THE ERDŐS DISTANCE PROBLEM: LECTURE NOTES

JULIA GARIBALDI AND ALEX IOSEVICH

Abstract. In these notes we describe many of the known advances on the Erdős distance problem in a fashion suitable for undergraduates and advanced high school students. An expanded version of these notes will become a book by the end of the summer. The book will be purely combinatorial and self-contained.

Introduction

These notes were written for the summer program on the Erdős distance problem, to be held at the University of Missouri, August 1-5, 2005. This is the second year of this program and our plan continues to be to introduce motivated high school students to accessible concepts of higher mathematics. Last year's theme was the Kakeya conjecture in finite fields. This year we concentrate on one of the most beautiful problems of geometric combinatorics, the Erdős distance conjecture.

The notes are heavily problem oriented. Most of the learning is meant to be done by doing the exercises interspersed throughout the lecture notes. Many of these exercises are recently published results by mathematicians working in the area. In a couple of places, steps are intentionally left out of proofs and the reader is then asked to fill them in in the process of working the exercises. On a number of occasions, solutions to exercises are used in later chapters in an essential way. Having said that, let us add that you should not rely solely on exercises in these notes. Create your own problems and questions! Modify the lemmas and theorems below, and, whenever possible, improve them! Mathematics is a highly personal experience and you will find true fulfillment only when you make the concepts in these notes your own in some way. Good luck!

Many theorems in mathematics say, one way or another, that it is very difficult to arrange mathematical object in such a way that they do not exhibit some interesting structure. The Erdős distance problem asks for the minimal number of distances determined by a set of N points in $\mathbb{R}^d$, $d \geq 2$. More precisely, let P be a finite subset of $\mathbb{R}^d$, $d \geq 2$, such that $\#P = N$. Let

$$\Delta(P) = \{|p - p'| : p, p' \in P\}, \tag{0.1}$$

and

$$|x| = \sqrt{x_1^2 + \cdots + x_d^2}, \tag{0.2}$$

the Euclidean distance.

The Erdős distance problem asks for the smallest possible size of $\Delta(P)$. Let us consider some simple examples. Let $P = \{(j, 0, \dots, 0) : j = 1, 2, \dots, N\}$. Then

Date: July 31, 2005.

The work was partly supported by a grant from the National Science Foundation.

$\Delta(P) = \{0, 1, 2, \ldots, N-1\}$. This simple example shows that the best general result we can hope for is

$$\#\Delta(P) \leq \#P. \tag{0.3}$$

This turns out to be too much. Let $P = \mathbb{Z}^d \cap [0, N^{\frac{1}{d}}]^d$, where N is a d'th power of an integer. Then $\Delta(P) = \{|p| : p \in P\}$ (why?) and $\#\Delta(P) = \#\{|p|^2 : p \in P\}$. Consider the set of numbers $p_1^2 + p_2^2 + \cdots + p_d^2$, $p = (p_1, \ldots, p_d) \in P$. All these numbers are positive integers no less than 0 and no more than $dN^{\frac{2}{d}}$. Now check that

$$\#\Delta(P) \leq dN^{\frac{2}{d}} + 1 \tag{0.4}$$

follows from this observation.

For dimension 2 the reality is even worse. It turns out (see Appendix 1) that $\#\Delta(P) \approx N^{\frac{2}{d}}$, if $d \geq 3$, and $\Delta(P) \approx \frac{N}{\sqrt{\log(N)}}$ if $d = 2$. Here, and throughout the notes, $X \lesssim Y$ means that there exists a positive constant C such that $X \leq CY$, and $X \approx Y$ means that $X \lesssim Y$ and $Y \lesssim X$. We take this notational game a step further and define $X \lessapprox Y$, with respect to the large parameter N, if for every $\epsilon > 0$ there exists $C_\epsilon > 0$ such that $X \leq C_\epsilon N^\epsilon Y$.

Erdős distance conjecture. Let P be a subset of $\mathbb{R}^d$, $d \geq 2$, such that $\#P = N$. Then

$$\#\Delta(P) \gtrapprox N, \text{ if } d = 2, \tag{0.5}$$

and

$$\#\Delta(P) \gtrsim N^{\frac{2}{d}}, \text{ if } d \geq 3. \tag{0.6}$$

□

The conjecture is nowhere near resolution, but much is known and we will come very close to the cutting edge of this beautiful problem in these notes.

Exercise 0.1. Define $\Delta_{l^1(\mathbb{R}^d)}(P) = \{|p_1 - p_1'| + \cdots + |p_d - p_d'| : p, p' \in P\}$. Prove that Erdős distance conjecture is false if $\Delta(P)$ is replaced by $\Delta_{l^1(\mathbb{R}^d)}(P)$. What should the conjecture say in this context? Can you prove this conjecture? Consider the case $d = 2$ first.

Exercise 0.2. Let K be a convex, centrally symmetric subset of $\mathbb{R}^2$, contained in the disk of radius 2 centered at the origin and containing the disk of radius 1 centered at the origin. Convex means that if x and y are in K, then the line segment connecting x and y is contained entirely inside K. Centrally symmetric means that if x is in K, then $-x$ is also in K.

Let $t = ||x||_K$ denote the number such that x is contained in tK, but is not contained in $(t - \epsilon)K$ for any $\epsilon > 0$. Define $\Delta_K(P) = \{||p - p'||_K : p, p' \in P\}$. If the boundary of K contains a line segment prove that one can construct a set P, with $\#P = N$, such that $\#\Delta_K(P) \lesssim N^{\frac{1}{d}}$.

1. Erdős' original argument

How does one prove that any set P of size N determines many distances? Let us start in two dimensions. Chose a point p_0 and draw circles around it that contains at least one point of P. Suppose that we have drawn t circles. If t is big enough then we are already doing very well. But what if t is happens to be small? Note that at least one of the t circles must contain at least N/t points. Draw the East-West line though the center of that circle. Then at least $N/2t$ are contained in either the Northern or Southern hemisphere. Without loss of generality suppose that there are $N/2t$ points in the Northern hemisphere. Fix the East-most point and draw segments from that point to all the other points of P in the Northern hemisphere. The length of these segments are all different, so at least $N/2t$ distances are thus determined. This proves that

$$\#\Delta(P) \geq \max\{t, N/2t\}. \tag{1.1}$$

There are several ways to proceed here. One way is to "guess" the answer. Since $t < \sqrt{N}$. Then $N/2t > \sqrt{N}/2$, so either way,

$$\#\Delta(P) \gtrsim \sqrt{N}. \tag{1.2}$$

A slightly less "sneaky" approach is to use the fact that

$$\max\{X, Y\} \geq \sqrt{XY} \text{ (why?)}. \tag{1.3}$$

This transforms (1.1) into (1.2). Summarizing, we have just proved the following.

Theorem 1.1 (Erdős [8]). *Suppose that $d = 2$ and $\#P = N$. Then (1.2) holds.*

What about higher dimensions? Let us try the same approach. Choose a point in P and draw all spheres that contain at least one point of P. As before, let t denote the number of spheres. If t is large enough, we are done. If not, then one of the spheres contains at least N/t points. Unfortunately, if $d > 2$, we cannot run the simple minded argument that worked in two dimensions. Or can we? Notice that if we are working in $\mathbb{R}^d$, the surface of each sphere is $(d-1)$-dimensional, whatever that means. This suggests the following approach.

Induction Hypothesis. Let P' be a subset of $\mathbb{R}^k$, $k \geq 2$, or S^k, $k \geq 1$. Suppose that $\#P' = N'$. Then

$$\#\Delta(P') \gtrsim (N')^{\frac{1}{k}}.$$

□

In the case of $\mathbb{R}^k$, the induction hypothesis holds if $k = 2$ as we have verified above. Similarly, we have verified the statement for S^k for $k = 1$. We are now ready to complete the higher dimensional argument. Then for the dimension d argument we end up with t $(d-1)$-spheres–one of which must have at least N/t points on it as in the $d = 2$ proof. By induction, these points determine $\gtrsim \left(\frac{N}{t}\right)^{\frac{1}{d-1}}$ distances. It follows that

$$\#\Delta(P) \gtrsim \max\left\{t, \left(\frac{N}{t}\right)^{\frac{1}{d-1}}\right\}. \tag{1.4}$$

We now use the fact that

$$\max\{X, Y\} \geq (XY^{d-1})^{\frac{1}{d}} \text{ (why?)}, \tag{1.5}$$

which implies that

$$\#\Delta(P) \gtrsim N^{\frac{1}{d}}. \tag{1.6}$$

We just proved the following result.

Theorem 1.2. *Let P be a subset of $\mathbb{R}^d$, $d \geq 2$, such that $\#P = N$. Then* (1.6) *holds.*

Exercise 1.1. Prove that the minimum of $\max\{t, N/2t\}$ is in fact $\sqrt{N}$. In other words, show that Erdős's method of proof cannot do better than $\#\Delta(P) \gtrsim \sqrt{N}$.

Exercise 1.2. Let K be a polygon in the plane. Let $\#P = N$. Prove that $\#\Delta_K(P) \gtrsim \sqrt{N}$. What about other convex K? This problem turns out to be surprisingly difficult. See a very nice article by Julia Garibaldi.

Exercise 1.3. We outline an alternate proof of Theorem 1.1. Let M_N denote the matrix constructed as follows. Fix $t \in \Delta(P)$ and let the entry $a_{pp'} = 1$ if $|p - p'| = t$, and 0 otherwise. Observe that for a fixed pair (p', p''), $p' \neq p''$, $a_{pp'} \cdot a_{pp''} = 1$ for at most one value of p (why?). Use this along with the Cauchy-Schwartz inequality to prove that $\sum_{p,p' \in P} a_{pp'} \lesssim N^{\frac{3}{2}}$. Conclude that for any $t \in \Delta(P)$, $\#\{(p, p') : |p - p'| = t\} \lesssim N^{\frac{3}{2}}$. Deduce that $\#\Delta(P) \gtrsim \sqrt{N}$. Can you make this idea run in higher dimensions?

Exercise 1.4. In the proofs of Theorems 1.1 and 1.2 we only used spheres centered at a single point. Is there any milage to be gained from considering, say, two points? Try it.

2. Moser's approach and the Erdős integer distance principle

Erdős' ingenious argument, described in the previous chapter, relies on spheres centered at a single point, and it stands to reason that one might gain something out of considering spheres centered at two points. This point of view was introduced by Moser in the early 1950s. Before presenting Moser's argument, we will present the Erdős integer distance principle where an idea similar to Moser's is already present, albeit in a different form and context.

Erdős integer distance principle (EIDP), [9]. Let A be an infinite subset of $\mathbb{R}^d$, $d \geq 2$. Suppose that $\Delta(A) \subset \mathbb{Z}$. Then A is contained in a line. □

To prove EIDP suppose that A is not contained in a line. Suppose that $d = 2$. Let a, a', a'' denote three points of A not lying on the same line. Let b be any other point of A. By assumption, $|a - b|$ and $|a' - b|$ are both integers, which means that $|a - b| - |a' - b|$ is also an integer. This means that every point of A is contained on hyperbolas with focal points at a and a'. (See Appendix 2 for a thorough description of basic theory of hyperbolas in the plane). How many such hyperbolas are there? Well, suppose that $|a - a'| = k$, which, by assumption is an integer. By the triangle inequality, $||a - b| - |a' - b|| \leq |a - a'| = k$. It follows that there are only $k + 1$ different hyperbolas with focal points at a and a'. Similarly, all the points of A are contained in $l + 1$ hyperbolas with focal points at a' and a''. Any hyperbola with focal points at a and a' and a hyperbola with focal points at a' and a'' intersect at at most 4 points (see Appendix 2 once again). It follows that the number of points in A cannot exceed $16(k + 1)(l + 1)$, which is a contradiction since A is assumed to be infinite. This proves the two-dimensional case of the Erdős integer distance principle. The higher dimensional argument is outlined in Exercise 2.1 below.

The following beautiful extension of the Erdős integer distance principle was proved by Jozsef Solymosi [23].

Theorem 2.1. *Suppose that P is a subset of $\mathbb{R}^2$, such that $\Delta(P) \subset \mathbb{Z}$ and $\#P = N$. Suppose that P is contained in a disk of radius R. Then $R \gtrsim N$.*

The proof of Solymosi's theorem is outlined in Exercise 2.5, and in Exercise 2.6 we ask you to verify that Theorem 2.1 would follow immediately from the Erdős distance conjecture.

We are now ready to introduce Moser's idea. Choose points X and Y in P such that

$$|X-Y| \leq \min\{|p-p'| : p, p' \in P\}. \tag{2.1}$$

Let O be the midpoint of the segment XY. Half the points of P are either above or below the line connecting X and Y. Call this set of points P'. Assume without loss of generality that at least half the points are above the line. Draw half annuli centered at O of thickness $|X-Y|$ until all the points of P' are covered. Keep only one third of the annuli in such a way that at least one third of the points of P' are there and such that if a particular annulus is kept, the next two consecutive annuli are discarded. (Prove that this can be done and explain why we are doing this as you read the rest of the argument!). Call the resulting set of points P''. Let n_j denote the number of points of P'' in the jth annulus. Let $\mathcal{A}_j$ denote the intersection of P'' with the jth annulus. Suppose that

$$\{|p-X| : p \in \mathcal{A}_j\} \cup \{|p-Y| : p \in \mathcal{A}_j\} = \{d_1, d_2, \dots, d_k\}. \tag{2.2}$$

Let

$$A_j = \{p \in \mathcal{A}_j : |p-X| = d_j\}, \tag{2.3}$$

and

$$B_i = \{p \in \mathcal{A}_i : |p-Y| = d_i\}. \tag{2.4}$$

By construction,

$$A_j = \cup_i \left(A_j \cap B_i\right), \tag{2.5}$$

since points of distance d_j from X are of some distance or another from Y. It follows that

$$\cup_j A_j = \cup_{i,j} \left(A_j \cap B_i\right). \tag{2.6}$$

Now,

$$\# \cup_j A_j = n_j, \tag{2.7}$$

while

$$\# \cup_{i,j} \left(A_j \cap B_i\right) \leq k^2 \max_{i,j} \# \left(A_j \cap B_i\right). \tag{2.8}$$

Now, A_j and B_i are contained on circles of approximately the same radius centered at different points, so $\max_{i,j} \#(A_j \cap B_i) \leq 1$. Plugging this into (2.8) we see that

$$k \geq \sqrt{n_j}, \tag{2.9}$$

from which we deduce that

$$\#\Delta(P) \geq \#\Delta(P'') \geq \sum_j \sqrt{n_j}. \tag{2.10}$$

We have

$$\frac{N}{6} \leq \sum_j n_j = \sum_j \sqrt{n_j} \cdot \sqrt{n_j} \leq \sqrt{n_{max}} \cdot \sum_j \sqrt{n_j}, \tag{2.11}$$

where

$$n_{max} = \max_j n_j. \tag{2.12}$$

Observe that by the proof of Theorem 1.1,

$$\#\Delta(P) \geq \#\Delta(P'') \geq n_{max}. \tag{2.13}$$

By (2.10),

$$\#\Delta(P) \geq \frac{N}{6\sqrt{n_{max}}}. \tag{2.14}$$

It follows that

$$(\#\Delta(P))^2 \cdot \#\Delta(P) \geq n_{max} \cdot \frac{N^2}{36 n_{max}} = \frac{N^2}{36}. \tag{2.15}$$

Which implies that

$$\#\Delta(P) \geq \frac{N^{\frac{2}{3}}}{(36)^{\frac{1}{3}}}, \tag{2.16}$$

and we have just proved the following theorem.

Theorem 2.2 (Moser [19]). *Let $d = 2$ and suppose that $\#P = N$. Then $\#\Delta(P) \gtrsim N^{\frac{2}{3}}$.*

Exercise 2.1. Why did we eliminate 2/3 of the annuli in the proof above? Where did we use this in the proof?

Exercise 2.2. What does Moser's method yield in higher dimensions? Can you use the two-dimensional result along with the induction argument used to prove Theorem 1.2 instead? Which approach yields better exponents?

Exercise 2.3. Let A be an infinite subset of $\mathbb{R}^d$, $d \geq 2$, with the following property. We assume that $|a - a'| \geq \frac{1}{100}$ for all $a \neq a' \in A$. We also assume that for every $m \in \mathbb{Z}^d$, $[0,1]^d + m$ contains exactly one point of A. Let $A_q = [0,q]^d \cap A$. What kind of a bound can you obtain for $\Delta(A_q)$ using Moser's idea? Why is this bound better than the one we obtain above?

Take this a step further. Instead of using two points as in Moser's argument, use d points. How should these points be arranged? What effect are we trying to achieve? Can you obtain a better exponent this way?

Exercise 2.4. Outline of proof of EIDP in higher dimensions.

Exercise 2.5. Deduce Solymosi' s Theorem from the following observation by using ideas from the proof of the EIDP.

Observation 1. For every set of n points in the plane with diameter Δ and with at most $n/2$ collinear points, there exists two pairs of points A,B and C,D such that each of the distances $\overline{AB}$ and $\overline{CD}$ are less than $6\Delta/n^{1/2}$.

Now prove the Observation 1. *Hint:* Show that there are fewer than $n/2$ points that are not within $6\Delta/n^{1/2}$ of other points.

Exercise 2.6. Deduce Solymosi's theorem from the Erdős distance conjecture.

3. Incidence theorems and graph theory

If you are familiar with basic theory of graphs, keep reading. If not, read Appendix 4 first where basic notions of graph theory are introduced and proved. We will also make use of some basic concepts from probability theory. Those are described in Appendix 5 below.

Let P be a finite set of n points in $\mathbb{R}^2$, and let L be a finite set of m lines. Define an incidence of P and L to be a pair $(p,l) \in P \times L : p \in l$. Let $I_{P,L}$ denote the total number of incidences between P and L. More precisely,

$$I_{P,L} = \#\{(p,l) \in P \times L : p \in l\}. \tag{3.1}$$

We already proved something about $I_{P,L}$ in Exercise 1.3, did we not? Let us think about it for a moment. Let $\delta_{lp} = 1$ if $p \in l$, and 0 otherwise. Then, by the Cauchy-Schwartz inequality,

$$\begin{aligned} I_{P,L} &= \sum_l \sum_p \delta_{lp} \le \sqrt{m} \left(\sum_l \left| \sum_p \delta_{lp} \right|^2 \right)^{\frac{1}{2}} \\ &= \sqrt{m} \left(\sum_l \sum_p \delta_{lp}^2 + \sum_l \sum_{p \neq p'} \delta_{lp}\delta_{lp'} \right)^{\frac{1}{2}} \\ &\le \sqrt{m} \left(mn + \sum_l \sum_{p \neq p'} \delta_{lp}\delta_{lp'} \right)^{\frac{1}{2}}. \end{aligned} \tag{3.2}$$

Now, for each $(p,p') \in P \times P$, $p \neq p'$, there is at most one l such that $\delta_{lp}\delta_{lp'} \neq 0$. This is because $\delta_{lp} = 1$ means that $p \in l$, and $\delta_{lp'} = 1$ means that $p' \in l$. Since two points uniquely determine a line, the expression $\delta_{lp}\delta_{lp'}$ cannot equal to one for any other l. It follows that

$$\sum_l \sum_{p \neq p'} \delta_{lp}\delta_{lp'} \le \#\{(p,p') \in P \times P : p \neq p'\} = n(n-1). \tag{3.3}$$

Now it can be shown that the following theorem holds. Check the details.

Theorem 3.1. *Let P be a set of n points in the plane, and let L be a set of m lines. Then $I_{P,L} \lesssim m\sqrt{n} + n\sqrt{m}$.*

As pretty as this result is, it turns out that we can do better. The following improvement on Theorem 3.1 is due to Szemeredi and Trotter [26].

Theorem 3.2. *Let P be a set of n points in the plane, and let L be a set of m lines. Then $I_{P,L} \lesssim n + m + (nm)^{\frac{2}{3}}$.*

We shall deduce this theorem from the following graph theoretic result.

Theorem 3.3. *Let G be a graph with n vertices and e edges. Suppose that $e \geq 4n$. Then*
$$cr(G) \gtrsim \frac{e^3}{n^2}$$
.

We now prove Theorem 3.2 using Theorem 3.3. In order to use Theorem 3.3 we construct the following graph. Let the points of P be the vertices of G and let the line segments connecting points of P on the lines L be the edges. You will prove in Exercise 3.2 below (not very difficult) that

$$e = I - m. \tag{3.4}$$

There are two possibilities. If $e < 4n$, then

$$I < m + 4n, \tag{3.5}$$

which is certainly alright with us.

If $e \geq 4n$, then Theorem 3.3 kicks in and we have

$$cr(G) \gtrsim \frac{e^3}{n^2} = \frac{(I-m)^3}{n^2}. \tag{3.6}$$

On the other hand, a crossing arises when two edges intersect not at a vertex. Since edges come from lines and there are m lines,

$$cr(G) \leq m^2. \tag{3.7}$$

Combining (3.5) and (3.6), we obtain the conclusion of Theorem 3.2.

It remains for us to prove Theorem 3.3. By Appendix 4,

$$cr(G) \geq e - 3n. \tag{3.8}$$

Choose a random subgraph H of G by keeping each vertex with probability p, a number to be chosen later. It follows that

$$\mathbb{E}(vertices\ in\ H) = np,$$
$$\mathbb{E}(edges\ in\ H) = ep^2,$$

and

$$\mathbb{E}(cr(H)) \leq cr(G)p^4, \tag{3.9}$$

where $\mathbb{E}$ denotes the expected value.

By (3.9) and linearity of expectation,

$$cr(G)p^4 \geq ep^2 - 3np. \tag{3.10}$$

Choosing $p = \frac{4n}{e}$, as we may, since $e \geq 4n$, we obtain the conclusion of Theorem 3.3.

One of the most misused words in mathematics is “sharp”. Nevertheless, we are about to use it ourselves. We will show that Theorem 3.2 is sharp in the sense that for any positive integer n and m, we can construct a set P of n points, and a set L of m lines, such that

$$I_{P,L} \approx n + m + (nm)^{\frac{2}{3}}. \tag{3.11}$$

We shall construct an example in the case $n = m$, but we absolutely insist that you work out the general case in one of the exercises below. Let

$$P = \{(i,j) : 0 \leq i \leq k-1; 0 \leq j \leq 4k^2 - 1\}. \tag{3.12}$$

Let L be the set consisting of lines given by equations $y = ax+b$, $0 \le a \le 2k-1$, $0 \le b \le 2k^2 - 1$. Thus we have n lines and n points. Moreover, for $x \in [0, k)$,

$$ax + b < ak + b < 4k^2, \tag{3.13}$$

and it follows that for each $i = 0, 1, \dots, k$, each line of L contains a point of P with x-coordinate equal to i. It follows that

$$I_{P,L} \ge k \cdot \#L = \frac{1}{4} n^{\frac{4}{3}}. \tag{3.14}$$

Exercise 3.1. Complete the details of the proof of Theorem 3.1.

Exercise 3.2. Prove (3.4) and write out the details.

Exercise 3.3. For each n and m, construct a set P of n points and a set L of m lines, such that (3.11) holds. Use the argument in the case $n = m$ above as the basis of your construction.

Exercise 3.4. Let P be a set of n points in the plane. Let L be a set of m curves. Let $\alpha_{pp'}$ denote the number of curves in L that pass through p and p'. Let $\beta_{ll'}$ denote the number of points of P that are contained in both l and l', Use the proof of Theorem 3.1 to show that

$$I_{P,L} \le n\sqrt{m} \left(\sum_{p \ne p'} \alpha_{pp'} \right)^{\frac{1}{2}} + m\sqrt{n} \left(\sum_{l \ne l'} \beta_{ll'} \right)^{\frac{1}{2}}. \tag{3.15}$$

Exercise 3.5. Prove a modified version of Theorem 3.3 which says that if α is the maximum number of edges connecting a pair of vertices in G, then

$$cr(G) \gtrsim \frac{e^3}{\alpha n^2}. \tag{3.16}$$

Hint: This can be proven by repeatedly using probabilistic arguments similar to those used in the proof of Theorem 3.3. First, delete edges independently with probability $1 - \frac{1}{k}$ and then delete all the remaining multiple edges–call this resulting graph G'. Calculate the probability p_e that a fixed edge e remains in G'. Now compare the expected number of edges and crossings in G' to the number in the original graph and use Theorem 3.3. Finally, use Jensen's inequality which says that $\mathbb{E}[x^a] \ge (\mathbb{E}[x])^a$ for $a \ge 1$.

Exercise 3.6. Let P be a set of n points in the plane. Let L be a set of m curves. Suppose that no more than α curves in L pass through a pair of points of P, and no more than β points of P are contained in the intersection of any two curves in L. What should Theorem 3.2 say under these hypotheses? Do it now because we will use this result in the next chapter. *Hint:* Use the result from Exercise 3.5.

Exercise 3.7. Is the weighted theorem given by Exercise 3.6 always stronger than the one given by Exercise 3.4? Give explicit examples to support your belief.

4. Bisectors enter the game: $n^{\frac{4}{5}}$ plateau is reached

In this section we shall use graph theory that already bore fruit in the previous chapter to improve the Erdős exponent from $2/3$ to $4/5$.

Suppose that a set P of n points determined t distinct distances. Draw a circle centered at each point of P containing at least one other point of P. By assumption, we have at most t circles around each point and thus the total number of circles is nt. By construction, these circles have $n(n-1)$ incidences with the points of P. The idea now is to estimate the number of incidences from above in terms of n and t and then derive the lower bound for t.

Delete all circles with at most two points on them. This eliminates at most $2nt$ incidences, and since we may safely assume that t is much smaller than n, the number of incidences of the remaining circles and the points of P is still $\gtrsim n^2$. Form a graph whose vertices are points of P and edges are circular arcs between the points. This graph G has $\approx n$ vertices, $\approx n^2$ edges, and the number of crossings is $\lesssim (nt)^2$.

Suppose for a moment that we can use Theorem 3.3. Then

$$\frac{e^3}{n^2} \lesssim cr(G) \lesssim (nt)^2, \tag{4.1}$$

and since $e \approx n^2$, it would follow that

$$n^4 \lesssim n^2 t^2, \tag{4.2}$$

which would imply the Erdős Distance Conjecture. Unfortunately, life is harder than that since Theorem 3.3 only applies if there is at most one edge connecting a pair of vertices. In our case we may assume that there is at most $2t$ edges connecting a pair of vertices (why? see Exercise 4.1 below). Applying Exercise 3.5 we see that

$$\frac{e^3}{tn^2} \lesssim cr(G) \lesssim n^2 t^2, \tag{4.3}$$

which implies that

$$t \gtrsim n^{\frac{2}{3}}, \tag{4.4}$$

the Moser's bound from Chapter 2. All this for $n^{\frac{2}{3}}$?! We must be able to do better than that! How can we possibly hope to do that? One way is to study edges of high multiplicity separately.

We try to take advantage of the following phenomenon. Let $p, p' \in P$. The centers of all the circles that pass through p and p' are located on the bisector, $l_{pp'}$, of the points p and p' in P [1]. Let us consider all the bisectors with at least k points on them. How many such bisectors are there? Recall that the Szemeredi-Trotter incidence bound (Theorem 3.2) says that the number of incidences between n points and m lines is $\lesssim (n+m+(nm)^{\frac{2}{3}})$. Let m_k denote the number of lines with at least k points. Then the number of incidences is at least km_k. It follows that

$$km_k \lesssim n + m_k + (nm_k)^{\frac{2}{3}}, \tag{4.5}$$

and we conclude that

$$m_k \lesssim \frac{n}{k} + \frac{n^2}{k^3}. \tag{4.6}$$

This implies that bisectors with at least k points on them have

$$\lesssim n + \frac{n^2}{k^2} \tag{4.7}$$

incidences with the points of P (see Exercise 4.3).

[1] The bisector of p and p' is the set of points that are equidistant to p and p'. Formally, $l_{pp'} = \{z \in \mathbb{R}^2 : |z-p| = |z-p'|\}$. In the Euclidean metric this turns out to be the line perpendicular to $\overline{pp'}$ through their midpoint. For more general metrics see Exercise 4.2.

Let P_k denote the set of pairs (p, p') of P connected by at least k edges. Let E_k denote the set of edges connecting those pairs. Each edge in E_k connecting a pair (p, p') corresponds to exactly one incidence of $l_{pp'}$ with a point p'' in P. However, an incidence of such a p'' with some $l_{pp'}$ corresponds to at most $2t$ edges in E_k since there at at most t circles centered at p''. It follows that

$$\#E_k \lesssim tn + \frac{tn^2}{k^2}. \tag{4.8}$$

Note that we are almost certainly over counting E_k here since we are removing all possible edges corresponding to incidences–not just those that contribute to high multiplicity.

Now, if we choose $k = c\sqrt{t}$, for an appropriate constant c, then

$$\#E_k \leq \frac{n^2}{2}. \tag{4.9}$$

If we now erase all the edges of E_k, there are still more than $\frac{n^2}{2}$ edges remaining. Applying Exercise 3.5 once again, we see that

$$\frac{e^3}{kn^2} \leq cr(G) \leq n^2t^2. \tag{4.10}$$

Since $k \approx \sqrt{t}$ and $e \approx n^2$, it follows that

$$t \gtrsim n^{\frac{4}{5}}. \tag{4.11}$$

We have just proved the following theorem of Szekely [25].

Theorem 4.1. *Let P be a set of n points in the plane. Then*

$$\#\Delta(P) \gtrsim n^{\frac{4}{5}}. \tag{4.12}$$

Exercise 4.1. Explain why there can be at most $2t$ edges connecting two vertices in the graph G from the above proof.

Exercise 4.2. Consider the l_1 metric defined in Exercise 0.1. Try to figure out what bisectors look like for this metric.

Exercise 4.3. Verify Equation 4.7. *Hint:* Define M_j to be the set of lines with between 2^j and 2^{j+1} points and observe that m_k can be written as a sum of such sets.

5. Arithmetic joins bisectors in the Erdős crusade!

In this chapter we present the Solymosi-Toth beautiful argument that will get us up to $n^{\frac{6}{7}}$ which opens the door to further important developments that we sketch in the next chapter. We start out with the following beautiful observation due to Jozsef Beck [3]. The proof we give is from [24].

Lemma 5.1. *Let P be a collection of n points in the plane. Then one of the following holds:*

1. *There exists a line containing $\approx n$ points of P.*
2. *There exist $\approx n^2$ different lines each containing at least two points of P.*

Proof. Let $L_{u,v}$ be the number of pairs of points of P which determine a line that goes through at least u but at most v points of P. Equation 4.6 and basic counting arguments tells us that $L_{u,v} \lesssim \frac{n^2v^2}{u^3} + \frac{nv^2}{u}$ (see exercise 5.3). Fix a constant C and consider $L_{C,N/C}$. Then

$$\begin{aligned}
L_{C,N/C} &\leq \sum_{i=0}^{\lfloor \log(N) \rfloor} N_{C2^i,C2^{i-1}} \\
&= \sum_{i=0}^{\lfloor \log(N) \rfloor} O\left(\frac{4N^2}{C2^i} + 4CN2^i\right) \\
&= O\left(\frac{N^2}{C}\sum_{i=0}^{\lfloor \log(N) \rfloor} 2^{-i} + NC\sum_{i=0}^{\lfloor \log(N) \rfloor} 2^i\right) \\
&= O\left(\frac{N^2}{C}\right).
\end{aligned} \tag{5.1}$$

In other words, for some $C_o > 0$ we have $L_{C,N/C} \leq C_o\left(N^2/C\right)$. Thus for the appropriate choice of C at least half of the pairs of points determine a line through fewer than C or at least C/N points. And consequently at least a fourth of the pairs go through fewer than C points or a fourth go through at least C/N points. In either case we are done.

□

Consider a set P of n points and let $\mathcal{L}$ denote the set of lines passing through at least two points of P. An averaging argument (see exercise 5.1) applied to Lemma 5.1 implies that there exists an absolute constant c_o such that at least c_on points of P are incident to at least c_on lines of $\mathcal{L}$. Then let B be the set of such points, and take some arbitrary point $a \in B$.

Draw in the lines through a that go through points of P. There must be at least c_on such lines. Choose one point other than a on each of these lines and draw in the circles around a that hit those chosen points (deleting those capturing fewer than 3 points). On each of these circles break the points in triples, possibly deleting as many as 2 from each. We still have $\gtrsim n$ points left by our hypotheses (check!).

We call a triple "bad" if all three bisectors formed from its points go through at least k points. And we call the initial point a from B "bad" if at least half of its triples are bad. We would like to choose k such that at least half the points of B are bad. Clearly, the smaller k is the "easier" it is to get k-rich lines and thus more bad points. However, it will become clear that we would like k as large as possible. You will show in Exercise 5.2 that we may take $k = \frac{c_2n^2}{t^2}$.

Then if we can get the following upper and lower bounds on the number of incidences $I(L_k, P)$ of k-rich lines and bad points we will be done:

$$n^2/t^{2/3} \lesssim I(L_k, P) \lesssim t^4/n^2. \tag{5.2}$$

Finding an upper bound on $I(L_k, P)$ is straight forward. We simply apply Equation 4.6 to find a bound on the number of k-rich lines and then use Theorem 3.1 to get that $I(L_k, P) \lesssim n^2/k^2$. Getting a lower bound on the quantity $I(L_k, p)$ in terms of n and t is somewhat harder. The following lemma is the key to the whole proof.

Lemma 5.2. *Let T be a set of N triples (a_i, b_i, c_i) of distinct real numbers such that $a_i < b_i < c_i$ for $i = 1, \dots, N$ and $c_i < a_{i+1}$ for all but at most $t-1$ of the i. Let $W = \{\frac{a_i+b_i}{2}, \frac{a_i+c_i}{2}, \frac{b_i+c_i}{2} : i = 1, \dots, N\}$. Then $|W| \gtrsim \frac{N}{t^{2/3}}$.*

Proof. Let the range of a triple $(a, b, c) \in T$ be defined as the interval $[a, c]$. By assumption, the sequence $(a_1, b_1, c_1, a_2, b_2, c_2, \dots, a_N, b_N, c_N)$ be partitioned into at most t contiguous monotone increasing subsequences. Partition the real axis into $N/(2t)$ open intervals so that each interval fully contains the ranges of t triples. These intervals are constructed from left to right. Let x denote the right endpoint of the rightmost interval constructed so far. Discard the at most t triples whose ranges contain x, and move to the right until you reach a point y that lies to the right of exactly t new ranges. We add (x, y) as a new open interval and continue in this manner until all triples are processed.

Let s be one of the open intervals defined in the previous paragraph. Each triple in T whose range is fully contained in s contributes three elements to $W \cap s$, and no two triples of T contribute the same triple to $W \cap s$. It follows that $|W \cap S| \geq t^{\frac{1}{3}}$, since otherwise the number of distinct triples of its elements would be smaller than t. Since the number of intervals s is $N/(2t)$, the conclusion of the lemma follows by the multiplication principle. This completes the proof of the lemma. □

For each point $p \neq a$ in a bad triple, map p to the orientation of the ray $\overrightarrow{ap}$. By construction this map is an injection, and W corresponds to k-rich lines. Therefore the number of k-rich lines incident to a is $\gtrsim n/t^{2/3}$. And since a was an arbitrary element of B, we get that $I(L_k, P) \gtrsim n^2/t^{2/3}$.

The only thing that remains is to show that if we take $k = \frac{c_2 n^2}{t^2}$ then half of the points of P are "bad". Construct a multigraph G out of the points that are part of the triples as in the proof of Theorem 4.1. Next apply the result of Exercise 3.5. Doing this we find that we can take $k = \frac{c_2 n^2}{t^2}$ and at least $c_o n/2$ points of B will be bad. We leave the details as an exercise to the reader. See [22] for the details.

Exercise 5.1. Write up the details of the averaging argument which tells us that "many" points go through "many" lines of $\mathcal{L}$. *Hint:* recall that we may assume that $t = o(n)$.

Exercise 5.2. Work out the details showing that we may take $k = \frac{c_2 n^2}{t^2}$ and at least $c_o n/2$ points of B will still be bad.

Exercise 5.3. Check that Equation 4.6 and basic counting arguments gives us that $L_{u,v} \lesssim \frac{n^2 v^2}{u^3} + \frac{n v^2}{u}$.

Exercise 5.4. Find the constants C and C_o in the proof of Theorem 5.1 and write up the details of why we are done in the case where at least a fourth of the pairs go through at least N/C points of P.

Appendix A. Sums of Squares

Appendix B. Hyperbolas in the Plane

The standard equation for a hyperbola in the plane that is centered at the origin and whose foci are $(-c, 0)$ and $(c, 0)$ is $x^2/a^2 - y^2/b^2 = 1$, where $a^2 + b^2 = c^c$. Fixing two points F_1 and F_2 in the plane a hyperbola can also be described as the set of points P such that $||PF_1| - |PF_2|| = 2a$ for some fixed number a. We will tend to use the latter definition. Check and see how these two definitions are related!

Appendix C. The Cauchy-Schwartz Inequality

Let $\{a_j\}_{j=1}^n$ and $\{b_j\}_{j=1}^n$ be sequences of real numbers. Our goal is to prove that

$$\sum_{j=1}^n a_j b_j \le \left(\sum_{j=1}^n a_j^2\right)^{\frac{1}{2}} \cdot \left(\sum_{j=1}^n b_j^2\right)^{\frac{1}{2}}. \tag{C.1}$$

Let

$$A = \left(\sum_{j=1}^n a_j^2\right)^{\frac{1}{2}} \text{ and } B = \left(\sum_{j=1}^n b_j^2\right)^{\frac{1}{2}}, \tag{C.2}$$

so it suffices to prove that

$$\sum_{j=1}^n \frac{a_j}{A}\frac{b_j}{B} \le 1. \tag{C.3}$$

Since

$$\left(\frac{a_j}{A} - \frac{b_j}{B}\right)^2 \ge 0, \tag{C.4}$$

we conclude that

$$\frac{a_j}{A} \cdot \frac{b_j}{B} \le \frac{1}{2}\frac{a_j^2}{A^2} + \frac{1}{2}\frac{b_j^2}{B^2}. \tag{C.5}$$

It follows that

$$\sum_{j=1}^n \frac{a_j}{A}\frac{b_j}{B} \le \frac{1}{2}\sum_{j=1}^n \frac{a_j^2}{A^2} + \frac{1}{2}\sum_{j=1}^n \frac{b_j^2}{B^2} = \frac{1}{2} + \frac{1}{2} = 1. \tag{C.6}$$

Thus we have proved the Cauchy Schwartz inequality:

Theorem C.1. *Let a_j, b_j be as above. Then* (6.3.1) *holds.*

Appendix D. Basic graph theory

To appear soon!

Appendix E. Basic probability theory

Also to appear soon. We promise.

References

[1] P. K. Agarwal, E. Nevo, J. Pach, R. Pinchasi, M. Sharir, and S. Smorodinsky, *Lenses in arrangements of pseudo-circles and their applications*, J. ACM, to appear.

[2] M. Ajtai, V. Chvatal, M. Newborn, and E. Szemeredi, *Crossing-free subgraphs*, Ann. Discrete Mathematics **12** (1982) 9–12.

Cutting circles into pseudo-segments and improved bounds for incidences, Discrete Comput. Geom. **28** (2002), no. 4, 475–490.

Falconer conjecture, spherical averages and discrete analogs, Towards a theory of geometric graphs, 15–24, Contemp. Math., 342, Amer. Math. Soc., Providence, RI, 2004.

[3] J. Beck, *On the lattice property of the plane and some problems of Dirac, Motzkin and Erdős in combinatorial geometry*, Combinatorica **3** (1983), no. 3-4, 281–297.
Unit distances, J. Combinatorical Theory A **37** (1984), no. 3, 231–238.
Hausdorff dimension and distance sets, Israel J. Math. **87** (1994), no. 1-3, 193–201.
On the Erdős-Volkmann and Katz-Tao ring conjectures, Geom. Funct. Anal. **13** (2003), no. 2, 334–365.

[4] P. Brass, *Erdős distance problems in normed spaces*, Discrete Comput. Geom. **17** (1997), no. 1, 111–117.

[5] F. R. K. Chung, *The number of different distances determined by n points in the plane*, J. Combin. Theory Ser. A **36** (1984), no. 3, 342–354.

[6] F. R. K. Chung, E. Szemerédi and W. T. Trotter, *The number of different distances determined by a set of points in the Euclidean plane*, Discrete Comput. Geom. **7** (1992), no. 1, 1–11.

[7] K. L. Clarkson, H. Edelsbrunner, L. Guibas, M. Sharir, and E. Welzl, *Combinatorial complexity bounds for arrangements of curves and spheres*, Discrete Comput. Geom. **5** (1990), no. 2, 99–160.

[8] P. Erdős, *On sets of distances of n points*, Amer. Math. Monthly **53** (1946) 248–250.

[9] P. Erdős, *Integral distances*, Bull. Amer. Math. Soc. **51** (1945).

[10] J. S. Garibaldi, *A Lower Bound for the Erdős Distance Problem for Convex Metrics*, preprint.

[11] A. Iosevich, *Curvature, combinatorics and the Fourier transform*, Notices Amer. Math. Soc. **48** (2001), no. 6, 577–583.
Fourier analysis and geometric combinatorics, (to appear in the Birkhauser volume dedicated to the annual Padova lectures in analysis).

[12] A. Iosevich and I. Łaba, *Distance sets of well-distributed planar sets*, Discrete Comput. Geom. **31** (2004), no. 2, 243–250.

[13] A. Iosevich and I. Łaba, *K-distance, Falconer conjecture, and discrete analogs*, preprint.

[14] N. H. Katz and G. Tardos, *A new entropy inequality for the Erdős distance problem*, Towards a Theory of Geometric Graphs, (ed. J Pach) Contempory Mathematics **342** (2004), 119–12.

[15] A. G. Khovanskiĭ, *A class of systems of transcendental equations*, Dokl. Akad. Nauk SSSR **255** (1980), no. 4, 804–807.

[16] A. G. Khovanskiĭ, *Fewnomials*, Translated from the Russian by Smilka Zdravkovska, Amer. Math. Soc., Providence, RI, 1991.

[17] T. Leighton, *Complexity Issues in VLSI, Foundations of Computer Series*, MIT Press, Cambridge, MA, 1983.

[18] L. Ma, *Bisectors and Voronoi diagrams for convex distance functions*, dissertation (unpublished).

[19] L. Moser, *On the different distances determined by n points*, Amer. Math. Monthly **59** (1952), 85–91.

[20] J. Pach and P.K. Agarwal, Combinatorial Geometry, Wiley, New York, NY, 1995.

[21] J. Pach and M. Sharir, *On the number of incidences of points and curves*, Combin. Probab. Comput. **7** (1998), no. 1, 121–127.

[22] J. Solymosi and C. Tóth, *Distinct distances in the plane*, Discrete Comput. Geom. **25** (2001), no. 4, 629–634.

[23] J. Solymosi, *Note on integral distances*, Discrete Comput. Geom. **30** (2003), no. 2, 337–342.

[24] J. Solymosi, G. Tardos and C. D. Tóth, *The k most frequent distances in the plane*, Discrete Comput. Geom. **28** (2002), no. 4, 639–648.

[25] L.A. Székely, *Crossing numbers and hard Erdős problems in discrete geometry*, Combin. Probab. Comput. **6** (1997), no. 3, 353–358.

[26] E. Szemerédi and W. T. Trotter, Jr., *Extremal problems in discrete geometry*, Combinatorica **3** (1983), no. 3-4,381–392.

[27] G. Tardos, *On distinct sums and distinct distances*, Adv. Math. **180** (2003), no. 1, 275–289.

Department of Mathematics, University of Missouri-Columbia, Columbia Missouri 65211 USA

E-mail address: iosevich math.missouri.edu