can be expressed in terms of degree-2 monomials via

$$\delta_{\boldsymbol{y}}(\boldsymbol{x}) = \prod_{i=1}^{n} \left[1 - (x_i - y_i)^2\right].$$

Suppose $P(\boldsymbol{x}, \boldsymbol{y}) = \sum_{\alpha,\beta} C_{\alpha,\beta} \boldsymbol{x}^{\alpha} \boldsymbol{y}^{\beta}$ is a polynomial in $2n$ variables. Then

- the *coefficient matrix* $C(P)$ is the $3^n \times 3^n$ matrix $(C_{\alpha,\beta})_{\alpha,\beta}$ with rows and columns indexed by power vectors $\in \{0,1,2\}^n$
- the *evaluation matrix* $E(P)$ is the $3^n \times 3^n$ matrix $(P(\boldsymbol{x}, \boldsymbol{y}))_{\boldsymbol{x},\boldsymbol{y}}$ with rows and columns indexed by n-variable input vectors $\in \mathbb{F}_3^n$.

Let V denote the *Vandermonde matrix* $(\boldsymbol{x}^{\boldsymbol{\alpha}})_{\boldsymbol{x},\boldsymbol{\alpha}}$. Then

$$E(P) = VC(P)V^T,$$

and hence $\operatorname{rank} E(P) \le \operatorname{rank} C(P)$. (In fact, they are equal since V is invertible.)

Let $L_{n,d}$ denote the subspace of $\mathbb{F}_3$-valued functions on $\mathbb{F}_3^n$ spanned by monomials of total degree $\le d$. Assume that n is a multiple of 3, and let $d = 2n/3$. Set $K = \dim(L_{n,d})$. We have

$$\dim(L_{n,d}) + \dim(L_{n,2d-1}) = 3^n.$$

(Scribe note: we can see this by considering the bijection between the set of monomials with coefficient vector $\boldsymbol{\alpha} \in \{0,1,2\}^n$ of total degree $\|\boldsymbol{\alpha}\|_1 \ge 2d = 4n/3$ and the set of monomials of total degree $\le d = 2n/3$ defined by $\boldsymbol{\alpha} \mapsto (2,\ldots,2) - \boldsymbol{\alpha}$.)

Let W be the subspace of $L_{n,2d-1}$ consisting of polynomials that vanish on $\bar{A}$. Since $\dim(L_{n,2d-1}) = 3^n - K$ and vanishing at each $\boldsymbol{y} \notin A$ puts one linear constraint on the coefficients, we have

$$\begin{aligned} \dim(W) &\ge (3^n - K) - (3^n - |A|) \\ &= |A| - K. \end{aligned}$$

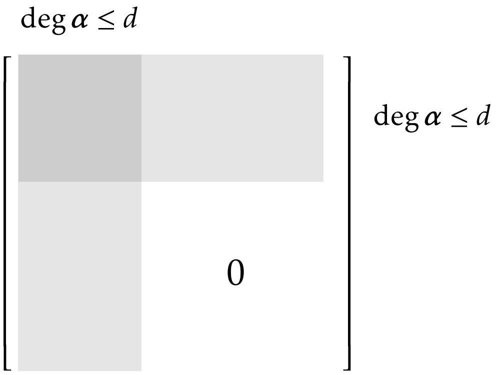
Then there exists a polynomial $Q(\boldsymbol{z}) \in W$ such that $Q(\boldsymbol{x}) = 1$ at $|A| - K$ points $\boldsymbol{x}$. Let S denote the set of these points, i.e.

$$A \supseteq S = \{\boldsymbol{x} \mid Q(\boldsymbol{x}) = 1\}.$$

If $\boldsymbol{x}, \boldsymbol{y} \in S$, $\boldsymbol{x} \ne \boldsymbol{y}$, then $(\boldsymbol{x} + \boldsymbol{y})/2 \notin S$. (Scribe note: otherwise the three elements would form a 3AP in S and hence in A.) Let $P(\boldsymbol{x}, \boldsymbol{y}) = Q\left(\frac{\boldsymbol{x}+\boldsymbol{y}}{2}\right)$. Then in $E(P)$, we have an $|S| \times |S|$ identity matrix:

$$\begin{array}{c|cc|c} & \boldsymbol{x} & \boldsymbol{y} & \\ \hline & 1 & 0 & \boldsymbol{x} \\ & 0 & 1 & \boldsymbol{y} \\ \hline \end{array} \quad \forall \boldsymbol{x}, \boldsymbol{y} \in S$$

Hence, $\operatorname{rank} E(P) \ge |S|$. On the other hand, if we assume monomials to be ordered by their total degree, $C(P)$ has the following structure:


deg $\boldsymbol{\alpha} \le d$
deg $\boldsymbol{\alpha} \le d$
0

Hence, $\operatorname{rank} C(P) \leq 2K$. (Scribe note: the first K rows may be linearly independent, and at most K of the following rows since each of them only has K nonzero entries.) Combining inequalities,

$$|A| - K \leq |S| \leq \operatorname{rank} E(P) \leq \operatorname{rank} C(P) \leq 2K$$

and hence

$$|A| \leq 3K.$$

Recall that K was the dimension of $L_{n,d}$, i.e. the subspace spanned by monomials of total degree at most $2n/3$ (and, as before, individual powers at most 2). We find that if n is sufficiently large, $3K \leq (2.756)^n$. □

Polynomial method: we used the fact that high-degree polynomials can be made to do whatever over finite fields, but low degree polynomials are much more constrained.