

NAMBOODIRI LECTURE 2: THE DISTINCT DISTANCE PROBLEM

1. OPEN QUESTIONS IN INCIDENCE GEOMETRY

Last time, we introduced incidence geometry and discussed a fundamental theorem. Let us recall this result.

Let \mathfrak{L} denote a set of L lines in \mathbb{R}^2 .

We consider the r -rich points of a configuration of lines \mathfrak{L} :

$$P_r(\mathfrak{L}) := \{x \in \mathbb{R}^2 \mid x \text{ lies in at least } r \text{ lines of } \mathfrak{L}\}.$$

Question 1. *For given L and r , estimate the maximum possible number of r -rich points that can be formed by L lines:*

$$\max_{\mathfrak{L} \text{ a set of } L \text{ lines}} |P_r(\mathfrak{L})|.$$

We saw that a grid configuration gives $\sim L^2 r^{-3}$ r -rich points, and a trivial star configuration gives $\sim L r^{-1}$ r -rich points. In the early 1980's, Szemerédi and Trotter proved that these examples are optimal up to a constant factor:

Theorem 2. ([ST], Szemerédi-Trotter) *For any configuration \mathfrak{L} of L lines in \mathbb{R}^2 ,*

$$|P_r(\mathfrak{L})| \leq C \max(Lr^{-1}, L^2 r^{-3}) = (\text{ST}).$$

Last lecture we introduced the cutting method, and we sketched the proof of Theorem 2.

We can pose variations on Question 1 by replacing lines with other curves in the plane. What happens for circles? Unit circles? Parabolas? Ellipses? These are four major open problems of incidence geometry!

Let us try to explain why these problems are hard, or at least why they cannot be solved by the same method that worked for lines. The short answer is that the cutting method gives some upper bounds in all four of these problems, but the bounds do not match any examples. We focus on the case of unit circles, which is the most intensively studied.

Suppose that Γ is a set of L unit circles. The bound for $|P_r(\Gamma)|$ from the cutting method is exactly the same as for lines. The cutting method combines a topological argument and a double counting argument, and both of these arguments work just the same for lines and for unit circles. In both cases it is easy to make a stars example with Lr^{-1} r -rich points. But the grid example does not work as well for circles. The problem is that a circle cannot contain very many lattice points. Given an $S \times S$ grid of points, there are lines that contain S grid points, but the number of grid points in any circle is $O(S^\epsilon)$ for any $\epsilon > 0$. We summarize these bounds and examples in the following table:

	lines	unit circles
Cutting method upper bound on $ P_r(\Gamma) $	(ST)	(ST)
Examples	stars: Lr^{-1}	stars: Lr^{-1}
	grid: L^2r^{-3}	grid: $\lesssim L^{1+\epsilon}r^{-1}$

This table suggests another question. Do mathematicians know of any other configurations of lines, besides the grid example, that produce a lot of r -rich points? For small values of r there are some other examples. For $r = 3$, there is an example involving elliptic curves which is actually slightly better than the grid and which goes back to the 19th century (cf. [GT]). For $r = 4$, only very recently, Solymosi and Stojaković [SS] constructed a really different example which is nearly as good as a grid. People had been working on this question for decades. Their construction works for other small values of r , but it stops working if r is a power of L . In the range $L^{.01} \leq r \leq L^{.49}$, all the currently known examples are based on grids.

(We should mention here that the problem is invariant under projective transformations, because projective transformations map lines to lines. So we can take a grid example, and modify it with a projective transformation to get an example which is not literally a grid. We will still refer to such a configuration as a “grid example”.)

It’s striking how few examples we know. It could be that there are more exotic examples we haven’t found yet, or it could be that integer grids are essentially the only examples when r is reasonably large. Understanding this issue is another major open problem in incidence geometry. If we did understand this issue, it might well help to understand the problem of unit circles.

(The cutting of edge of research in this direction is a recent paper of Green and Tao, [GT], which determines the exact maximum value of $|P_3(\mathcal{L})|$ for a set of L lines. They prove that the elliptic curve example mentioned above gives the exact maximum value of $|P_3(\mathcal{L})|$, and they also prove that if $|P_3(\mathcal{L})|$ is very close to this maximum value, then \mathcal{L} must have an elliptic curve structure.)

That gives a sense of some of the main open problems in the field. I wanted to say a little now about the history of these problems. The beginning of this field of inquiry owes a lot to Paul Erdős and his collaborators. They were attracted to questions that were simple to state but surprisingly hard to resolve. Erdős began his career in number theory, studying prime numbers and diophantine equations. One of the cool features of this part of number theory is that there are really deep problems that can be explained to a high school student. When Erdős was beginning his career, in the early 20th century, this probably seemed like a unique feature of number theory. Erdős worked hard to find other sources of simply stated but really hard problems, and over his career he found several sources of such problems, with a very different flavor from the examples in number theory. The problems we just discussed are one of the new areas that Erdős helped develop. Once a new source of problems is found, then many people can join in and work on variations. In my view, finding a really new source of hard simple problems is a big achievement. It marks out an interesting direction, and it can be the source of a new field of mathematics.

2. THE DISTINCT DISTANCE PROBLEM

Erdős’s first paper about incidence geometry is a 1946 paper [E] on the distances between points in a finite set. Suppose $X \subset \mathbb{R}^2$ is a set of N points. Let $d(X)$ denote the set of distances between points of X :

$$d(X) := \{|p_1 - p_2|\}_{p_1, p_2 \in X; p_1 \neq p_2}.$$

The distinct distance problem asks about the minimum possible number of distinct distances among all sets of N points.

If X is a set of N points evenly spaced along a line, then $|X| = N - 1$. If X is a set of N points arranged in a square grid, then the set of distinct distances is a little smaller. Using some number theory, [E] computes that $|d(X)| \sim N(\log N)^{-1/2}$. This is the smallest known example, and Erdős conjectured that it is optimal up to a constant factor. But he could only prove a much weaker bound: $|d(X)| \geq cN^{1/2}$. Many people worked on this problem, and the exponent $1/2$ was improved many times. For instance, ideas connected with the Szemerédi-Trotter theorem led to substantial improvements, but they didn't solve the problem. Before the polynomial method, the best proven lower bound was $|d(X)| \geq cN^{.864}$. The polynomial method has led to a lower bound which is sharp up to log factors:

Theorem A. *For any set $X \subset \mathbb{R}^2$ of N points,*

$$|d(X)| \geq cN(\log N)^{-1}.$$

(This theorem was proven by Nets Katz and myself [GK] building on some ideas by Elekes-Sharir [ES].)

Let us try to visualize a set X with $|d(X)|$ much smaller than N . Let p be a point in X , and consider circles centered at p with each radius in $d(X)$. All the other points of X lie in these circles. Since the number of circles is much smaller than N , each circle must contain many points of X . See Figure 1a for a picture of this situation. Now let p_2 be a second point of X , and draw a second family of circles centered at p_2 . Every point of X also lies in this second family of circles, so the points of X all lie at the intersections of the first two families of circles. See Figure 1b for a picture of this situation. Now consider a third point $p_3 \in X$ and draw a third family of circles centered at p_3 . Every point of X needs to lie in one circle from each family. Unless we choose the first two families very carefully, it will not be possible to satisfy this condition. And we have only considered circles centered at three points of X – we still have another $N - 3$ centers to go!

This discussion shows how the distinct distance problem is related to the intersection patterns of circles in the plane. It is connected with some of the major open problems we discussed in Section 1, and indeed some of those problems were invented to help understand the distinct distance problem.

Our goal for the rest of the lecture is to describe the main ideas of the proof of Theorem A. There are two main steps in the proof. The first step is a novel way of getting started on the problem that was developed by Elekes and Sharir. Instead of thinking about circles in the plane, they are able to connect the distinct distance problem to a problem about lines in \mathbb{R}^3 in the spirit of the joints problem. The second step is to solve this problem using polynomials.

- (1) Using symmetries (relates the problem to lines in \mathbb{R}^3).
- (2) Polynomial partitioning (to study lines in \mathbb{R}^3).

3. USING SYMMETRIES

The distinct distance problem is connected to incidence problems about circles in the plane, but these problems are still wide open and look very hard. Elekes and Sharir proposed a different way to get started, which connects to questions about lines in \mathbb{R}^3 .

Suppose that X is a set of N points in \mathbb{R}^2 with few distinct distances. It follows that the same distance must occur many times. In other words, there must be many solutions to the equation $|p_1 - q_1| = |p_2 - q_2|$ where $p_i, q_i \in X$. We let $Q(X)$ denote the set of such distance quadruples:

$$Q(X) := \{(p_1, q_1, p_2, q_2) \in X^4 \text{ so that } |p_1 - q_1| = |p_2 - q_2| \neq 0\}.$$

The distance problem is connected to the size of $|Q(X)|$, so it's natural to ask how big $|Q(X)|$ can be. To prove Theorem A, we first establish a bound on $|Q(X)|$: $|Q(X)| \lesssim N^3 \log N$. (The number $N^3 \log N$ is the number of quadruples in a square grid.) It's not hard to deduce Theorem A from this bound. But how can we estimate $|Q(X)|$? Elekes and Sharir gave a very interesting way to estimate $|Q(X)|$, involving the symmetry group of the problem.

Let G denote the group of orientation-preserving rigid motions of the plane.

Lemma 3. *If $|p_1 - q_1| = |p_2 - q_2| \neq 0$, then there is a unique $g \in G$ so that*

$$g(p_1) = p_2 \text{ and } g(q_1) = q_2.$$

Proof idea. First consider the translation sending p_1 to p_2 . We can get any rigid motion g with $g(p_1) = p_2$ by composing this translation with a rotation around p_2 . Since $|p_1 - q_1| = |p_2 - q_2| \neq 0$, there is exactly one such rotation which maps q_1 to q_2 . (On the other hand, if $|p_1 - q_1| \neq |p_2 - q_2|$, then there is no rigid motion g taking p_1 to p_2 and q_1 to q_2 .)

So every quadruple in $Q(X)$ corresponds to a unique rigid motion $g \in G$. We are going to try to view the problem from the point of view of G instead of \mathbb{R}^2 .

For any points $p_1, p_2 \in \mathbb{R}^2$, define

$$S(p_1, p_2) := \{g \in G \mid g(p_1) = p_2\}.$$

As we saw in the proof sketch above, $S(p_1, p_2)$ is a 1-dimensional curve: it consists of taking the translation that sends p_1 to p_2 and composing with a rotation around p_2 . So $S(p_1, p_2)$ is a 1-dimensional curve inside of the 3-dimensional Lie group G .

The key point is that $(p_1, q_1, p_2, q_2) \in X^4$ is a distance quadruple if and only if the curves $S(p_1, p_2)$ and $S(q_1, q_2)$ intersect each other. In Figure 2, we illustrate a distance quadruple from the point of view of \mathbb{R}^2 and then from the point of view of G .

Let S_X denote the set of all curves $S(p_1, p_2)$ with $p_1, p_2 \in X$. So S_X is a set of $L = N^2$ curves in G . The quadruples $Q(X)$ correspond to intersections between the curves of S_X . So to estimate $|Q(X)|$, we want to prove a bound of the form

$$|P_2(S_X)| \lesssim N^3 = L^{3/2}.$$

(It can happen that a rigid motion g lies in more than two of the curves of S_X . If g lies in r curves of S_X , it means geometrically that $|X \cap gX| = r$. Such an r -rich rigid motion corresponds to many quadruples: about r^2 quadruples. So to bound the number of quadruples, we actually need to bound $|P_r(S_X)|$ for all r . In this lecture, we just focus on $r = 2$, but the ideas that we discuss also apply to other values of r .)

There is a clever choice of coordinates on the group G in which every curve of the form $S(p_1, p_2)$ is a straight line. In these coordinates, S_X is a set of L lines in \mathbb{R}^3 , which we denote \mathfrak{L}_X , and we want to show a bound of the form

$$|P_2(\mathfrak{L}_X)| \lesssim L^{3/2}.$$

This approach gives a very different way of getting started on the distinct distance problem, and it connects it to the incidence geometry of lines in \mathbb{R}^3 . The clever change of coordinates we just mentioned is convenient, but it is not the heart of the matter. If we use less clever coordinates, then

the curves S_X will be algebraic curves of degree 2. All of the ideas in the next section would work as well for algebraic curves of fixed degree, but they are somewhat easier to explain for straight lines. So in the next section, we return to the incidence geometry of lines in \mathbb{R}^3 .

4. LINES IN \mathbb{R}^3 AND POLYNOMIAL PARTITIONING

Given a set X of N points in the plane, we associated a set \mathfrak{L}_X of $L = N^2$ lines in \mathbb{R}^3 . To prove the distinct distance estimate, it is necessary to establish a bound of the form

$$|P_2(\mathfrak{L}_X)| \leq CL^{3/2}.$$

A set of L lines in the plane may have $\binom{L}{2} \sim L^2$ 2-rich points. This can happen if all L lines lie in a plane. So we should check if all the lines of \mathfrak{L}_X can lie in a plane. It turns out that at most $N = L^{1/2}$ lines of \mathfrak{L}_X can lie in any plane.

This leads us to try to estimate

$$\max_{|\mathfrak{L}|=L, \text{ at most } L^{1/2} \text{ lines in any plane}} |P_2(\mathfrak{L})|.$$

Surprisingly, this maximum is still on the order of L^2 . This happens in a beautiful example involving a degree 2 algebraic surface. The surface $z = xy$ contains many lines. For any $b \in \mathbb{R}$, it contains a “horizontal line” H_b parametrized by

$$t \mapsto (t, b, tb).$$

Also, for any $a \in \mathbb{R}$, it contains a “vertical line” V_a parametrized by

$$t \mapsto (a, t, at).$$

For any $a, b \in \mathbb{R}$, the lines H_b and V_a intersect at the point (a, b, ab) . Now if we let \mathfrak{L} be a union of $L/2$ vertical lines and $L/2$ horizontal lines, then $|P_2(\mathfrak{L})| = L^2/4$. On the other hand, because all the lines lie in the surface $z = xy$, it is impossible for any plane to contain many lines. The intersection of any plane with this surface is a degree 2 curve, and so it can contain at most two lines.

Based on this example, we should check if all the lines of \mathfrak{L}_X can lie in a degree 2 surface. It turns out that any degree 2 algebraic surface contains at most $100L^{1/2}$ lines of \mathfrak{L}_X . More generally, it turns out that any degree D algebraic surface contains at most $C_D L^{1/2}$ lines of \mathfrak{L}_X . (The fact that not too many lines of \mathfrak{L}_X lie in a low degree surface is not hard to check. Given a particular surface $Z(P) \subset \mathbb{R}^3$, the condition that a certain number of lines of \mathfrak{L}_X lie in $Z(P)$ can be translated into a condition about the set X . Given any threshold M , it is fairly straightforward to describe all the sets X so that \mathfrak{L}_X contains at least M lines in some degree 2 surface.)

If \mathfrak{L} contains few lines in any plane or any low degree surface, can we get a better estimate for $|P_2(\mathfrak{L})|$?

Theorem 4. ([GK]) *If \mathfrak{L} is a set of L lines in \mathbb{R}^3 with at most $100L^{1/2}$ lines in any plane or degree 2 algebraic surface, then*

$$|P_2(\mathfrak{L})| \leq CL^{3/2}.$$

This theorem gives the desired bound on $|P_2(\mathcal{L}_X)|$. This question was not particularly investigated before the stimulus by Elekes and Shair, who connected it to the distinct distance problem, but I think that it's a really natural question. It can be interpreted as a structure theorem for configurations of lines with many 2-rich points: it implies that if $|P_2(\mathcal{L})|$ is much bigger than $L^{3/2}$, then there must be some planes and/or degree 2 surfaces which account for most of the 2-rich points. Because of the example above with the surface $z = xy$, we see that degree 2 surfaces are necessary in the statement of this theorem. This degree 2 example shows that algebraic surfaces play an important role in the possible intersection patterns of lines in \mathbb{R}^3 , and this helps to motivate the polynomial method.

In the rest of the lecture I want to describe how to use polynomials to prove this theorem. The proof I will describe is based on polynomial partitioning, which we discussed at the end of Lecture 1. We use a polynomial surface to divide space into pieces, and then add up the 2-rich points in each piece.

We illustrate a polynomial partitioning in Figure 3. The red surface is $Z(P)$. The complement of $Z(P)$ is a union of cells O_i . If P has degree D , then there can be up to $\sim D^3$ cells. (This happens, for instance, if $Z(P)$ is a union of D planes in general position.) Each line enters at most $D + 1$ cells. Therefore, if the lines are evenly distributed, each cell would intersect about $D^{-2}L$ lines.

There is an equipartition theorem which says that we can choose the polynomial P to get this bound on the number of lines in each cell.

Theorem 5. ([GK], [G]) *If \mathcal{L} is a set of L lines in \mathbb{R}^3 and $D \geq 1$, then there exists a non-zero $P \in \text{Poly}_D(\mathbb{R}^3)$ so that $\mathbb{R}^3 \setminus Z(P)$ is a disjoint union of $\sim D^3$ open cells O_i , and each cell intersects at most $CD^{-2}L$ lines of \mathcal{L} .*

This equipartition theorem is a fancy version of the ham sandwich theorem. The classical ham sandwich theorem, proven by Banach in the 1930's, says that given three finite volume open sets $U_1, U_2, U_3 \subset \mathbb{R}^3$, there is a plane that bisects all three sets. Note that the set of planes in \mathbb{R}^3 forms a 3-parameter family, and by counting dimensions, it is at least plausible that we could tune those parameters to bisect three sets. The proof uses ideas from topology. In the early 1940's, Banach's ham sandwich theorem was greatly generalized by Stone and Tukey. They proved a version of the ham sandwich theorem in higher dimensions, but they also proved a generalization of the ham sandwich theorem where planes are replaced by other types of surfaces, such as algebraic hypersurfaces of a given degree. The space of algebraic surfaces in \mathbb{R}^3 of degree at most D has dimension $\text{Dim Poly}_D(\mathbb{R}^3) - 1 \sim D^3$. By counting dimensions, it is at least plausible that we could tune the parameters in order to divide something into $\sim D^3$ equal pieces. Algebraic topology, especially the work of Banach, Stone and Tukey, provides tools to make this rigorous. Using those tools, it's not hard to establish Theorem 5.

The full proof of Theorem 4 is pretty complicated, but to give a flavor we state and prove a weaker version of Theorem 4. For this weaker version, we can give a one page proof using polynomial partitioning.

Theorem 6. *If \mathcal{L} is a set of L lines in \mathbb{R}^3 with at most S lines in any surface of degree at most D , then*

$$|P_2(\mathcal{L})| \leq K(D, S)L^{\frac{3}{2} + \epsilon(D)}, \quad (*)$$

and $\lim_{D \rightarrow \infty} \epsilon(D) = 0$.

Here $K(D, S)$ is a constant depending on D and S , which works out to $2(D + S^{1/2})$.

If D and S are large fixed constants and $L \rightarrow \infty$, then we get a bound of the form $|P_2(\mathfrak{L})| \lesssim L^{\frac{3}{2}+\epsilon}$ for a tiny $\epsilon > 0$, which is nearly as good as Theorem 4. The main improvement which is necessary for the distinct distance problem is to get good bounds when S is around $L^{1/2}$, which is quite large. If we plug in $S = L^{1/2}$ in our weak theorem, we get a bound of the form $|P_2(\mathfrak{L})| \lesssim L^{\frac{7}{4}+\epsilon}$, which is too weak to give a sharp estimate in the distinct distance problem. This improvement takes some further work, but Theorem 6 still gives a good sense of how polynomial partitioning works, and it gives the flavor of the proof of Theorem 4. See [GK] or [G2] for full proofs.

Proof of Theorem 6. The proof is by induction on the number of lines. There are three types of points $x \in P_2(\mathfrak{L})$.

Type 1. x is in one of the cells O_i .

Type 2. $x \in Z(P)$, but one of the lines through x is not contained in $Z(P)$.

Type 3. $x \in Z(P)$, and all the lines through x lie in $Z(P)$ also.

The three types of intersection are illustrated in Figure 4.

We let RHS denote the right-hand side of (*). The points of type 2 and type 3 are easy to bound. Since each line intersects $Z(P)$ in at most D points,

$$|\text{Type 2}| \leq DL.$$

By assumption there are at most S lines of \mathfrak{L} in any degree D surface, and so we know that there are at most S lines of \mathfrak{L} in $Z(P)$. Therefore,

$$|\text{Type 3}| \leq S^2 \leq S^{1/2} L^{3/2}.$$

Motivated by this, we define $K(D, S) = 2(D + S^{1/2})$. It follows easily that

$$|\text{Type 2} + \text{Type 3}| \leq \frac{1}{2} K(D, S) L \leq \frac{1}{2} \text{RHS}.$$

Finally, we have to bound $|\text{Type 1}|$, which we do by induction on the number of lines. We let \mathfrak{L}_i be the set of lines of \mathfrak{L} that intersect O_i . We know that $|\mathfrak{L}_i| \leq CD^{-2}L$. Therefore,

$$|\text{Type 1}| \leq \sum_i |P_2(\mathfrak{L}_i)| \leq CD^3 \max_i |P_2(\mathfrak{L}_i)|.$$

We can bound $|P_2(\mathfrak{L}_i)|$ by using induction. Now we get

$$|\text{Type 1}| \leq CD^3 K(D, S) (CD^{-2}L)^{\frac{3}{2}+\epsilon(D)} = CD^{-2\epsilon(D)} \text{RHS}.$$

To make the induction close, we need to choose D and ϵ so that $CD^{-2\epsilon} \leq 1/2$. The constant C is an absolute constant (independent of D or ϵ), and so as $D \rightarrow \infty$, we can choose $\epsilon(D) \rightarrow 0$. □

REFERENCES

- [ES] Gy. Elekes and M. Sharir, Incidences in three dimensions and distinct distances in the plane, Proceedings 26th ACM Symposium on Computational Geometry (2010) 413-422.
- [E] P. Erdős, On sets of distances of n points, Amer. Math. Monthly (1946) 53, 248-250.
- [GT] B. Green and T. Tao, On sets defining few ordinary lines, Disc. Comput. Geom., see also arXiv:1208.4714.
- [G] L. Guth, Polynomial partitioning for a set of varieties, arXiv:1410.8871
- [G2] L. Guth, Distinct distances estimates and low degree polynomial partitioning, accepted by Disc. Comput. Geom., see also arXiv:1404.2321
- [GK] L. Guth and N. Katz, On the Erdős distinct distance problem in the plane, Annals of Math 181 (2015), 155-190. See also arXiv:1011.4105.

- [SS] J. Solymosi and M. Stojaković, Many collinear k -tuples with no $(k+1)$ collinear points, *Discrete and Comp. Geom.*, (2013) 50, 811-820.
- [ST] E. Szemerédi and W. T. Trotter Jr., Extremal Problems in Discrete Geometry, *Combinatorica* (1983) 3, 381-392.