# Math 616 (Additive Combinatorics) Notes

January → April 2025

# Contents

# Lecture 1

Consider a finite set $A$ inside a group/field, $|A| = n$. The sumset is $|A+A| = \{a_i + a_j : a_i, a_j \in A\}$. The questions here often concern the size of this sumset depending on properties of $A$. Let's start by considering $A \subset \mathbb{R}$.

**Theorem 1.1.** $|A + A| \geq 2|A| - 1$. *If equality holds, then $A$ is an aritmetic progression (AP).*

*Proof.* Write $A = \{a_1 < a_2 < ... < a_n\}$, ordered. The sumset has elements

$$a_1 + a_1, a_1 + a_2, ..., a_1 + a_n, a_2 + a_n, a_3 + a_n, ..., a_n + a_n$$

which are ordered as well.

We want to show $A$ is an arithmetic progression, so $A = \{a_1 + kd : k = 0, 1, ..., n-1\}$. We can show $a_i + a_{j+1} = a_{i+1} + a_j$. Essentially consider the same list of ordered sums from above, but note that you can take any taxicab path in the matrix below and get an ordered path. You can take a path to $a_i + a_{j+1}$, and then just switch to $a_{i+1}a_j$ without knowing these are equal and continue. This gives $2|A| - 1$ elements, which forces $a_i + a_{j+1}$ to equal $a_{i+1}a_j$.

$$\begin{array}{ccccc} a_n + a_1 & ... & a_n + a_i & ... & a_n + a_n \\ ... & ... & ... & ... & ... \\ a_i + a_1 & ... & a_i + a_j & ... & a_i + a_n \\ ... & ... & ... & ... & ... \\ a_1 + a_1 & ... & a_1 + a_j & ... & a_1 + a_n \end{array}$$

□

What about products? $A \cdot B = \{a_i b_j : a_i \in A, b_j \in B\}$. If $A, B \in \mathbb{R}_+$, the same theorem as before holds (because you can just take log of all the elements). But if negative numbers are allowed, you can take $A = \{-1, 0, 1\}$ which gives $|AA| = 3$, not $2|A| - 1 = 5$. We will need a tool to explore this further.

**Theorem 1.2** (Hilbert's Nullstellensatz)**.** *If $Z$ is the common zero set (the intersection of zero sets) of polynomials $g_1, g_2, ..., g_k$ in an algebraically closed field, and $f$ is a polynomial that vanishes exactly on $Z$, then we can write*

$$f^m = \sum_i h_i g_i$$

*for some $m \in \mathbb{N}$ and polynomials $h_i$.*

**Theorem 1.3** (Combinatorial Nullstellensatz)**.** *If $\mathbb{F}$ is a field and $F(x, y) \in \mathbb{F}[x, y]$ is a polynomial that vanishes on the common zero set of $g(x) \in \mathbb{F}[x]$ and $h(y) \in \mathbb{F}[y]$, then we can write $f(x, y) = k(x, y)g(x) + \ell(x, y)h(y)$, where $\deg k \leq \deg f - \deg g$ and $\deg \ell \leq \deg f - \deg h$.*

Note that $f(x, y)$ vanishing on the common zero set of $g(x)$ and $h(y)$ really means that it is vanishing on a Cartesian product.

**Theorem 1.4** (Cauchy-Davenport)**.** *Let $A, B \subseteq \mathbb{F}_p$. Then $|A + B| \geq \min\{p, |A| + |B| - 1\}$.*

*Proof.* Let $c_i$ be the elements of $A+B$, and suppose $d := |A+B| < p$. Define

$$F(x,y) = \prod_i (x+y-c_i)$$

The degree of $F$ is $d = |A+B|$. Let's define $g(x) = \prod_i (x-a_i)$ and $h(y) = \prod_i (y-b_i)$, which satisfy the conditions of Theorem 1.3. The degrees of these are $|A|$ and $|B|$. Write

$$F(x,y) = k(x,y)g(x) + \ell(x,y)h(y)$$

Consider a term of $F(x,y)$, $x^{|A|-1}y^{d-|A|+1}$, which has coefficient $\binom{d}{|A|-1}$. The maximal degree of $k$ is $d-|A|$ by the theorem, so the $y$-degree of this term implies that this term must be present in the $\ell(x,y)h(y)$ part. In particular, the $x$-degree of $\ell$ satisfies $\deg_x \ell \geq |A|-1$. The largest $y$-degree of $\ell$ is $\leq d-|B|$ from the theorem since $h$ has degree $|B|$. Together this means $|A|-1 \leq \deg \ell \leq d-|B|$, and rearranging yields $d \geq |A|+|B|-1$. □

# Lecture 4

We are talking about generalized arithmetic progressions ($\mathrm{GAP}_d$ where $d$ is the dimension), and know that

$$|\mathrm{GAP}_d| \leq \prod_{i=1}^{d} \ell_i.$$

Freiman's dimension Lemma says that if $A \subseteq \mathbb{R}^d$, and $A$ is truly $d$-dimensional, (i.e., $A$ is not contained in a $d-1$ dimensional flat), then

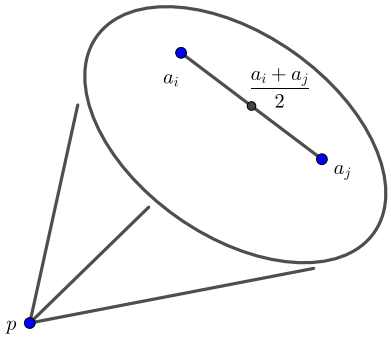
$$|A+A| \geq (d+1)|A| - \frac{d(d+1)}{2}$$

*Proof.* Induction, both on the size of $A$ and $d$. The base case $d=1$ is done. Specifically, suppose it is true for any set in $d-1$ dimensions and try to prove it for $d$ by doing induction on $|A|$. Recall $\mathrm{Conv}(A)$ is the convex hull of $A \subset \mathbb{R}^d$. Choose a point $p$ from the vertex set of the polytope defined by $\mathrm{Conv}\, A$.

Consider the convex hull of all the vertices with $p$ removed (i.e., $\mathrm{Conv}(A \backslash p)$). From $p$, we can "see" at least $d$ vertices of this new convex hull, unless it is $d-1$ dimensional. In other words, there are at least $d$ vertices $q_1, \ldots, q_d$ of $\mathrm{Conv}(A \backslash p)$ such that the $\overline{q_i p}$ interval has only one point $q_i$ from $\mathrm{Conv}(A \backslash p)$. There are two cases

1. $\dim(\mathrm{Conv}(A \backslash p) = d$. In this case, the induction hypothesis is

$$|(A \backslash p) + (A \backslash p)| \geq (d+1)(|A|-1) - \frac{d(d+1)}{2}$$



All the halving points are in the convex hull of $A\backslash p$. $\frac{p+q_i}{2}$ are outside ($d$ of them).

2. $\dim(\text{Conv}(A\backslash p) = d-1$. In this case, there are still $d$ vectros visible from $p$, otherwise everything can be expressed as $d-1$ vectors but $A$ was supposed to be $d$ dimensional. Then W.L.O.G., suppose $p = 0$ and the other points lie in the plane $x_1 = 1$. The sumset has elements where $x_0 = 0$, 1, or 2. The first class comes only from $p+p$, the second class comes from $p+q_i$, which are $|A|-1$ in number, and the third are handled by the induction, giving

$$|(A\backslash p) + (A\backslash p)| \geq d(|A|-1) - \frac{d(d-1)}{2}.$$

Adding these together gives the right bound.

□

**Corollary 4.1.** *Let $A \subset \mathbb{R}^d$ and $|A+A| \leq C|A|$. Then $A$ is contained in a flat of dimension at most $C$.*

**Corollary 4.2.** *Let $A \subset \mathbb{N}$ and $|A \cdot A| \leq C|A|$. Then the* rank *of the multiplicative group generated by $A$ (which we henceforth refer to as $\langle A \rangle$ is at most $C$.*

To see this second one, let $B = \{\log a : a \in A\}$ and apply Freiman's lemma to show that $|B+B| \leq C|B|$.

**Theorem 4.3** (Schmidt's Subspace)**.** *Let $a, b \in \mathbb{C}$, and we looking for the solutions to $ax+by = 1$ for $x, y \in A$, where* $\text{rank}\langle A \rangle \leq r$. *Then the number of solutions is $\leq B(r)$.*

# Lecture 5

**Theorem 5.1** (Baker's)**.** *Unstated. (this would make a good presentation if you want to!)*

Today we return to theorem 4.3. But first, a problem: what can you say about $|2^n - 3^m|$ as $n$ goes to infinity, and $m$ is chosen to minimize this value? What is a lower bound: $|2^n - 3^m| \geq$?

Fix an integer $k$, we want to show that the number of solutions to $2^n - 3^m = k$ is bounded. Let $A = \{2^n 3^m : n, m \in \mathbb{Z}\}$. We take $a, b = \frac{1}{k}$, and note $\langle A \rangle = 2$. So the solutions to this are at most $B(2)$ for each $k$. This means There is an $f(n) \to \infty$ so that $|2^n - 3^m| \geq f(n)$ for any $m$. We have $f(n) = \Omega(n)$.

Without using this tool, can we show $2^n - 3^m = 1$ only happens finitely many times? You can show this by noting that that $2^n - 3^m = 1 \mod 8$, which doesn't happen if $n \geq 3$.

**Theorem 5.2** (Subspace – Schmidt, Evertse, ...)**.** *Let $a_1, \ldots, a_n \in \mathbb{C}$, and*

$$a_1 x_1 + a_2 x_2 + \ \ldots \ + a_n x_n = 1 \tag{5.1}$$

*be a linear form. There exists a bound $B(n, r)$ such that the number of non-trivial solutions to (5.1) when $x_i \in S$ where* $\text{rank}\langle S \rangle \leq r$ *is $\leq B(n, r)$.*

Note that subsets where subsets add up to 0 give infinitely many solutions, like consider $n = 3$, you can get lots of the form $x_1 = -x_2$ and $x_3 = 1$. So 'non-trivial' here means that no subsets add up to 0.

**Conjecture 5.3** (Erdős-Szemerédi Sum-Product)**.** *Let $A \subset \mathbb{N}$ be a finite set (there are versions for $\mathbb{Q}$, $\mathbb{C}$, $\mathbb{F}_p$, $\mathbb{F}_q$...). For any $\varepsilon > 0$, there is $n_0(\varepsilon)$ so that if $n \geq n_0$, then $|A+A|+|A\cdot A| \geq |A|^{2-\varepsilon}$.*

For now, let $A \subset \mathbb{R}^+$, and let $|A \cdot A| < C|A|$. We can show that $|A + A| = \frac{1}{2}|A|^2 + O(|A|)$. To do this, first note that Freiman's dimensional lemma implies that $\operatorname{rank}\langle A\rangle \leq C$. Now, we want to say the number of solutions too $x_1 + x_2 = x_3 + x_4$ is very small. For $x_4 \neq 0$, we can rewrite this as

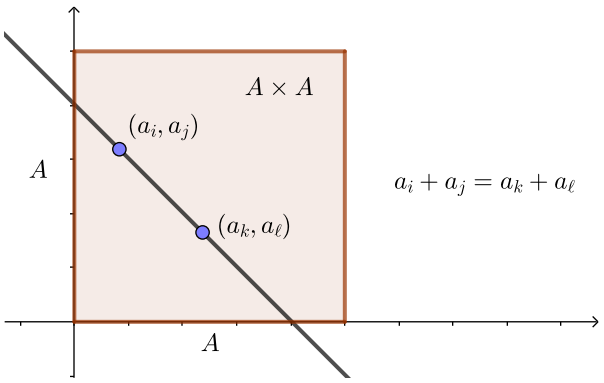
$$\frac{x_1}{x_4} + \frac{x_2}{x_4} - \frac{x_3}{x_4} = 1.$$

Which is $Y_1 + Y_2 - Y_3 = 1$. From the subspace theorem, the number of non-trivial solutions is at most $B(3, C)$. This has some trivial solutions $Y_1 = 1$ and $Y_2 - Y_3 = 0$, and $Y_2 = 1$ and $Y_1 - Y_3 = 0$. Looking back to where these came from, the number of trivial solutions is $O(|A|)$.

*Recommended reading*: Combinatorial applications of the subspace theorem.

# Lecture 6

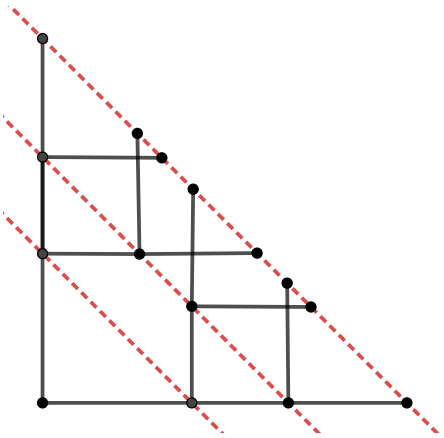
Suppose $|A + A|$ is small (that is, $\leq |A|^{1+c}$). Consider $A_1 = A \times A$. The lines $x + y = c$ are lines with slope $-1$, and notice that lines of this slope cover all the points and we have an equality of sum if and only if two points lie on the same line. So, each line contains $\approx \frac{n^2}{n^{1+c}} = n^{1-c}$ points.




Choose $\ell_1$, a slope $-1$ line with $\geq n^{1-c}$ points on it. Select points $A_{x1} = \operatorname{proj}_x(A \times A \cap \ell_1)$ and $A_{y1} = \operatorname{proj}_y(A \times A \cap \ell_1)$. Let $B_1 = A_{x1} \times A_{y1}$. The size of $|B_1| \geq n^{2-2c}$ (roughly), and the sumset still satisfies $|A_{x1} + A_{x1}| \leq n^{1+c}$. Notably,

$$n^{1+c} = \left(n^{(1-c)}\right)^{\frac{1+c}{1-c}} \approx |A_{x1}|^{\frac{1+c}{1-c}} = |A_{x1}|^{1+\frac{2c}{1-c}},$$

so by the above, we have that $|A_{x_1} + A_{y_1}| \leq |A_{x_1}|^{1+c_1}$ with $c_1 = \frac{2c}{1-c}$. Now, repeat this process with $B_1$ and $c_1$ to create a new set $C_1$. Repeat this over and over finding $A_{x2}$ and $A_{x3}$ and so on until we find some $A_{xi}$ where $|A_{xi} \times A_{yi}| \leq n^{1+c}$, i.e. the cartesian product has become so small that the sumset bound doesn't say anything. (If $c = 0$, you can do $\log \log n$ steps.)



The $x$-projection is a 'Hilbert cube' generated by $\{x_0, x_1, \dots, x_d\}$, i.e. the set

$$H_d = \left\{ x_0 + \sum_{i=1}^{d} \eta_i x_i : \eta_i \in \{0,1\} \right\}. \tag{6.1}$$

You can see this by starting at $x_0$ (bottom left corner in diagram) which comes from the very last step in the process above, and moving towards the top-right. Each step has distance $x_i$, and $\eta_i$ depends on whether you go up (came from an $x$ projection) or right ($y$ projection).

**Corollary 6.1.** *For $d \in \mathbb{N}$, there is a $c_d > 0$ such that if $|A + A| \leq n^{1+c_d}$, then $A$ contains a Hilbert cube of dimension $d$.*

**Theorem 6.2** (Falting's (special case))**.** *Let $y^2 = f(x)$, a polynomial (over the rationals) such that the curve defined over $\mathbb{CP}^2$ has genus $q \geq 2$, then the number of rational solutions is finite.*

We apply this next time to the polynomial $y^2 = (x^2 + a)(x^2 + b)(x^2 + a + b)$ with $a \neq b$.

# Lecture 7

We begin the class with a claim: For the complex polynomial $f(x)$ with distinct roots and degree $\geq 5$, the algebraic curve $y^2 = f(x)$ has genus $\geq 2$. Today we are talking a bit about about hyperelliptic curves, but not going far into the algebraic geometry.

*Recommended reading*: Heights in Diophantine Geometry by Bombieri and Gubler.

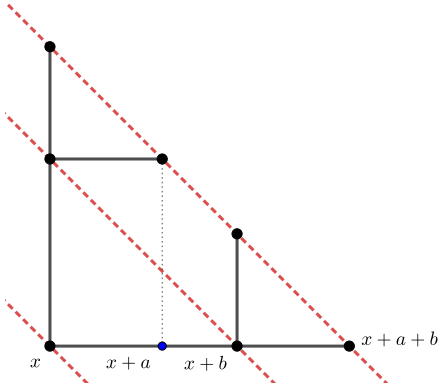
**Conjecture 7.1** (Bombieri-Lang/Uniformity (special case))**.** *In Falting's theorem (6.2), the number of solutions is uniformly bounded by the genus.*

*Recommended watching*: a Wiles interview wherein he is sad. His problem was Fermat's Last Theorem: to show that $x^n + y^n = z^n$ has no solutions for $n \geq 3$.

**Conjecture 7.2.** *Let $A \subset \{$squares in $\mathbb{N}\}$. Then $|A + A| \geq |A|^{2-\varepsilon}$.*

**Exercise 7.3.** *prove that $|A + A| \geq |A|^{1+c}$ for some $c > 0$, assuming Bombieri-Lang.*

*Solution.* If $A + A$ is smaller than this, then there is a Hilbert cube of dimension $d$, for any $d$ we want. What we need is that there are "many" $x \in \mathbb{N}$ so that $x, x + a, x + b, x + a + b$. You find as many such $x$ by choosing $d$ larger... in particular, the number of such $x$ is at least $d$.




We then consider the polynomial

$$y^2 = (w^2 + a)(w^2 + b)(w^2 + a + b).$$

This has six distinct roots, hence genus two[1]. The uniformity conjecture says that we have

$$|\{x \in \mathbb{N} : x, x + a, x + b, x + a + b \in A\}| \leq B$$

where $B$ is independent of $a$ and $b$. So for the bound, we are choosing the dimension of the Hilbert cube to be $d = B$. □

**Theorem 7.4** (Mei-Chu-Chang)**.** *In the problem above, we have the lower bound* $|A + A| > |A| \cdot \log |A|$

*Proof.* (Steps only)

- Freiman's theorem
- Szemerédi's theorem (about $AP_4$s in dense subsets of a larger AP)
- Squares contain no AP of length 4 (Fermat)

□

The last statement, 'Squares contain no AP of length 4', can be proven with 'Fermat descent': suppose there are some of those, find the smallest one, and then you can do some operations to find an even smaller one (but it is tricky).

**Problem 7.5.** *What is the max number of squares in an AP of length $n$?*

Note: This problem has an upper bound of $c_1 N^{3/5} (\log N)^{c_2}$ for two positive absolute, and computable, constants $c_1$, $c_2$ by Bombieri and Zannier. The conjecture is that $O(\sqrt{n})$ is right.

[1]We didn't discuss genus too precisely, so could be helpful to see the wiki: `https://en.wikipedia.org/wiki/Hyperelliptic_curve`

**Problem 7.6.** ***(Homework Q1)*** *Assume the uniformity conjecture and use it to give a $n^{1-c}$ type upper bound on the previous problem (7.5).*

*A (bad) homework Q1 solution.* Call the arithmetic progression $A$, and consider a set $S \subset A$ of squares. We know that the sumset of $A$ is small, $|A+A| = 2n-1$. So $|S+S| \leq 2n-1$. From exercise 7.3 we know that $|S+S| > |S|^{1+c'}$, so we get $|S| < (2n-1)^{\frac{1}{1+c'}} \approx n^{1-\frac{c'}{1+c'}}$. So our constant is $c = \frac{c'}{1+c'}$. From the proof of 7.3, the constant $c' = c_B$ in the notation of Corollary 6.1, where $B$ is the bound coming from the uniformity conjecture. □

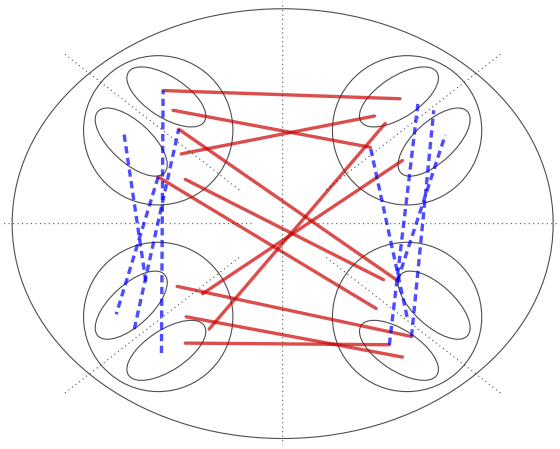
# Lecture 9

We began by proving Ramsey's theorem, that a two-colouring of $K_n$ contains a monochromatic clique of any size we want, if we take $n$ to be large enough. We get $R(k,k) \leq 4^k$, which has recently been beaten with $\approx 3.99^k$ with some advanced tools.

By random colouring, using Lovasz-Local-Lemma, we get the lower bound of $2\sqrt{2}^k$.

Let $R(\overbrace{3,3,\ldots,3}^{k})$ be the number such that the $k$ colouring of the edges of $K_{R(3,3,\ldots,3)}$ results in a monochromatic triangle. We abbreviate this notation with $R_k(3)$, and can prove the result for this similar to what we did last time: choose any vertex and take the most popular color among edges, say, red in this case. The size of edges with the same color is $\geq \frac{n-1}{k}$, and suppose there is no triangle; then there is no more of that color in the subgraph. We then have $R_k(3) \leq kR_{k-1}(3)$, and so $R_k(3) \lesssim k!$

For the other direction we can consider that the complete bipartite graph has no triangle, so colour edges joining two halves one colour, then use a new colour between the bipartite classes, and repeat this to get a lower bound of $2^k$. This is depicted in the following figure.



A related question: how many colours can we have at most so there is a mono-$\chi$ solution to the equation $x+y=z$? The idea behind this is: Say we have a colouring of the integers. Let's create a graph on the first $n$ integers, where the edge joining $i$ and $j$ is the colour of the vertex $i-j$. A triangle made out of three edges, say between vertices $i<j<k$ was coloured according to the points $a=k-j$, $b=j-i$ and $c=k-i$. Thus if the triangle is mono-$\chi$, then $a,b,c$ received the same same colour, and we can see that $a+b=c$. So the number of colours needed is at least $2^k$ from the previous stuff.

We can use this info to prove Fermat's conjecture: For any $n \in \mathbb{N}$ there is a $p'$ in $\mathbb{Z}$ so that for any prime $p \geq p'$, the equation $x^n + y^n = z^n \mod p$ has a non-trivial solution. The proof will be based on colouring $\mathbb{Z}_m$ according to cosets. All elements in the coset receive the same colour, so we are using $(p-1)/n$ colours.

# Lecture 10

Recall the Hilbert cube (equation (6.1)):

$$H_d = \left\{ x_0 + \sum_{i=1}^{d} \varepsilon_i x_i : \varepsilon_i \in \{0,1\} \right\}.$$

This is kind of a weaker structure than a GAP, it's sort of like a GAP with only one point in each direction. We showed that sets with small sumsets have Hilbert cubes, and last time we showed some things about colouring.

We can combine them to show for example: A stronger statement is:

**Theorem 10.1** (Hindman's)**.** *For any $k$-colouring of $\mathbb{N}$, there is an $S \subset \mathbb{N}$ where $|S| = \infty$ so that elements of $S$ and all its finite sums have the same colour.*

**Theorem 10.2** (Finite Union/Folkman's)**.** *For any $k, d \in \mathbb{N}$, there is a threshold $n_0(k,d)$ such that for any $k$-colouring of $2^S$ (the subsets of $S$) where $|S| \geq n_0$, there will be $S_1, S_2, \ldots, S_d$, disjoint subsets of $S$, such that $\cup_{i=1}^{d} \varepsilon_i S_i$ have the same colour.*

**Exercise 10.3.** *Use Theorem 10.2 to show: For any $d, k \in \mathbb{N}$, there is an $n_0(k,d)$ such that any $k$-colouring of $[n]$ where $n \geq n_0$, there is a mono-$\chi$ Hilbert cube of dimension $d$ with $x_0 = 0$.*

*Solution.* Take $m = dn$ to be 'large enough' (larger than $n_0^2/2$) where $n_0$ comes from 10.2. Colour the subsets of $[m]$ according to the colour of the sum of its elements. The theorem says that there will be $d$ disjoint subsets of $[n]$ of the same colour and the union of them is also the same colour. Such a set system corresponds to a mono-$\chi$ Hilbert cube.

Alternatively: we have a $k$-colouring of $[n]$. Colour a subset of $[n]$ according to the colour of its cardinality. □

Now, I think it was stated that an exercise is to show that with non-empty $S_0$, You can show there is a monochromatic set

$$S_0 \cup \bigcup \varepsilon_i S_i.$$

Towards proving the Union theorem, the first step/idea is: let $S_0 \cup \bigcup_{i=1}^{d_1} S_i$ be monochromatic, and colour $S_0$ by its colour... Consider the $d_1$ elements $S_i$ as the atoms, and now repeat. Find a monochromatic $S_0^{(1)} \cup \bigcup_{i=1}^{d_2} S_i^{(1)}$ and repeat again.

# Lecture 11

In this lecture we construct a large subset of $[n]$ that has no 3-APs. To begin, something you might think of is to take all numbers which have no 2 in their base-3 expansion. How many numbers is this? Well, it is $n^{\log_3 2}$. That's pretty good, but we can do even better! To summarize the results of this section, we have the following theorem.

**Theorem 11.1.** *There exists a subset of $[n]$ which has no 3-term arithmetic progressions, and whose size is at least:*

- $n \cdot e^{-\frac{c \log n}{\log \log n}}$ *from Salem-Spencer's construction, or even larger,*
- $n \cdot e^{-c\sqrt{\log n}}$ *from Behrend's Construction.*

## Salem-Spencer

Consider two numbers $m$ and $k$ to be chosen later. The $n$ we are using is $n = (2k)^m$. Let $\ell = \frac{m}{k}$. We want to choose numbers from $[n]$ in where

- all digits in base $2k$ are less than $k$,
- all digits are used the same number of times, i.e. each digit from 0 to $k-1$ appears exactly $\ell$ times.

Why is this 3AP free? Think about how we could have $a + b = 2c$ from numbers in this set. Because all digits are less than $k$, the only way this equation can hold is if we have an arithmetic progression in each digit $a_i + b_i = 2c_i$. In particular, $c_i$ will be 0 iff $a_i$ and $b_i$ are both 0. Therefore, the 0s of $a$ and $b$ must occur in the same spots, else $c$ would have fewer 0s. With this established, you can see the same thing holds for the digit 1, and 2, and so on. So the only way this equation holds is trivially, when $a$, $b$, and $c$ are all the same number.

Now we count how many numbers there are in our set. We have

$$\binom{m}{\ell} \cdot \binom{m-\ell}{\ell} \cdot \binom{m-2\ell}{\ell} \cdot ... = \frac{m!}{\ell!(m-\ell)!} \cdot \frac{(m-\ell)!}{\ell!(m-2\ell)!} \cdot \frac{(m-2\ell)!}{\ell!(m-3\ell)!} \cdot ... = \frac{m!}{(\ell!)^k}$$

such numbers. Call the set of these $X$. Using Sterling's approximation, we get

$$\frac{m!}{(\ell!)^k} \approx \frac{\sqrt{2\pi m}\,(m/e)^m}{\left(\sqrt{2\pi\ell}\,(\ell/e)^\ell\right)^k} = \sqrt{2\pi m}(2\pi\ell)^{-k/2}k^m = \sqrt{k}(2\pi\ell)^{(1-k)/2}k^m$$

So we want to maximize roughly $|X| \approx (\gamma\frac{m}{k})^{-k/2}k^m$ for $\gamma = 2\pi$, subject to $(2k)^m = n$. Salem and Spencer suggest defining $k$ to be the number where $(2k)^{k\log^2 k} = n$. So then $m = k\log^2 k$, and $\frac{m}{k} = \log^2(k)$. We compute the size of $X$ to be

$$|X| \approx \gamma^{-k/2}(\log^{-k}(k))k^{k\log^2 k} = \gamma^{-k/2}(\log^{-k}(k))2^{-k\log^2 k}n. \tag{11.1}$$

Now, from the definition of $k$, we have the following two identities,

$$\log n = k\log^2 k \log(2k),$$

$$\log\log n = \log k + 2\log\log k + \log\log(2k).$$

The most significant term in $\log\log n$ is the $\log k$ term, which leads us to

$$\frac{\log n}{\log\log n} > k\log^2 k(1 + o(1)).$$

Now, using (11.1), we have

$$\log|X| = -\frac{k}{2}\log\gamma - k\log(\log(k)) - k\log^2 k\log 2 + \log n > \log n - \frac{\log n}{\log\log n}(1+o(1)),$$

and finally, raising $e$ to the power of both sides, we have

$$|X| \approx n^{1-\frac{c}{\log\log n}}.$$

**Remark 11.1.** *Alternatively, you can take $k = \frac{1}{2}\log^2 n$. Then*

$$m = \log_{2k} n = \frac{\log n}{2\log\log n}, \qquad \textit{and} \qquad \ell = \frac{m}{k} = \frac{\frac{\log n}{2\log\log n}}{\frac{1}{2}\log^2 n} = \frac{1}{\log\log n\log n},$$

*so that*

$$|X| = (\log\log n\log n)^{\frac{1}{4}\log^2 n}(2)^{-\frac{\log n}{2\log\log n}} n > n\cdot e^{-\frac{c\log n}{\log\log n}}.$$

## Behrend

Behrend had a different approach which does better, and involves some geometry. His major idea is that strictly convex surfaces in higher dimension do not have collinear triples, so we could try mapping one of those with many points on it to our integers.

Again pick integers $k$ and $m$, and set $n = (2k)^m$. Consider an $k$-dimensional lattice cube $[k]^m \cap \mathbb{Z}^m$ and family of spheres $x_1^2 + x_2^2 + \cdots + x_m^2 = t$ for $t = 1, \ldots, mk^2$. Each point in the cube is contained in one of the spheres, so by the pigeonhole principal, some sphere has $k^m/mk^2$ lattice points. Call the set of these points $A \subset \mathbb{R}^m$.

Since spheres do not contain any collinear triples, $A$ does not contain any either. Now we map the set $A$ to a subset $X \subset [n]$ by sending

$$a = (a_1, a_2, \ldots, a_m) \qquad \text{to} \qquad x = a_1 + a_2(2k) + a_3(2k)^2 + \cdots + a_m(2k)^{m-1},$$

that is, we treat $a_i$ as $i$'th digit of $x$ in base $2m$. Just like in Salem-Spencer, a 3AP here would be a 3AP in each digit, but that would correspond to a 3AP in $A$. Setting $m = c\sqrt{\log n}$, we get $k = \frac{1}{2}e^{\frac{1}{c}\sqrt{\log n}}$, so there cardinality of the construction is

$$|X| = \frac{k^m}{mk^2} = \frac{(\frac{1}{2})^{c\sqrt{\log n}}n}{c\sqrt{\log n}\cdot\frac{1}{4}\cdot e^{\frac{2}{c}\sqrt{\log n}}} = \frac{n}{\frac{c}{4}\sqrt{\log n}e^{\sqrt{\log n}(c\log 2 + 2/c)}} = ne^{-\sqrt{\log n}(c\log 2 + 2/c) - \frac{1}{2}\log\log n},$$

so there is a progression-free subset of $\{1, 2, \ldots, n\}$ of size at least $ne^{-\sqrt{\log n}(c\log 2 + 2/c + o(1))}$. You can choose $c = \sqrt{2/\log(2)}$ to maximize the constant.

# Lecture 12

A circulant matrix is an $n \times n$ matrix where the rows are shifts of each other. So they look like

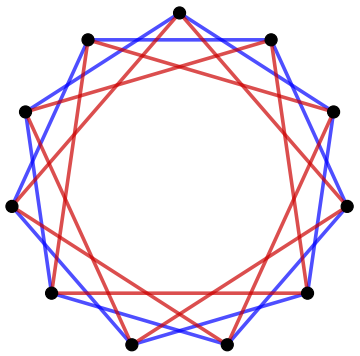
$$W = \begin{pmatrix} c_1 & c_2 & \ldots & c_{n-1} & c_n \\ c_n & c_1 & \ldots & c_{n-2} & c_{n-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ c_2 & c_3 & \ldots & c_n & c_1 \end{pmatrix}.$$

Let $\omega = e^{\frac{2\pi i}{n}}$. The eigenvalues of this matrix are

$$\sum_{j=1}^{n} c_j w^{kj}, \qquad k = 0, 1, \ldots, n-1$$

which each have multiplicity 1. Something we will see is that Cayley graphs have circulant adjacency matrices.

Let's consider $\mathbb{Z}/11\mathbb{Z}$, and draw the Cayley graph of the generators $\{2, 3\}$. The graph looks like this, where blue edges come from generator 2 and red from 3.



We can compute the adjacency matrix of this, which is

$$W = \begin{pmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \end{pmatrix}.$$

Writing down the eigenvalues in this case, we have Eigenvalues of $\omega^{3k} + \omega^{4k} + \omega^{9k} + \omega^{10k}$ for $k = 0, \ldots 10$.

Simpler question: write the Eigenvalues of $C_5$, the cycle on five vertices. Solution: $C_5$ is also circulant, so we get from the sum formula $1 + \omega^{2k}$ for $k = 1, 2, 3, 4$.

**Definition 12.1.** *The largest (also called "second") eigenvalue of a graph with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ is $\lambda = \max\{|\lambda_2|, |\lambda_n|\}$.*

Note that if $G_n$ is $d$-regular, then $\lambda_1 = d$ since $(1, 1, \ldots, 1)$ is an eigenvector. Why can't we have a larger one? Consider another eigenvector $v$, the entries of this are all at most 1. A row of $G$ has exactly $d$ 1s, so we can see the eigenvalue for this $v$ cannot be larger than $d$.

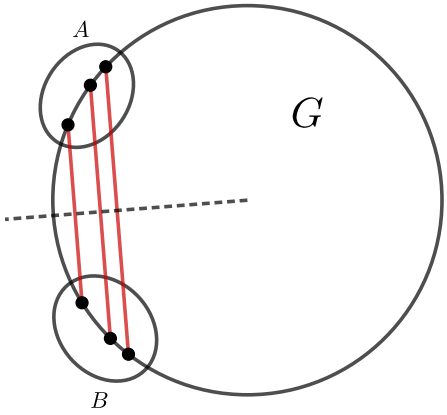
**Lemma 12.2** (Mixing)**.** *If $G_n$ is $d$-regular graph on $n$-vertices and $\lambda$ is its second eigenvalue, then for any distinct vertex sets $A, B \subset V(G_n)$, we have*

$$\left| e(A, B) - \frac{d|A||B|}{n} \right| \leq \lambda \sqrt{|A||B|} \tag{12.1}$$

This is also called an "isoperimetric inequality" or "Cheeger inequality".

**Theorem 12.3** (Roth's). *If $S$ is a subset of the first $n$ integers and $|S| \geq \frac{cn}{\log\log n}$, then $S$ has a 3-term arithmetic progression.*

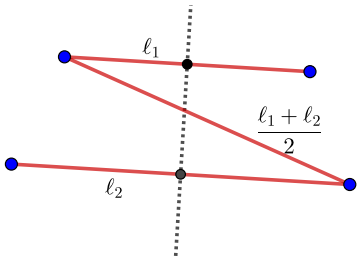
Towards proving this, we can actually solve this mod p since a progression from the middle third (which is a positive fraction) of $n \mod p$ will be a genuine 3AP in $[n]$. Consider the Cayley graph on $p$ vertices and generated by $p$, and connect two vertices with an edge if their difference is in $S$.




The key steps will be

- Show that if $S$ is 3AP free, then $G$ has large second eigenvalue.
- Density increment argument.

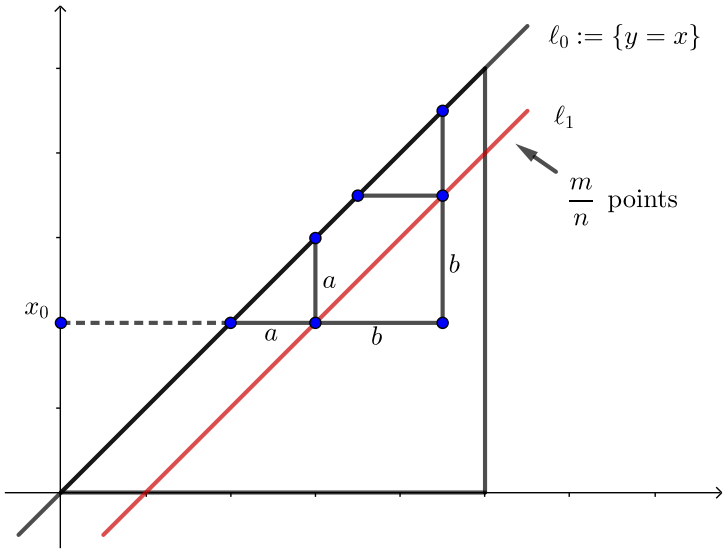
We will begin but not finish the proof today. Note this graph is $2|S|$-regular. Every edge has its midpoint somewhere, and there are $2p$ possible halving points. Take one halving point, which has many edges, and collect the two endpoints of these into two sets $A$ and $B$. There cannot be any other edge between $A$ and $B$ now, so we just have this matching of parallel edges. This is because if we had $\ell_1$ and $\ell_2$ between $A$ and be in the parallel-matching, and we also had another edge joining them, then we'd have a threeAP.




# Lecture 13

Today we went over in detail a proper solution to Homework 1 (Problem 7.6). What is the bound required on the sumset before you see a small Hilbert cube $H_{a,b} = \{x_0, x_0+a, x_0+b, x_0+a+b\}$? Precisely, if $S \subset [n]$ and $|S| \geq n^{1-\delta}$, for which $\delta$ is $S$ forced to have an $H_{a,b}$? Note that $2n-1$ lines cover the $n \times n$ grid, and $|S \times S| = n^{2-2\delta} =: m$, so there is a line containing $\approx \frac{1}{2}n^{1-2\delta}$ points. But in fact, you can find a line ($\{x = y\}$) which contains actually $|S|$ points, since we are starting

with an arithmetic progression. So we start with $|S|$ on a line and then we have $n-1$ lines below, and $\binom{|S|}{2}$ points below. So there is a line below this which contains $\frac{m}{n}$ points, call it $\ell_1$.




Out of the $\geq \frac{m}{n}$ points on $\ell_1$, at least $\frac{\binom{(m/n)}{2}}{n}$ pairs are at the same distance apart by the pigeonhole principle, so their difference is the same. Therefore, the number of $H_{a,b}$ with some $x_0$ is $\geq \frac{m^2}{n^3}$. The uniformity conjecture says that the number of solutions to

$$y^2 = (w^2 + a)(w^2 + b)(w^2 + a + b).$$

is bounded, so in particular, we have

$$|\{x \in \mathbb{N} : x, x + a, x + b, x + a + b \in A\}| \leq B$$

So this means $\frac{m^2}{n^3} \leq B$, which simplifies to $|S| \lesssim n^{1-\frac{1}{4}}$.

# Lecture 14*

## Elementary Sum-Product Phenomena

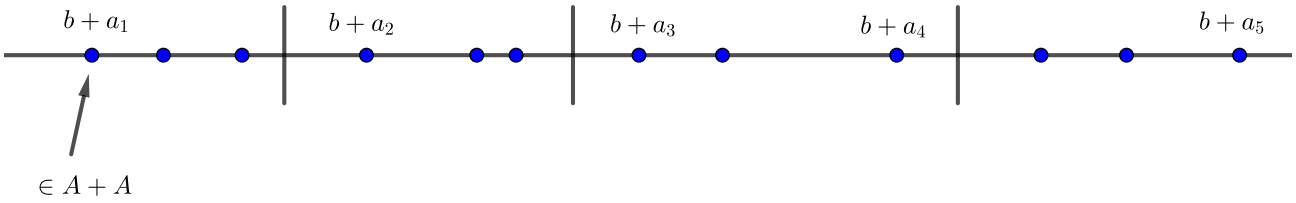
Under what circumstances do we know that $A + A$ (or $A \cdot A$) is large? 'Large' usually means $\geq |A|^{1+c}$.

**Definition 14.1.** *Let $A = \{a_1 < \cdots < a_n\} \subseteq \mathbb{R}^n$ be a set where $a_2 - a_1 < a_3 - a_2 < a_4 - a_3 < \ldots$. Then $A$ is called* convex.

A maybe conjecture: convex sets have close to maximal size sumset.

**Theorem 14.2** (Elekes, Nathanson, Ruzsa)**.** *Let $A$ be convex. Then $|A + A| \gtrsim |A|^{3/2}$.*

*Proof.* Divide $A+A$ into at most $|A|/2$ intervals, each containing at most $\frac{4|A+A|}{|A|}$ elements of $A+A$.




Count pairs $p, q \in A + A$ that have the following properties:

- $p, q$ are in the same interval.
- $p = b + a_i$ and $q = b + a_{i+1}$ for some $b \in A$, $q \leq i \leq n = 1$.

For a fixed $b \in A$, there are $n - 1$ elements $b + a_i$.

Towards a a lower bound, at most $n/2$ pairs $b + a_i$ and $b + a_{i+1}$ can be split up by one of the dividers. So at least $n - 1 - \frac{n}{2} \geq \frac{n}{4}$ are in the same interval. Summing over all the possibilities $b$, we get $n \cdot \left(\frac{n}{4}\right)$ pairs. However, have we counted the same thing multiple times? Notice that for a given $p = b + a_i$, and $q = b + a_{i+1}$, the difference $q - p = a_{i+1} - a_i$. These differences are unique, so from $p$ and $q$, we can recover $a_i$ and therefore $b$. That is, you cannot generate the same pair $p$, $q$ from different $a_i$ and $b$. So in the counting, we only counted each thing once.

Towards the upper bound, in each interval, we have at most $\frac{|A+A|^2}{|A|^2}$ such pairs. We have $|A|/2$ intervals, so the total is at most $\frac{|A+A|^2}{|A|}$. The two bounds together give $|A + A| \gtrsim |A|^{3/2}$. □

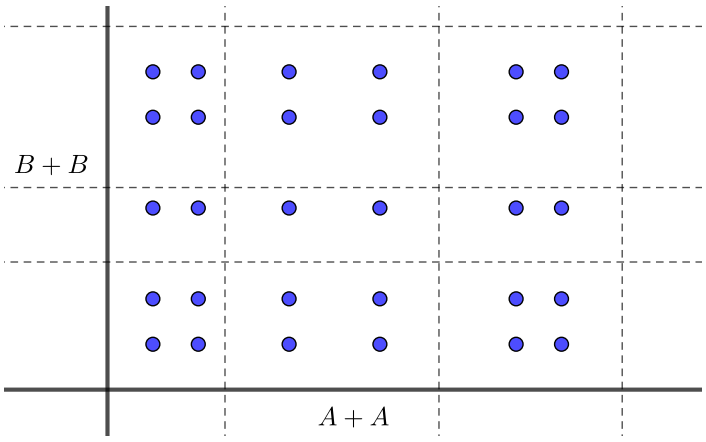
Note that we only needed that consecutive pairs have distinct differences, not convexity! There are a lot of generalizations of this. You can replace the second set with an arbitrary set $B$ for example and bound $|A + B|$.

We now show a 2D version of this, and give an application to the sum product problem (Conjecture 5.3).

**Theorem 14.3.** *Let $A = \{a_1 < \cdots < a_n\}$ and $B = \{b_1 < \cdots < b_n\}$, and suppose all vectors $(a_{i+1} - a_i, b_{i+1} - b_i)$ are distinct. Then*

$$|A + A| \cdot |B + B| \gtrsim n^{5/2}. \tag{14.1}$$

*Proof.* Divide $\mathbb{R}$ into at most $\frac{n}{4}$ intervals of at most $\frac{8|A+A|}{|A|}$ elements, and do the same for $B$. This divides the cartesian product $(A + A) \times (B + B)$ into at most $\frac{|A||B|}{4} = \frac{n^2}{16}$ intervals, each with at most $\frac{64|A+A||B+B|}{n^2}$ elements of $(A + A) \times (B + B)$.



# Lecture 15*

Today we continue the proof of 14.3. We want to count pairs $p, q \in (A+A) \times (B+B)$ such that

1. $p$ and $q$ are in the same cell
2. $\exists c \in A, d \in B$, $i \in [n-1]$ such that

$$p = (c + a_i, d + b_i), \quad q = (c + a_{i+1}, d + b_{i+1})$$

For the lower bound, there are $n$ options of $c, d$ and $n-1$ options for $i$. This gives $n^2(n-1)$ pairs $p$ and $q$. We have $\frac{n}{4}$ vertical divisions and $\frac{n}{4}$ horizontal. For a fixed $c, d$, this can divide at most $\frac{n}{2}$ of our pairs. So for each $c, d$, there are at least $\frac{n}{4}$ remaining pairs, for a total of $n^2\frac{n-1}{4}$.

As in the previous proof, we need to check we are not counting the same pair $p$ and $q$ multiple times though. Notice that $p - q = (a_{i+1} - a_i, b_{i+1} - b_i)$, so each pair $p$, $q$ we recover $i$ and so $b, c$.

Towards the upper bound, there are at most $\lesssim \frac{|A+A|^2|B+B|^2}{n^4}$ pairs in each cell, and $\lesssim n^2$ cells, so there are $\lesssim \frac{|A+A|^2|B+B|^2}{n^2}$ pairs. Altogether, simplifying,

$$n^3 \lesssim \{\#\text{pairs } p, q\} \lesssim \frac{|A+A|^2|B+B|^2}{n^2}$$

so $|A+A| \cdot |B+B| \gtrsim n^{5/2}$. □

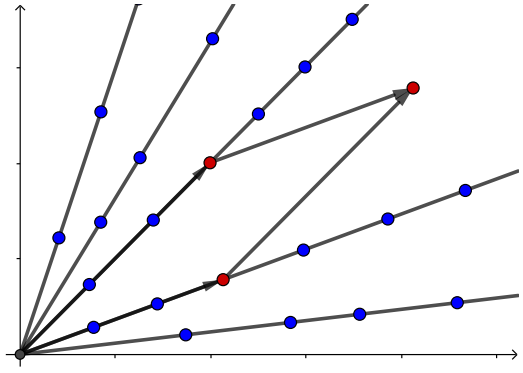
We now show the Elekes bound on sum-product (though, he didn't use this method).

**Theorem 15.1.** $\max\{|A+A|, |A \cdot A|\} \gtrsim |A|^{5/4}$

*Proof.* We may assume $A$ is positive, otherwise choose the larger half of $A$ that lies in the positive/negatives, and work just with that. Let $B := \log(A)$, so $|B+B| = |A \cdot A|$. Note that log is a convex curve, so the difference vectors of points on this curve are all distinct because the slope is decreasing. Applying the previous theorem, $|A+A| \cdot |A \cdot A| \gtrsim n^{5/2}$, so at least one of them is size $n^{5/4}$. □

The goal now is to show $\max\{|A+A|, |A \cdot A|\} \gtrsim \frac{|A|^{4/3}}{\log^{1/3}|A|}$ due to Solymosi. To start, we will show $|A+A|^2|A/A| \gtrsim |A|^4$, which means $\max\{|A+A|, |A/A|\} \gtrsim |A|^{4/3}$. Observe that:

- $(A+A) \times (A+A) = (A \times A) + (A \times A)$. This implies
- $|A+A|^2 = |(A \times A) + (A \times A)|$.
- A point $(a, b)$ in the cartesian product is on the line through the origin of slope $\frac{b}{a}$. So,
- We can cover $A \times A$ with $\leq |A/A|$ lines through the origin.
- All sums of vectors from consecutive lines are distinct, and lie between the lines.



Intuitively, if there are very few lines (so $|A/A|$ is small), the lines have many points on them, which will make the sumset large.

# Lecture 16

We have a set $A$, what can we say if $|A+A|$ is small ($\leq |A|$)? One reason we are interested in these things is convolution. For two discrete valued functions $f$ and $g$, the convolution is

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

Now, triangle inequalities for sumsets. If we have two sets $A, B \subseteq G$, then

$$|A-B| \leq \frac{|A-C||C-B|}{|C|} \tag{16.1}$$

Note that the sum of $a-c$ and $c-b$ is $a-b$.

$$|A+B| \leq \frac{|A+C||C+B|}{|C|} \tag{16.2}$$

Interestingly, (16.1) is relatively easy to prove, but (16.2) is hard. From these two we can get all combinations, so any choice of plus/minuses in the following inequality holds.

$$|A \pm B| \leq \frac{|A \pm C||C \pm B|}{|C|} \tag{16.3}$$

Now, can we use these to show something of the form $|A+A| \leq k|A| \implies |A-A| \leq$ __$|A+A|$? The answer is:

$$|A-A| \leq \frac{|A+A||A+A|}{|A|} \leq \frac{k^2|A|^2}{|A|} = k^2|A| \tag{16.4}$$

Now, what is the best exponent you can get showing that $|A-A| \leq |A+A|^\alpha$? One thing we can show is

$$|A-A| \leq \frac{|A+A|^2}{|A|} \leq \frac{|A+A|^2}{|A-A|^{1/2}}.$$

Simplifying, we get $|A-A| \leq |A+A|^{4/3}$. This is the best known bound. The best construction gives us a set where $\alpha$ has to be at least 1.18... though.

Consider $\mathbb{Z}^n$ and a radius $m$. Take the set $\left\{(x_1, x_2, \dots, x_n) : \sum_{|x_i| \le m}\right\}$. This set is the octahedron $O(n, m)$, and we now count the number of points in it. To do so, we first consider the positive quadrant $O^+$ where $x_i \ge 0$. We have $O^+ = \binom{n+m}{n}$, so the whole octahedron is

$$O = \sum_{k=0}^{\min(m,n)} 2^k \binom{m}{k}\binom{n}{k}.$$

If the positive quadrant is our $A$, then $|A| = \binom{n+m}{n}$ and $|A+A| = \binom{n+2m}{n}$. The content of the whole octahedron is contained in the difference set. To see that, any vector in the set can be broken into a sum of its positive and negative entries. That vector is the difference of those two vectors in the positive quadrant. Optimizing the $n$ and $m$ leads to the number 1.18...

In particular,

$$A - A = \left\{(x_1, \dots, x_n) : \sum_{i \in P} |x_i| \le m, \sum_{i \in N} |x_i| \le m\right\}.$$

So

$$A - A \approx \binom{n}{n/2}\left(2\binom{\frac{n}{2}+m}{m}\right) \ge \frac{2^n}{\sqrt{n}}\binom{\frac{n}{2}+m}{m}$$

Optimizing, let $r = \frac{n}{m}$. Then $|A| = \binom{(r+1)m}{rm}$, $|A-A| \ge \frac{2^{rm}}{\sqrt{rm}}\binom{(\frac{1}{2}r+1)m}{\frac{1}{2}r}$.

(Update: using the simplex, you can actually achieve around $n^{1.24\dots}$)

# Lecture 17*

More sum product today!

**Theorem 17.1** (Li, Shen, Solymosi)**.** *Let $A \subset \mathbb{R}$, then $|A+A|^2|A/A| \gtrsim |A|^4$. In particular, we get $\max|A+A|, |A/A| \gtrsim |A|^{4/3}$.*

*Proof.* Recall: cover $A \times A$ with a bunch of lines through the origin. $|A/A|$ of them cover the set. The sums of elements on two lines are in $(A\times A)+(A\times A)$, whose size is equal to $(A+A)\times(A+A)$.

We now throw away lines with fewer than $\frac{|A|^2}{2|A/A|}$ points on them. We lose at most $|A|^2/2$ points doing this. Index the remaining lines by $i_1, \dots, i_k$. Then

$$\frac{|A|^2}{2} \le \sum_{j=1}^{k} \left|L_{i_j} \cap (A \times A)\right|$$

Now, since $\left|L_{i_j} \cap (A \times A)\right| \le |A| \le \frac{|A|^2}{4}$, we can also let go of the top element. I.e.,

$$\frac{|A|^2}{4} \le \sum_{j=1}^{k-1} \left|L_{i_j} \cap (A \times A)\right|$$

Now,

$$|A+A|^2 = |(A\times A)+(A\times A)| \ge \sum_{j=1}^{k-1} \left|L_{i_j} \cap (A \times A)\right| \left|L_{i_{j+1}} \cap (A \times A)\right|$$

so

$$|A+A|^2 \geq \frac{|A|^2}{2|A/A|}\sum_{j=1}^{k-1}\left|L_{i_j}\cap(A\times A)\right| \gtrsim \frac{|A|^4}{2|A/A|}$$

which completes the argument. □

**Definition 17.2.** *For a set $A\subset\mathbb{R}$, the* additive energy *and* multiplicative energy *of $A$ are*

$$E^+(A) := \left|\left\{(a,b,c,d)\in A^4 : a+b=c+d\right\}\right|, \qquad E^\times(A) := \left|\left\{(a,b,c,d)\in A^4 : a+b=c+d\right\}\right|$$

Note that the trivial bounds are $|A|^2 \lesssim E^+(A) \lesssim |A|^3$, from the number of trivial quadruples and the fact that picking three elements determines the fourth respectively. Random sets will have energies around $|A^2|$, and arithmetic/geometric progressions have respective energies around $|A^3|$.

This is going to be useful for us, since multiplicative and 'divisive' energies are equivalent, i.e.,

$$E^\times(A) := \left|\left\{(a,b,c,d)\in A^4 : \frac{a}{c}=\frac{d}{b}\right\}\right|$$

**Lemma 17.3.** $E^\times(A)\geq \frac{|A|^4}{|A\times A|}$.

*Proof.* Let $h_1,\dots,h_{\mathcal{A}\cdot A}$ index the different products in $A\cdot A$. The number of pairs of elements sharing the same product is minimized when all products are represented a roughly equal number of times (this is Cauchy-Schwartz). More specifically, define

$$r_A^\times(h) = \left|\{(a,b)\in A\times A : a\cdot b = h\}\right|.$$

Then we have

$$|A|^4 = (|A\times A|)^2 = \left(\sum_{h\in A\cdot A} r_A^\times(h)\right)^2 \leq \left(\sum_{h\in A\cdot A} 1^2\right)\left(\sum_{h\in A\cdot A}(r_A^\times(h))^2\right) = |A\cdot A||E^\times(A)|,$$

using Cauchy Schwartz. □

We now complete the proof that $\max\{|A+A|,|A\cdot A|\}\gtrsim \frac{|A|^{4/3}}{\log^{1/3}|A|}$ from before. We showed previously that $|A/A|$ being small implied $|(A\times A)+(A\times A)|$ is large. Now, the goal is to show that $|A/A|$ being small implies $|(A\times A)+(A\times A)|$ is large. We see that $E^\times(A)=\sum_{j=1}^m \left||L_j\cap(A\times A)|\right|^2$.

From the proof of 17.1, recall

$$|(A\times A)+(A\times A)| \geq \sum_{j=1}^{k-1}\left|L_{i_j}\cap(A\times A)\right|\left|L_{i_{j+1}}\cap(A\times A)\right|.$$

We can regularize this this, losing a log factor using 'dyadic pigeonholing'

# Lecture 18

Last time we had the triangle inequality on $A,B,C\subset G$, this time we prove it.

$$|A-B|\leq \frac{|A-C||C-B|}{|C|}. \tag{18.1}$$

*Proof.* We'll define a map $M : (A-B) \times C \mapsto (A-C) \times (C-B)$. For an $x \in A-B$, choose representative functions $a(x) \in A, b(x) \in B$ so that $a(x) - b(x) = x$. Then our map is $M(x,c) = (a(x)-c, c-b(x))$. This map is injective, because we can recover $x$ and $c$ from the output: $x = (a(x)-c)+(c-b(x))$ and then $c = (c-b(x))+b(x)$. □

The following lemma was used in Ruzsa's simpler proof of Freiman's theorem.

**Lemma 18.1** (Ruzsa's covering)**.** *For any $A, B$, there exists an $X \subset B$ such that $B \subset (A-A)+X$ and $|X| \leq \frac{|A+B|}{|A|}$.*

*Proof.* Let $X$ be a maximal subset of $B$ such that $\{A+x : x \in X\}$ are disjoint. Let $b \in B$. Then note $A+X$ is intersected by $A+b$ otherwise $X$ wouldn't be maximal. So $b \in (A+X)-A$. We see $|A+X| \leq |A+B|$. □

**Corollary 18.2.** *If $|A+A| \leq k|A|$ then there is an $X : |X| \leq k^4 : 2A-A \subset A-A+X$.*

*Proof.* Define $B = 2A-A$, then there is an $X$ in there so that $|X| \leq \frac{|3A-A|}{|A|} \leq \frac{|4A||A+A|}{|A|^2} \leq k\frac{|4A|}{|A|}$. □

**Theorem 18.3** (Plünecke-Ruzsa(-Petridis))**.** *If $|A+B| \leq k|A|$ then there is an $X \subseteq A$ such that $|X+B+C| \leq k|X+C|$ for any set $C$.*

**Theorem 18.4** (Balog-Szemerédi-Gowers)**.** *Suppose you have a graph which has many edges, $|A| = |B| = n$ and $cn^2$ edges. $|A+B| \leq kn$. Then there are are $A' \subseteq A$, $B' \subseteq B$ such that $|A'| \geq \delta|A|$ and $|B'| \geq \delta|B|$ and $|A'+B'| \leq k'n$.*

**Theorem 18.5** (Freiman's theorem)**.** *(for bounded torsion group) $G$ is of torsion $r \neq 0$ if each $x \in G$ satisfies $rx = 0$.*

Note $(r-1)x = -x$. $\langle A \rangle = \bigcup_{s \geq 0} sA$. If $|A+A| \leq k|A|$ then $A$ is contained in a subgroup of $G$ of size at most $f(r,k)|A|$.

# Lecture 19*

**Theorem 19.1** (Solymosi)**.** *If $A \subseteq \mathbb{R}$, then $|A+A|^2|A \cdot A| \gtrsim \frac{|A|^4}{\log|A|}$.*

Assume $A$ is all positive, and cover $A \times A$ with lines $L_1, \ldots L_{|A/A|}$ of increasing slope. For any subset $i_1, \ldots i_k$, of these indices, recall

$$|A+A|^2 \geq \sum_{j=1}^{k-1} \left|L_{i_j} \cap (A \times A)\right| \left|L_{i_{j+1}} \cap (A \times A)\right|.$$

On the other hand,

$$E^{\times}(A) = \sum_{d \in A/A} \left| \left\{ (a,b) \in A \times A : \frac{b}{a} = d \right\} \right|^2 = \sum_{i \geq 1} \left|L_i \cap (A \times A)\right|^2.$$

To combine these things, we want to find some indices $i_1 < \cdots < i_k$ such that

$$\left|L_{i_j} \cap (A \times A)\right|^2 \approx \left|L_{i_j} \cap (A \times A)\right| \left|L_{i_{j+1}} \cap (A \times A)\right|.$$

So we use dyadic pigeonholing: write the energy as

$$E^{\times}(A) = \sum_{j=0}^{\lceil \log |A| \rceil} \sum_{2^j \leq |L_i \cap (A \times A)| < 2^{j+1}} |L_i \cap (A \times A)|^2.$$

So by pigeonhole principal, there is a $J$ where

$$\frac{E^{\times}(A)}{\log |A|} \lesssim \sum_{2^J \leq |L_i \cap (A \times A)| < 2^{J+1}} |L_i \cap (A \times A)|^2,$$

so let $i_1 < \cdots < i_k$ be the indices where $2^J \leq |L_{i_k} \cap (A \times A)| < 2^{J+1}$. In this case we have that $\left|L_{i_j} \cap (A \times A)\right| \leq 2 \left|L_{i_{j+1}} \cap (A \times A)\right|$, so then

$$\frac{E^{\times}(A)}{\log |A|} \lesssim \sum_{j=1}^{k} |L_{i_j} \cap (A \times A)|^2 \approx \sum_{j=1}^{k-1} |L_{i_j} \cap (A \times A)|^2,$$

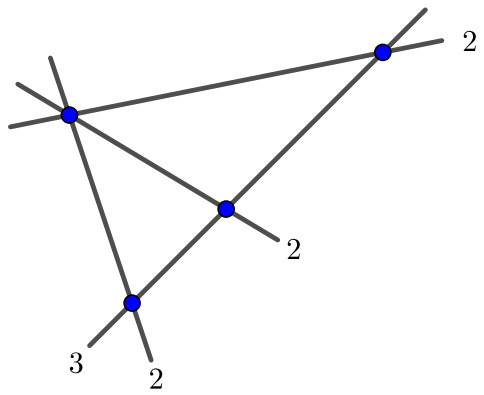
so we just threw away the top line, because of the pigeonholing, no single line is too large here so this loses at most a constant factor. We also assumed $k > 1$ doing this. If $k = 1$, $|L_{i_k} \cap (A \times A)|^2 \leq |A|^2 \leq |A + A|^2$ so we're good. Now,

$$\begin{aligned} \sum_{j=1}^{k-1} |L_{i_j} \cap (A \times A)|^2 &\lesssim \sum_{j=1}^{k-1} \left|L_{i_j} \cap (A \times A)\right| \left|L_{i_{j+1}} \cap (A \times A)\right|, \\ &\leq |A + A|^2. \end{aligned} \tag{19.1}$$

Combining that with $\frac{|A|^4}{|A \cdot A|} \leq E^{\times}(A)$ from before, we get the result.

## Incidence Geometry

Given a set of points $P$, and a set of curves $C$, an *incidence* is a pair $(p, c) \in P \times C$ such that $p \in C$. For example, in the following figure, there are 9 incidences. We can count this by visiting each line and counting how many points lie on it.




In this field, we are interested in upper bounds on the number of possible incidences for various types of $P$ and $C$. Denote by $I(P, C)$ the set of incidences.

**Theorem 19.2** (Szemerédi-Trotter)**.** *Let $P$ be a set of points, and $L$ a set of lines in $\mathbb{R}^2$. Then*

$$|I(P,L)| \lesssim |P|^{2/3}|L|^{2/3} + |P| + |L|.$$

An equivalent version of this theorem is in terms of $r$-rich lines (a line containing at least $r$ points of $P$.

**Theorem 19.3** (Szemerédi-Trotter (two alternate version))**.** *Let $P$ be a set of points in $\mathbb{R}^2$. The number of $r$-rich lines satisfies*

$$|L_r| \lesssim \frac{|P|^2}{r^3} + \frac{|P|}{r}$$

*Also, let $L$ a set of lines in $\mathbb{R}^2$. The number of $r$-rich points satisfies*

$$|P_r| \lesssim \frac{|L|^2}{r^3} + \frac{|L|}{r}$$

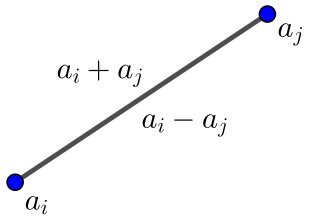
These versions are both equivalent to Theorem 19.2. We will first prove this in the special case that the point set is a cartesian product. Many applications only use this case, and the full proof works very similarly. For now, in fact we make two simplifying assumptions,

- assume $P$ is a catesian product, and
- assume "everything" is "evenly distributed".

So if $P = \sqrt{n} \times \sqrt{n}$

# Lecture 20

Today we studied something with surprising relevance to Kakeya, where we have an addition of elements in a set only along some graph edges $|A \overset{G}{+} A| \leq |A|$. I.e., the vertices of a graph are labelled with some $a_i$ and edges are labelled with $a_i + a_j$ if it joins those vertices. This is based on a paper of Katz-Tao from 2000.

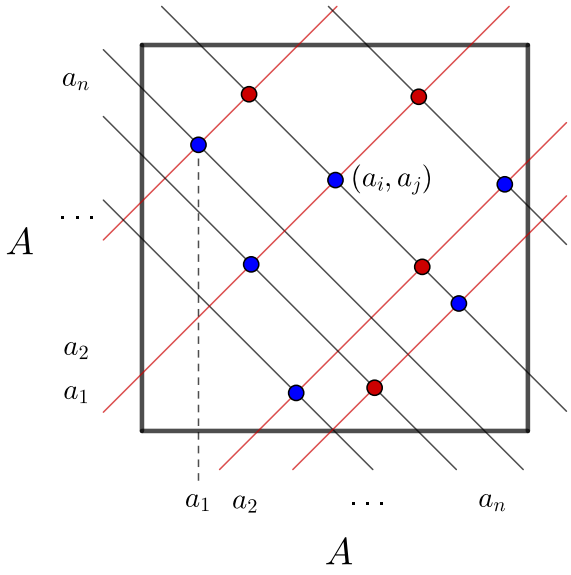



Trivially, we have $|A \overset{G}{-} A| \leq |e(G)|$. Today, we are going to show:

**Theorem 20.1.** *For any graph $G$ with labelled vertices, if $|A \overset{G}{+} A| \leq |A|$ then $|A \overset{G}{-} A| \leq |A|^{11/6}$.*

There is a construction showing that we can get an exponent of 1.75.

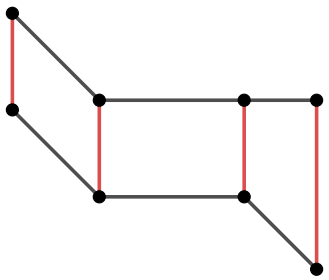
*Proof.* Consider $A \times A$. Collect the lattice points where $(a_i, a_j) \in e(G)$.




Looking at this product, the slope $-1$ lines cover all the sums, and the slope 1 lines cover all the differences. We select just one point from all of the slope 1 lines, and thus create a new graph $G' \subseteq G$ which has all unique differences.

If you have an $n \times n$ lattice with $m$ points in it, let $d_i$ be the number of points in column $i$. We have $m = \sum d_i$. By Cauchy Schwartz, the number of pairs sharing a column $\approx \sum d_i^2 \geq \frac{1}{n} m^2$. We use this fact to define a new object $W$ which comes from these selected points, it's the collection of all pairs of lattice points within one column.

We now consider a particular configuration of points that can appear in the grid.



We count the number of times it will appear in the point set. We know from a combinatorial counting (not shown today) that $\#(\text{config}) \geq \frac{|W|^4}{|A|^6}$. You can find how to do this in "Bounds on Arithmetic Projections, and Applications to the Kakeya Conjecture" by Katz and Tao. Use Lemma 2.1.

On the other hand, by fixing the parameters of one segment on the top left, and the coordinates of two points from the two rightmost segments, we completely fix the the structure. So we fixed one element of $W$ and two of $A$, so combining the bounds, we get $|W|^3 \leq |A|^8$. By the definition of $W$, we have

$$\left(\frac{|G'|^2}{|A|}\right)^3 \leq |W|^3 \leq |A|^8$$

and simplifying this leads to $|A \overset{G}{-} A| = |G'| \leq |A|^{11/6}$. □

# Lecture 21*

We are talking about 19.2 in cartesian products some more. We have a point set $P$ which is size $\sqrt{n} \times \sqrt{n}$, $L$ a set of $m$ lines, and $I$ incidences. On an average line, we have $\approx \frac{I}{m}$. A naive attempt is to count pairs of points in $P$. There are about $n^2$ pairs of points, and each line gives about $\left(\frac{I}{m}\right)^2$ pairs. so we have

$$\left(\frac{I}{m}\right)^2 \leq n^2,$$

and rearranging gives $n\sqrt{m}$. In the balanced case, this is an exponent of $\frac{3}{2}$, which is not good enough.

Instead, we want to count pairs of points at $x$ and $y$-distance at most $\frac{\sqrt{n}m}{I}$ (where the 'distance' is in units of the cartesian product itself). This is because the average distance between consecutive pairs of points is

$$\frac{\sqrt{n}}{\left(\frac{I}{m}\right)} = \frac{\sqrt{n}m}{I}.$$

Because of the 'even-ness' of the distribution, the number of close pairs on a given line is just about the number of points on that line. On the upper end, since this is a cartesian product, for every point we find the close pairs involving it, which is just the points in a box of height and width $\frac{\sqrt{n}m}{I}$. Combining both these bounds, we have

$$m\left(\frac{I}{m}\right) \leq \#\text{ of "close pairs"} \leq n\left(\frac{\sqrt{n}m}{I}\right)^2$$

rearranging this gives us the bound $I^3 \leq n^2m^2$,

**Theorem 21.1.** *Let $P = A \times B$ and $r \leq |A|, |B|$. The number of $r$-rich lines is $\lesssim \frac{|P|^2}{r^3}$.*

*Proof.* ... □

The Szemeredi-Trotter Theorem is tight (up the constants). To see this, we can construct a grid example. Take the $r \times \frac{n}{r}$ grid, and define

$$L = \left\{y = mx + b \,:\, 1 \leq m \leq \frac{n}{2r^2};\, 1 \leq b \leq \frac{n}{2r};\, m, b \in \mathbb{Z}\right\}$$

We see $|L| = \frac{n^2}{4r^3} \approx \frac{n^2}{r^3}$, and all of them hit $r$ points.

Using this, we can give another proof of Elekes' sum-product bound: $|A + A||A \cdot A| \lesssim |A|^{5/2}$. The following is more similar to his original proof.

*Proof.* Define the cartesian product $(A + A) \times (A \cdot A)$. Let $L = \{y = a_i(x - a_j) : a_i, a_j \in A\}$. For every $a_j + a_k$ with $a_k \in A$, we have an incidence, so these lines are all $|A|$-rich. There are $|A|^2$ lines, all $|A|$-rich, and $|A + A||A \cdot A|$ points. Szemeredi-Trotter completes the proof. □

# Lecture 23*

## Sum-product and incidence geometry over $\mathbb{F}_q$

Quick facts about $\mathbb{F}_q^2$:

- There are $q^2$ points.
- A line is just the solution set of a linear equation $y = mx + b$.
- There are $q + 1$ slopes, and thus $q(q + 1)$ lines total.
- Each line has $q$ points.

**Theorem 23.1** (Vinh)**.** *Let $P \subseteq \mathbb{F}_p^2$, and $L$ a set of lines. Then*

$$\left| I(P, L) - \frac{|P||L|}{q} \right| \leq q^{1/2} |P|^{1/2} |L|^{1/2} \tag{23.1}$$

The second term on the left side is the "expected" # of incidences. That is, there is a $1/q$ chance $p \in \ell$ for every $p, \ell$ pair, so this theorem says the number of incidences is not too far from the expected value, as long as the expected value is smaller than the error term. So the equation is useful when $|P||L| \geq q^3$.

*Proof.*

$$\begin{aligned}
\left| I(P, L) - \frac{|P||L|}{q} \right| &= \left| \sum_{\ell \in L} |\ell \cap P| - \frac{|P|}{q} \right| \\
&\leq \sum_{\ell \in L} \left| |\ell \cap P| - \frac{|P|}{q} \right| \cdot 1 \\
&\leq \left( |L| \sum_{\ell \in L} \left( |\ell \cap P| - \frac{|P|}{q} \right)^2 \right)^{1/2} \\
&\leq \left( |L| \sum_{\ell \in L} \left( |\ell \cap M| - \frac{|P|}{q} \right)^2 \right)^{1/2} = *
\end{aligned}$$

Where $M$ is the set of all lines in $\mathbb{F}_q^2$. We focus on the sum:

$$\sum_{\ell \in L} \left( |\ell \cap M| - \frac{|P|}{q} \right)^2 = \sum_{\ell \in L} |\ell \cap M|^2 - 2|\ell \cap M| \frac{|P|}{q} + \frac{|P|^2}{q^2} \tag{23.2}$$

The final term doesn't depend on $\ell$, it gives us $\frac{|P|^2 q(q+1)}{q^2}$. The first term can be counted as the number of pairs on lines. Distinct points determine a line, so the first term below comes from distinct points and the second from individual points (pairs of the same point).

$$\sum_{\ell \in L} |\ell \cap M|^2 = |P|(|P| - 1) + |P|(q + 1).$$

For the middle term is multiplied by $\frac{|P|}{q}$, not depending on $\ell$. Ignoring that, it counts the number of points on each line. We can switch that into counting the number of lines through each

point. This gives us: $\sum_{\ell \in L} -2|\ell \cap M| \frac{|P|}{q} = -2|P|(q+1)\frac{|P|}{q}$. Putting this all together, we have

$$\begin{aligned}
\sum_{\ell \in L} \left( |\ell \cap M| - \frac{|P|}{q} \right)^2 &= \frac{|P|^2 q(q+1)}{q^2} - 2|P|(q+1)\frac{|P|}{q} + |P|(|P|-1) + |P|(q+1) \\
&= |P|^2 + |P|q - \frac{|P|^2(q^2+q)}{q^2} \\
&= |P|q - \frac{|P|^2}{q} \\
&\leq |P|q
\end{aligned}$$

We can get rid of the extra term since $|P|q \geq \frac{|P|^2}{q}$ for all sizes of $P$. Going back to the main equation now,

$$* \leq (|L||P|q)^{1/2}$$

as claimed. □

Note that if $|P| = |L| = N$, this gives $|I(P,L)| \leq q^{1/2}N$ when $N \leq q^{3/2}$, and $|I(P,L)| \approx \frac{N^2}{q}$ when $N \geq q^{3/2}$. When $N \approx q^{3/2}$, the bound is $|I(P,L)| \leq N^{4/3}$.

Suppose $q = p^2$ and $N = p^2 = q$. Let $P = \mathbb{F}_p^2$, and $L = \{y = mx + b : m, b \in \mathbb{F}_p\}$. We have $N$ points and lines, and $N^{3/2}$ incidences (since there is $p = N^{1/2}$ points in each line). So When $N = p^2 = q$, this theorem is tight. For $q \leq N \leq q^{3/2}$, we can take translates of this. Take $\frac{N}{q}$ disjoint translates $\mathcal{P} + (0, a_i)$ for $1 \leq i \leq \frac{N}{q}$. Then there are $N$ point and lines, and $Np - Nq^{1/2}$ incidences. For $q \leq N \leq q^{3/2}$ and $q = p^2$ this is also sharp, but for example when $q = p$ we don't know.