Incidence Geometry

Tristan Shin

16–23 Oct 2024

These notes were written in preparation for a reading seminar on the polynomial method. They were mainly intended for me to use as a guide, so some details (in particular, diagrams), are omitted.

Main themes:

- Double-counting/extremal graph theory is very useful in incidence geometry.
- Topology of $\mathbb{R}^2$ is required to obtain best bounds.
- Lines are special somehow, other shapes might be more difficult to understand.
- Square grids have a lot of structure in incidence geometry.

Note: These notes mainly follow Chapter 7 of Larry Guth's *Polynomial Methods in Combinatorics*. Additionally, I also partially follow sections of Yufei Zhao's *Graph Theory and Additive Combinatorics*, particularly Section 1.4 and Chapter 8. I also use Larry's notes for his class: https://math.mit.edu/~lguth/PolynomialMethod.html.

1 Concurrencies

Question 1.1: Concurrencies problem

Let $\mathcal{L}$ be a set of L lines in $\mathbb{R}^2$. How many points can lie on $\geq r$ lines of $\mathcal{L}$?

Define

$$\begin{aligned} \mathcal{P}_r(\mathcal{L}) &\coloneqq \{x : x \text{ on } \geq r \text{ lines of } \mathcal{L}\} \\ P_r(L) &\coloneqq \max_{|\mathcal{L}|=L} |\mathcal{P}_r(\mathcal{L})| . \end{aligned}$$

Example 1.2: Stars config

Take N points, draw r lines through each ($N = Lr^{-1}$). Then $|\mathcal{P}_r(\mathcal{L})| = N = Lr^{-1}$.

Example 1.3

Take L generic lines, then you get $\binom{L}{2}$ points that lie on 2 lines. So $|\mathcal{P}_2(\mathcal{L})| \asymp L^2$.

Example 1.4

For $r = 3$, take an $n \times n$ square grid with NE diagonals. Then $L = 4n - 1$, and $|\mathcal{P}_3(\mathcal{L})| = n^2 \asymp L^2$.

Example 1.5: Grid config

Generalising, for any r, take an $n \times n$ square grid. Choose r different rational slopes of height $\lesssim r^{1/2}$ (check that this is possible using some NT bound).

Let $I(\mathcal{P}, \mathcal{L}) = \{(p, \ell) \in \mathcal{P} \times \mathcal{L} : p \in \ell\}$. Each point of the grid lies on r lines, so $|I(\mathcal{P}, \mathcal{L})| = n^2 r$. Each line contains $\gtrsim nr^{-1/2}$ points, so $|I(\mathcal{P}, \mathcal{L})| \gtrsim Lnr^{-1/2}$. Combining these gives that $n \gtrsim Lr^{-3/2}$, so $|\mathcal{P}_r(\mathcal{L})| \geq n^2 \gtrsim L^2 r^{-3}$.

Can we find better examples? It turns out, not really.

Theorem 1.6: Szemerédi–Trotter

$P_r(L) \lesssim L^2 r^{-3} + Lr^{-1}$

So in the regime $r \gtrsim L^{1/2}$, the stars configs is sharp; in the regime $r \lesssim L^{1/2}$, the grid config is sharp.

1.1 Simple bounds

Let us try proving bounds on $P_r(L)$.

For each $x \in \mathcal{P}_r(\mathcal{L})$, at least $\binom{r}{2}$ pairs of lines intersect at x. But over all x, these pairs must be distinct, so

$$|\mathcal{P}_r(\mathcal{L})| \cdot \binom{r}{2} \leq \binom{L}{2}.$$

This gives:

Proposition 1.7

$P_r(L) \lesssim L^2 r^{-2}$

This is not quite at the Szemerédi–Trotter bound yet. Let us try a different simple argument.

Take a subset $\mathcal{P}' \subseteq \mathcal{P}_r(\mathcal{L})$ of size t, where t will be chosen later. For each $x \in \mathcal{P}'$, there are at least $r - (t-1)$ lines of $\mathcal{L}$ that don't pass through another element of $\mathcal{P}'$. But over all x, these lines must be distinct, so

$$L \geq t \cdot (r - t + 1).$$

Now choose t such that $t \cdot (r - t + 1) > L$, and this gives a contradiction, implying that $|\mathcal{P}_r(\mathcal{L})| < t$. In particular, $t = \lceil 2Lr^{-1} \rceil$ gives:

Proposition 1.8

If $r \geq 2L^{1/2}$, then $P_r(L) < 2Lr^{-1}$.

This gives one part of Szemerédi–Trotter. How do we get the other part?

1.2 Crossing Lemma

So far, we have been using the Euclidean axiom of plane geometry: two lines intersect in at most one point. This cannot carry us to the full Szemerédi–Trotter, as the following example shows.

Example 1.9

Work in $\mathbb{F}_q^2$, where the Euclidean axiom still holds. Consider the grid config with the whole plane as our grid, and every possible (finite) slope. We get $L = q^2$, $r = q$, and $|\mathcal{P}| = q^2$. So $P_q(q^2) = q^2$, while Szemerédi–Trotter would imply that $P_q(q^2) \lesssim q$.

So we need to distinguish $\mathbb{R}^2$ and $\mathbb{F}_q^2$ somehow. We will use the topology of $\mathbb{R}^2$ via the following lemma.

Lemma 1.10: Crossing lemma

Let G be a graph with n vertices and m edges. If $m \geq 4n$, then

$$\operatorname{cr}(G) \geq \frac{m^3}{64n^2},$$

where $\operatorname{cr}(G)$ is the **crossing number** of G (minimum number of crossings of edges in a drawing of G in $\mathbb{R}^2$).

See the book for a proof of this lemma. The gist is to use Euler's formula ($f - e + v = 2$ on planar graphs) to prove that $\operatorname{cr}(G') \geq e - 3v$ for all graphs G'. Then take expected value over random induced subgraphs of G (put vertices in G' with probability $p = 4n/m$).

1.3 Proof of Szemerédi–Trotter

We can use the crossing lemma to prove Szemerédi–Trotter.

Theorem 1.11: Szemerédi–Trotter, effective

$P_r(L) \leq \max\{2Lr^{-1}, 256L^2r^{-3}\}$

Proof. Consider the graph G formed by taking $V = \mathcal{P}_r(\mathcal{L})$ and edges between vertices that are consecutive along a line. This graph has crossing number $\operatorname{cr}(G) \le \binom{L}{2}$ because any crossing must correspond to the intersection of two lines. Let us try applying the crossing lemma.

The number of vertices is $N = |\mathcal{P}_r(\mathcal{L})|$. Each vertex in the graph has degree $2r$, except for those which are the "first" or "last" along a line (because then the ray extending to infinity from that point is not an edge of the graph). So the sum of the degrees is $2rN - 2L$, so the number of edges is $rN - L$.

So as long as $rN - L \ge 4N$, we have that

$$\binom{L}{2} \ge \operatorname{cr}(G) \ge \frac{(rN-L)^3}{64N^2}.$$

The bound follows from analysis of various cases:

- If $r \le 7$, then use the first bound we obtained: $N \le \binom{L}{2}\binom{r}{2}^{-1} < \frac{49}{6}L^2r^{-3}$.
- If $rN - L \le \frac{1}{2}rN$, then $N \le 2Lr^{-1}$.
- If $rN - L < 4N$ and $r \ge 8$, then $N < \frac{L}{r-4} \le 2Lr^{-1}$.
- Otherwise, $rN - L \ge 4N$ and $rN - L > \frac{1}{2}rN$. Then

  $$\binom{L}{2} > \frac{r^3N^3/8}{64N^2},$$

  so $N < 256L(L-1)r^{-3}$. ∎

1.4 Remarks on the bound

Let us discuss some thoughts about the bound. Consider the same problem, but instead of using lines, use unit circles. With a little bit of work, one can show the same result as the line case. But our best construction is still based on the square grid, and the bound is far away from the construction. So to improve our understanding of the problem, we need to somehow distinguish lines and circles.

This problem becomes tougher because of the following thought experiment. Again consider the same problem, but now use unit parabolas (i.e. graphs $y = x^2 + ax + b$). Again, we get the same bound. But our understanding of this problem is actually quite good, because it is actually the same problem as the lines problem! To see why, consider the function $\Phi\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $(x, y) \mapsto (x, y + x^2)$. Then Φ is a bijection that maps lines to unit parabolas. So everything we know about the lines problem applies to the unit parabolas problem, and vice versa.

So we need to distinguish lines and circles, but we can't use some of the obvious things (e.g. algebraic degree, convexity, parameters, intersections).

We will very soon see the problem of point-line incidences, and we will later see the unit distance problem. The reason this discussion is useful is because the unit distance problem is equivalent of a version of the incidence problem for unit circles. So the difficulties of understanding the incidence geometry of circles will translate to difficulties in resolving the unit distance problem.

2 Incidences

Here is a closely related problem. Let $\mathcal{P}$ be a set of P points and $\mathcal{L}$ be a set of L lines. Recall that we defined $I(\mathcal{P}, \mathcal{L}) = \{(p, \ell) \in \mathcal{P} \times \mathcal{L} : p \in \ell\}$. Note that

$$\begin{aligned} |I(\mathcal{P}, \mathcal{L})| = \sum_{p \in \mathcal{P}} |\{\ell \in \mathcal{L} : p \in \ell\}| &= \sum_{\ell \in \mathcal{L}} |\{p \in \mathcal{P} : p \in \ell\}| \\ &= \sum_{\ell \in \mathcal{L}} |\ell \cap \mathcal{P}| . \end{aligned}$$

(We used this fact when analyzing the grid config.)

Question 2.1: Incidence problem

How big can $I(\mathcal{P}, \mathcal{L})$ be?

Proposition 2.2

$|I(\mathcal{P}, \mathcal{L})| \le \min\{P^2 + L, L^2 + P\}$

Proof. Split $\mathcal{L}$ into

$$\begin{aligned} \mathcal{L}_1 &= \{\ell \in \mathcal{L} : |\ell \cap \mathcal{P}| = 1\} \\ \mathcal{L}_{>1} &= \{\ell \in \mathcal{L} : |\ell \cap \mathcal{P}| > 1\} . \end{aligned}$$

Note that $|I(\mathcal{P}, \mathcal{L})| = |I(\mathcal{P}, \mathcal{L}_1)| + |I(\mathcal{P}, \mathcal{L}_{>1})|$. But $|I(\mathcal{P}, \mathcal{L}_1)| \le |\mathcal{L}_1| \le L$, and

$$|I(\mathcal{P}, \mathcal{L}_{>1})| = \sum_{p \in \mathcal{P}} |\{\ell \in \mathcal{L}_{>1} : p \in \ell\}| \le \sum_{p \in \mathcal{P}} (P - 1) = P^2 - P,$$

so $|I(\mathcal{P}, \mathcal{L})| \le P^2 + L$. The other inequality is similar; swap the roles of points and lines. ■

One can similarly get a slightly stronger bound.

Proposition 2.3

$|I(\mathcal{P}, \mathcal{L})| \le \min\{PL^{1/2} + L, LP^{1/2} + P\}$

Proof. Consider the bipartite point-line **incidence graph**, i.e. the bipartite graph with bipartition $\mathcal{P} \sqcup \mathcal{L}$ and edges corresponding to incidences. Then this graph is $K_{2,2}$-free because 2 lines intersect ≤ 1 time.

Count the number of $K_{2,1}$ in the graph with 2 vertices in $\mathcal{P}$ and one vertex in $\mathcal{L}$. This gives

$$L\binom{M/L}{2} \overset{\text{convexity}}{\leq} \sum_{\ell \in \mathcal{L}} \binom{\deg \ell}{2} \overset{K_{2,2}\text{-free}}{\leq} \binom{P}{2} \cdot 1,$$

so $M \leq L + PL^{1/2}$. Again, the other inequality is similar. ■

If we try to improve these bounds just using the Euclidean axiom again, we will struggle. As before, there is an example over finite fields where this bound is tight (in fact, this is the construction used to verify that $\operatorname{ex}(n, K_{2,2}) \asymp n^{3/2}$).

It turns out that this is because bounding $|I(\mathcal{P}, \mathcal{L})|$ is basically equivalent to bounding $P_r(L)$.

Theorem 2.4: Incidence Szemerédi–Trotter

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

We derive this from the first Szemerédi–Trotter that we proved earlier.

Proof. Use dyadic decomposition. For $j = 1, \ldots, k$, let

$$\mathcal{P}_j := \left\{p \in \mathcal{P} \,:\, p \text{ lies on } [2^{j-1}, 2^j) \text{ lines of } \mathcal{L}\right\} \subseteq \mathcal{P}_{2^{j-1}}(\mathcal{L}),$$

and let

$$\mathcal{P}_{\text{high}} := \left\{p \in \mathcal{P} \,:\, p \text{ lies on } \geq 2L^{1/2} \text{ lines of } \mathcal{L}\right\} \subseteq \mathcal{P}_{2L^{1/2}}(\mathcal{L}).$$

So letting $2^{k-1} < 2L^{1/2} \leq 2^k$, we get that

$$\mathcal{P} = \mathcal{P}_1 \cup \mathcal{P}_2 \cup \cdots \cup \mathcal{P}_k \cup \mathcal{P}_{\text{high}}$$

(not necessarily disjoint union), so

$$|I(\mathcal{P}, \mathcal{L})| \leq |I(\mathcal{P}_{\text{high}}, \mathcal{L})| + \sum_{j=1}^{k} |I(\mathcal{P}_j, \mathcal{L})|.$$

We can use the weaker incidence bound on the high term:

$$|I(\mathcal{P}_{\text{high}}, \mathcal{L})| \leq |\mathcal{P}_{\text{high}}| L^{1/2} + L.$$

But

$$|\mathcal{P}_{\text{high}}| \leq P_{2L^{1/2}}(L) < L^{1/2},$$

so $|I(\mathcal{P}_{\text{high}}, \mathcal{L})| < 2L$.

For the dyadic parts, note that each point in $\mathcal{P}_j$ is on at most 2^j lines, so

$$|I(\mathcal{P}_j, \mathcal{L})| \le 2^j |\mathcal{P}_j|.$$

But $|\mathcal{P}_j| \le P_{2^{j-1}}(L) \lesssim L^2 2^{-3j} + L2^{-j} \asymp L^2 2^{-3j}$ (since $L > 2^{2j-4}$), and $|\mathcal{P}_j| \le P$, so we can bound the j term by $2^j \min\{L^2 2^{-3j}, P\}$. Thus

$$\begin{aligned}|I(\mathcal{P}, \mathcal{L})| &\lesssim L + \sum_{j=1}^{k} 2^j \min\{L^2 2^{-3j}, P\} \\ &= L + \sum_{2^j < L^{2/3} P^{-1/3}} 2^j P + \sum_{2^j \ge L^{2/3} P^{-1/3}} L^2 2^{-2j}.\end{aligned}$$

The bound follows from analysis of various cases:

- If $L^{2/3} P^{-1/3} \le 1$, then the first sum disappears, and the second sum is $\asymp L^2 \le P$.
- If $L^{2/3} P^{-1/3} > 2^k$, then the second sum disappears, and the first sum is $\asymp 2^k P < L^{2/3} P^{2/3}$.
- If $1 < L^{2/3} P^{-1/3} \le 2^k$, then the first sum is $\asymp L^{2/3} P^{-1/3} \cdot P = L^{2/3} P^{2/3}$, while the second sum is $\asymp L^2 (L^{2/3} P^{-1/3})^{-2} = L^{2/3} P^{2/3}$. ■

Remark. This version also implies the first version, so these two versions are indeed equivalent. You can also prove this version directly using the crossing lemma, as one might expect.

Note that all of the bounds we proved were symmetric in P and L. In fact, you can convince yourself that this must be true using point-line duality.

2.1 An application to the sum-product problem

We can quickly apply Szemerédi–Trotter to a problem in additive combinatorics.

Question 2.5: Sum-Product Problem

Let A be a finite set of real numbers. How small can $A + A$ and $A \cdot A$ simultaneously be?

Think about a set with small doubling (e.g. arithmetic progression); it has terrible multiplicative structure. Think about a set with small multiplying ratio (e.g. geometric progression); it has terrible additive structure. The conjecture is that we are always somewhat close to one of these two, namely that

$$\max\{|A + A|, |A \cdot A|\} \ge |A|^{2-o(1)}.$$

For $A = [N]$, we have that $|[N] + [N]| \sim 2N$ but expect $[N] \cdot [N]$ to be large. One can use probabilistic number theory to show that $N^2/\log N \lesssim |[N] \cdot [N]| \lesssim N^2/\log_2 N$, so this would be tight. In fact, the exact order was shown by Ford to be

$$|[N] \cdot [N]| \asymp \frac{N^2}{(\log N)^{1-(1+\log_2 2)/\log 2}(\log_2 N)^{3/2}}.$$

We can prove a bound using Szemerédi-Trotter.

Proposition 2.6: Elekes

$|A + A| \, |A \cdot A| \gtrsim |A|^{5/2}$

Proof. Let $\mathcal{P} = (A + A) \times (A \times A) \subseteq \mathbb{R}^2$ and $\mathcal{L}$ be the set of lines through $(a_1, 0)$ with slope a_2 for $a_1, a_2 \in A$. We have that $P = |A + A| \, |A \cdot A|$ and $L = |A|^2$. By Szemerédi–Trotter, this implies that

$$|I(\mathcal{P}, \mathcal{L})| \lesssim |A + A|^{2/3} |A \cdot A|^{2/3} |A|^{4/3}$$

(using the fact that $|A + A|$ and $|A \cdot A|$ are between $|A|$ and $|A|^2$). But for each $\ell \in \mathcal{L}$, if ℓ passes through $(a_1, 0)$ with slope a_2, then ℓ passes through $(a_1 + a, a_2 a) \in \mathcal{P}$ for all $a \in A$, so $|\ell \cap \mathcal{P}| \geq |A|$. Thus

$$|I(\mathcal{P}, \mathcal{L})| \geq L \cdot |A| = |A|^3.$$

Combining these bounds gives the desired result. ■

Remark. You can also go directly from the first Szemerédi–Trotter to this result; see book for details. It's basically the same, which you can imagine because the two versions are equivalent.

Corollary 2.7

$\max\{|A + A|, |A \cdot A|\} \gtrsim |A|^{5/4}$

It turns out that this is not the best known bound. Solymosi used dyadic pigeonhole on **multiplicative energy** to prove that

$$|A + A|^2 |A \cdot A| \gtrsim \frac{|A|^4}{\log |A|},$$

so $\max\{|A + A|, |A \cdot A|\} \gtrsim |A|^{4/3 - o(1)}$. But still not at the conjectured correct exponent.

3 Unit distance problem

Here is another problem in incidence geometry.

Question 3.1: Unit distance problem

Consider a set $\mathcal{P}$ of N points. What is the maximum number of segments p_1p_2 with $p_1, p_2 \in \mathcal{P}$ such that $d(p_1, p_2) = 1$?

Example 3.2

Suppose $u(N)$ is the answer for N. Take an optimal set $\mathcal{P}_{N/2}$ of size $N/2$, make a copy, and translate by some generic unit vector. Then

$$u(N) \geq 2u(N/2) + N/2,$$

so $u(N) \gtrsim N \log N$.

Example 3.3

Take an $n \times n$ unit square grid ($N = n^2$). For suitable r that is a product of many 1 mod 4 primes, this grid has $N^{1+\Omega(1/\log\log N)}$ vectors of length $\sqrt{r}$.

It is conjectured that this square grid example is close to sharp. We expect that $u(N) \asymp N^{1+o(1)}$. Can we prove any bounds towards this?

Proposition 3.4: Erdős

$u(N) \lesssim N^{3/2}$

Proof. Consider the **unit distance graph**, i.e. the graph whose vertices are the points of $\mathcal{P}$ and whose edges correspond to unit distances. Then this graph is $K_{2,3}$-free because 2 circles intersect ≤ 2 times.

Count the number of $K_{2,1}$ in the graph. This gives

$$N\binom{2M/N}{2} \overset{\text{convexity}}{\leq} \sum_v \binom{\deg v}{2} \overset{K_{2,3}\text{-free}}{\leq} \binom{N}{2} \cdot 2,$$

so $M \lesssim N^{3/2}$. ∎

Remark. This analysis generalises to show that if any graph G is $K_{s,t}$-free, then $e(G) \lesssim_{s,t} n^{2-1/s}$ (Kővári–Sós–Turán). You need to be careful when doing convexity, because $\binom{x}{s}$ is not convex. It is for $x > s-1$ though, so just zero out the $x < s-1$ region.

This is not at the conjectured bound yet. We can try applying the crossing lemma as we did with the previous incidence geometry problems.

We would like to construct a graph by drawing unit circles centered at points in $\mathcal{P}$ and have edges correspond to arcs of circles between points of $\mathcal{P}$. This runs into two issues: if a circle only has one point, we get self-loops; and there might be multiple "edges" between two vertices (e.g. if there are only two points on a circle, or if there are two circles through two points). It turns out that these issues are manageable.

Lemma 3.5: Multigraph crossing lemma

Let G be a **multigraph** with n vertices, m edges, and at most k edges between any two vertices. If $m \geq 4kn$, then

$$\operatorname{cr}(G) \geq \frac{m^3}{64k^3n^2}.$$

Proof. Let $\tilde{G}$ be the (standard) graph formed by contracting all edges between two vertices into one edge. Then $e(\tilde{G}) \geq m/k$, and obviously $\operatorname{cr}(G) \geq \operatorname{cr}(\tilde{G})$. So the standard crossing lemma on $\tilde{G}$ gives this bound. ∎

Using this, we can prove a stronger bound on $u(N)$.

Proposition 3.6: Spencer–Szemerédi–Trotter

$u(N) \lesssim N^{4/3}$

Proof. Let G be the multigraph whose vertices are points of $\mathcal{P}$, and edges are between two consecutive (distinct) points along a unit circle centered at a point of $\mathcal{P}$. Then this multigraph has at most 4 edges between any two vertices.

How many edges are in this multigraph? Let u be the number of unit distances. If we allowed self-loops (i.e. consider *all* circles with a point on them, not just those with ≥ 2), then there are $2u$ edges, by associating each unit distance with two clockwise arcs at each endpoint. We delete at most N of these edges by removing the self-loops, so there are at least $2u - N$ edges in G.

If $u < 9N$ then we are clearly good, so assume $u \geq 9N$. Then $2u - N \geq 16N$ and $2u - N \geq u$, so the multigraph crossing lemma gives that

$$\operatorname{cr}(G) \geq \frac{(2u-N)^3}{4096N^2} \geq \frac{u^3}{4096N^2}.$$

But the edges are along N circles, and each pair of circles intersects at most twice, so $\operatorname{cr}(G) \leq 2\binom{N}{2}$, so

$$2\binom{N}{2} \geq \frac{u^3}{4096N^2}.$$

The conclusion follows. ∎

Remark. The earlier discussion about parabolas is extremely relevant here. Recall that the Szemerédi–Trotter bounds for point-line incidences are tight. Further recall that the incidence geometry of lines transfers to that of unit parabolas via the bijection Φ. So the number of incidences between N points and N unit parabolas is bounded by $N^{4/3}$. The Spencer–Szemerédi–Trotter result is saying the same but with unit circles. But the unit parabola bound is tight, while we expect the unit circle bound to be far from the correct answer (which we expect to be $N^{1+o(1)}$). So to improve this bound, we need some understanding of what differentiates circles and parabolas here.

4 Distinct distances problem

This problem is closely related to the unit distance problem.

Question 4.1: Distinct distances problem

Consider a set $\mathcal{P}$ of N points. What is the minimum number of values in the set $\{d(p_1, p_2) : p_1, p_2 \in \mathcal{P}^2\}$?

To see why this is closely related to the previous problem, let $d(N)$ be the answer to this problem. Take an optimal set $\mathcal{P}_N$ of size N. Then each of the $d(N)$ distances appears at most $u(N)$ times, so $\binom{N}{2} \leq d(N)u(N)$. Thus, we get a bound on $d(N)$ immediately for free.

Proposition 4.2: Spencer–Szemerédi–Trotter

$d(N) \gtrsim N^{2/3}$

What about constructions? It turns out that the square grid is again our best construction, this time with $\lesssim N/\sqrt{\log N}$ distinct distances. It is conjectured that this is sharp.

We can actually improve on the previous bound by again applying the crossing lemma.

Proposition 4.3: Székely

$d(N) \gtrsim N^{4/5}$

See the book for a sketch of the proof.

But this is actually not the best that we know. This crossing lemma technique was pushed further by Katz and Tardos to improve the exponent to 0.864. But the best known bound by Guth and Katz uses the polynomial method (which is what we are mainly studying in this book) to show that $d(N) \gtrsim N/\log N$.