Chapter 1: A Taste of Classical Discrete Geometry

Adam Sheffer

April 16, 2015

"My most striking contribution to geometry is, no doubt, my problem on the number of distinct distances. This can be found in many of my papers on combinatorial and geometric problems." / Paul Erdős, in a survey of his favorite contributions to mathematics, compiled for the celebrations of his 80'th birthday [4].

We begin the course with a quick tour of some classical discrete geometry, from before the introduction of the polynomial method. This chapter is rather different than the rest of the course (e.g., it involves graphs), but it also introduces several tricks that we will use in the following chapters.

1 Introduction

We use standard asymptotic notation. That is, $f(n) = O(g(n))$ implies that there exist constants c, n_0, such that for any $n \geq n_0$, we have $f(n) \leq c \cdot g(n)$. For example, $10n^2 + 1000 = O(n^2)$ holds since we can take $c = 100$ and $n_0 = 20$. Similarly, $f(n) = \Omega(g(n))$ implies that there exist constants c, n_0, such that for any $n \geq n_0$, we have $f(n) \geq c \cdot g(n)$. Finally, $f(n) = \Theta(g(n))$ implies that both $f(n) = O(g(n))$ and $f(n) = \Omega(g(n))$ hold.

Given a set $\mathcal{P}$ of points and a set $\mathcal{L}$ of lines, both in $\mathbb{R}^2$, an *incidence* is a pair $(p, \ell) \in \mathcal{P} \times \mathcal{L}$ such that the point p is contained in the line ℓ. We denote by $I(\mathcal{P}, \mathcal{L})$ the number of incidences in $\mathcal{P} \times \mathcal{L}$. For any m, n, Erdős constructed a set $\mathcal{P}$ of m points and a set $\mathcal{L}$ of n lines with $\Theta(m^{2/3}n^{2/3} + m + n)$ incidences. Erdős also conjectured that no point-line configuration has an asymptotically larger number of incidences. This conjecture has been proven by Szemerédi and Trotter [15] in 1983.

Theorem 1.1 (The Szemerédi-Trotter theorem). *Let $\mathcal{P}$ be a set of m points and let $\mathcal{L}$ be a set of n lines, both in $\mathbb{R}^2$. Then $I(\mathcal{P}, \mathcal{L}) = O(m^{2/3}n^{2/3} + m + n)$.*

Szemerédi and Trotter's original proof is rather involved. Instead, we present an elegant proof by Székely [14].

2 The crossing lemma

The *crossing number* of a graph $G = (V, E)$, denoted $\mathrm{cr}(G)$, is the smallest integer k such that we can draw G with k edge crossings. Figure 1(a) depicts a drawing of K_5 with a single crossing (where the *complete graph* K_m has m vertices and an edge between every two vertices). Since it is known that K_5 cannot be drawn without crossings, we have $\mathrm{cr}(K_5) = 1$. Given a graph a graph $G = (V, E)$, we are interested in a lower bound for $\mathrm{cr}(G)$ with respect to $|V|$ and $|E|$.

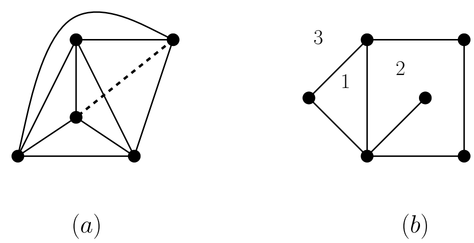

Figure 1: (a) A drawing of K_5 with a single crossing. (b) Two bounded faces and one unbounded.

We consider a connected graph $G = (V, E)$ with v vertices, e edges, and $\mathrm{cr}(G) = 0$ (that is, a *planar graph*). More specifically, we consider a drawing of G in the plane with no crossings. The *faces* of this drawing are two-dimensional maximal connected regions that are bounded by the edges (including one outer, infinitely large region; e.g., see Figure 1(b)). Denote by f the number of faces in the drawing of G. According to *Euler's formula* (also known as Euler's polyhedron formula), we have

$$v + f = e + 2. \tag{1}$$

The formula does not hold for graphs that are not connected.

Every edge is either on the boundary of two faces or has both of its sides on the boundary of the same face. Moreover, the boundary of every face of G consists of at least three edges. Thus, we have $2e \geq 3f$. Plugging this into (1) yields

$$e = v + f - 2 \leq v + \frac{2e}{3} - 2.$$

That is, for any graph $G = (V, E)$ with $\mathrm{cr}(G) = 0$, we have[1]

$$|E| \leq 3|V| - 6. \tag{2}$$

This leads to our first lower bound on $\mathrm{cr}(G)$.

Lemma 2.1. *For any graph $G = (V, E)$, we have* $\mathrm{cr}(G) \geq |E| - 3|V| + 6$.

Proof. Consider a drawing of G in the plane that minimizes the number of crossings. Let $E' \subset E$ be a maximum subset of the edges such that no two edges of E' intersect in the drawing. By (2), we have $|E'| \leq 3|V| - 6$. Since every edge of $E \setminus E'$ intersects at least one edge of E', and since $|E \setminus E'| \geq |E| - 3|V| + 6$, there are at least $|E| - 3|V| + 6$ crossings in the drawing. □

[1] This is also valid for non-connected graphs, since the number of edges in Euler's formula decreases when the graph is not connected.

Since K_5 has five vertices and ten edges, Lemma 2.1 yields the correct value $\text{cr}(K_5) = 1$. However, in general the bound of the lemma is rather weak. For example, it is known that $\text{cr}(K_n) = \Theta(n^4)$ while Lemma 2.1 implies only $\text{cr}(K_n) = \Omega(n^2)$. We now amplify the lower bound of Lemma 2.1 by combining it with a probabilistic argument.

Lemma 2.2 (The crossing lemma). *Let $G = (V, E)$ be a graph with $|E| \geq 4|V|$. Then* $\text{cr}(G) = \Omega\left(|E|^3/|V|^2\right)$.

Proof. Consider a drawing of G with $\text{cr}(G)$ crossings. Set $p = \frac{4|V|}{|E|}$, and notice that by the assumption we have $0 < p \leq 1$. We remove every vertex of V from the drawing with probability $1 - p$ (together with the edges that are adjacent to it). Let $G' = (V', E')$ denote the resulting subgraph, and let c' denote the number of crossings that remain in the drawing.

To avoid confusion with the set E, we denote expectation of a random variable as $\mathbb{E}[\cdot]$. Since every vertex remains with probability p, we have $\mathbb{E}[|V'|] = p|V|$. Since every edge remains if and only if its two endpoints remain, we have $\mathbb{E}[|E'|] = p^2|E|$. Finally, since each crossing remains if and only if the two corresponding edges remain, we have $\mathbb{E}[c'] = p^4\text{cr}(G)$. By linearity of expectation

$$\begin{aligned}\mathbb{E}[c' - |E'| + 3|V'|] &= p^4\text{cr}(G) - p^2|E| + 3p|V| \\ &= \frac{4^4|V|^4}{|E|^4}\text{cr}(G) - \frac{4^2|V|^2}{|E|^2} \cdot |E| + \frac{4|V|}{|E|} \cdot 3|V| \\ &= \frac{4^4|V|^4}{|E|^4}\text{cr}(G) - \frac{4|V|^2}{|E|}. \end{aligned} \tag{3}$$

Since this is the expected value, there exists a subgraph $G^* = (V^*, E^*)$ with c^* crossings remaining from the drawing of G, such that

$$c^* - |E^*| + 3|V^*| \leq \frac{4^4|V|^4}{|E|^4}\text{cr}(G) - \frac{4|V|^2}{|E|}.$$

By Lemma 2.1, we have $c^* \geq |E^*| - 3|V^*| + 6$. Combining this with (3) implies

$$0 < 6 \leq c^* - |E^*| + 3|V^*| \leq \frac{4^4|V|^4}{|E|^4}\text{cr}(G) - \frac{4|V|^2}{|E|}.$$

Tidying up this inequality yields the bound asserted in the lemma. □

It can be easily checked that the bound of Lemma 2.2 indeed implies $\text{cr}(K_n) = \Omega(n^4)$. This lemma was originally derived in [1, 8].

3 Szemerédi-Trotter via the crossing lemma

We are now ready to prove Theorem 1.1. To help the reader, we first repeat the statement of the theorem.

Theorem 1.1. *Let $\mathcal{P}$ be a set of m points and let $\mathcal{L}$ be a set of n lines, both in $\mathbb{R}^2$. Then* $I(\mathcal{P}, \mathcal{L}) = O(m^{2/3}n^{2/3} + m + n)$.

Proof. We write $\mathcal{L} = \{\ell_1, \ldots, \ell_n\}$ and denote by m_i the number of points of $\mathcal{P}$ that are on ℓ_i. Notice that $I(\mathcal{P}, \mathcal{L}) = \sum_{i=1}^n m_i$. We may remove any line ℓ_i that satisfies $m_i = 0$, since this would have no effect on the number of incidences.

We build a graph $G = (V, E)$ as follows. Every vertex of V corresponds to a point of $\mathcal{P}$. For $v, u \in V$, we have $(v, u) \in E$ if v and u correspond to consecutive points along one of the lines of $\mathcal{L}$. Notice that ℓ_i corresponds to exactly $m_i - 1$ edges of E. Thus, we have $|V| = m$ and $|E| = \sum_{i=1}^n (m_i - 1) = I(\mathcal{P}, \mathcal{L}) - n$.

If $|E| < 4|V|$, then we immediately have $I(\mathcal{P}, \mathcal{L}) = O(m + n)$, as required. If $|E| \geq 4|V|$, then by Lemma 2.2 we have

$$\text{cr}(G) = \Omega\left(\frac{(I(\mathcal{P}, \mathcal{L}) - n)^3}{m^2}\right). \tag{4}$$

We next draw G according to the point-line configuration — every vertex is at the corresponding point and every edge is the corresponding line segment. Since every crossing in the drawing corresponds to an intersection of two lines of $\mathcal{L}$, and since every two lines intersect at most once, we have $\text{cr}(G) \leq \binom{n}{2} = O(n^2)$. Combining this with (4) implies

$$\frac{(I(\mathcal{P}, \mathcal{L}) - n)^3}{m^2} = O(n^2).$$

Rearranging this equation yields $I(\mathcal{P}, \mathcal{L}) = O(m^{2/3}n^{2/3} + n)$, as asserted. □

In the proof of Theorem 1.1 we used a common combinatorial method called *double counting*. The method is based on counting some quantity in two different ways, to obtain new information about a different quantity. In the above proof we counted the number of crossings in two different ways, and by comparing these two bounds we obtained a bound for the number of incidences. We will encounter this technique rather frequently in this course.

4 The unit distances problem

The *unit distances problem* is one of the main open problems of discrete geometry. Although it has proven to be extremely difficult to solve, this problem is very easy to state: How many pairs of points in a planar set of n points could be at unit distance from each other? We denote the maximum number of such pairs as $u(n)$. By taking a set of n points equally spaced on a line, we immediately obtain $u(n) \geq n - 1$. Erdős [3] introduced the problem in 1946 and derived the bounds $u(n) = O(n^{3/2})$ and $u(n) = \Omega(n^{1+c/\log\log n})$ (for some constant c). Even though this is such a central problem in discrete geometry, in over seven decades the lower bound was never improved and the upper bound was improved only once. The bound $u(n) = O(n^{4/3})$ was derived by Spencer, Szemerédi, and Trotter [13] in 1984.

Consider a set $\mathcal{P} \subset \mathbb{R}^2$ of n points such that the number of unit distances between pairs of points of $\mathcal{P}$ is $u(n)$. We draw a unit circle (i.e., a circle of radius one) around each point of $\mathcal{P}$, and denote the set of these n circles as $\mathcal{C}$. Every two points $p, q \in \mathcal{P}$ that determine a unit distance correspond to two incidences in $\mathcal{P} \times \mathcal{C}$ — the circle around p is incident to q and vice versa. Thus, to bound $u(n)$ it suffices to bound the maximum number of incidences between n points and n unit circles (it is not hard to show that the two expressions are in fact asymptotically equivalent).

Theorem 4.1. *Let $\mathcal{P}$ be a set of n points and let $\mathcal{C}$ be a set of n unit circles, both in $\mathbb{R}^2$. Then $I(\mathcal{P},\mathcal{C}) = O(n^{4/3})$.*

Notice that the theorem immediately implies the best known bound $u(n) = O(n^{4/3})$.

Proof. We imitate the proof of Theorem 1.1. We write $\mathcal{C} = \{c_1, \ldots, c_n\}$ and denote by m_i the number of points of $\mathcal{P}$ that are on c_i. Notice that $I(\mathcal{P},\mathcal{C}) = \sum_{i=1}^{n} m_i$. We may remove any circle c_i that satisfies $m_i < 3$, since these circles yield at most $2n$ incidences.

We build a graph $G = (V, E)$ as follows. Every vertex of V corresponds to a point of $\mathcal{P}$. For $v, u \in V$, we have $(v, u) \in E$ if v and u are consecutive points along at least one of the circles of $\mathcal{C}$. Notice that c_i corresponds to exactly m_i edges of E, and that every edge originates from at most two unit circles. Thus, we have $|V| = n$ and $|E| \geq \sum_{i=1}^{n} m_i/2 = I(\mathcal{P},\mathcal{C})/2$.

If $|E| < 4|V|$, then we immediately have $I(\mathcal{P},\mathcal{C}) = O(n)$, as required. If $|E| \geq 4|V|$, then by Lemma 2.2 we have

$$\mathrm{cr}(G) = \Omega\left(\frac{I(\mathcal{P},\mathcal{C})^3}{n^2}\right). \tag{5}$$

We next draw G according to the point-circle configuration — every vertex is at the corresponding point and every edge is the corresponding circle arc. Since every crossing in the drawing corresponds to an intersection of two circles of $\mathcal{C}$, and since every two circles intersect at most twice, we have $\mathrm{cr}(G) \leq 2\binom{n}{2} = O(n^2)$. Combining this with (5) implies

$$\frac{I(\mathcal{P},\mathcal{C})^3}{n^2} = O(n^2).$$

Rearranging this equation yields $I(\mathcal{P},\mathcal{C}) = O(n^{4/3})$, as asserted. □

The common belief seems to be that the following conjecture holds.

Conjecture 4.2. $u(n) = O(n^{1+\varepsilon})$ *for any* $\varepsilon > 0$.

5 The distinct distances problem

The *distinct distances* problem can be considered as the "twin" problem of the unit distances problem, and it was introduced in the same 1946 paper of Erdős [3]. The question asks for the minimum number of distinct distances that can be determined by a set of n points in the plane (that is, we consider the minimum cardinality of the set of distances that are determined by at least one pair of points). We denote this quantity as $d(n)$.

It can be easily verified that a set of n points that are equally spaced on a line determines $n-1$ distinct distances. Thus, we have $d(n) \le n-1$. A better bound was already in Erdős' original paper. Specifically, Erdős considered a $\sqrt{n} \times \sqrt{n}$ integer lattice. The number of distances that are determined by this set is an immediate corollary of a result from number theory.

Theorem 5.1. (Landau and Ramanujan [2, 7]) *The number of positive integers smaller than n that are the sum of two squares is $\Theta(n/\sqrt{\log n})$.*

Every distance in the $\sqrt{n} \times \sqrt{n}$ integer lattice is the square root of a sum of two squares between 0 and n. Thus, Theorem 5.1 implies that the number of distinct distances in this case is $\Theta(n/\sqrt{\log n})$.

Theorem 5.2 (Erdős 1946). $d(n) = O(n/\sqrt{\log n})$.

For the lower bound on $d(n)$, we begin by deriving Erdős' original bound (although with a different proof).

Claim 5.3. $d(n) = \Omega(n^{1/2})$.

Proof. Consider an n point set $\mathcal{P}$ and two points $v, u \in \mathcal{P}$. Let d_v denote the number of distinct distances between v and $\mathcal{P} \setminus \{v\}$. Notice that the points of $\mathcal{P} \setminus v$ are contained in d_v circles that are centered at v. We denote this set of circles as $\mathcal{C}_v$. We define d_u and $\mathcal{C}_u$ symmetrically. Each of the $n-2$ points of $\mathcal{P} \setminus \{v, u\}$ is contained in the intersection of a circle of $\mathcal{C}_v$ and a circle of $\mathcal{C}_u$. Since the number of such intersections is at most $2|\mathcal{C}_v||\mathcal{C}_u| = 2d_v d_u$, we have $2d_v d_u \ge n-2$, which in turn implies $\max\{d_v, d_u\} = \Omega(n^{1/2})$. (An example is depicted in Figure 2.) $\square$

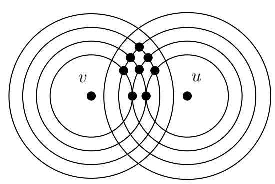

Figure 2: The points of $\mathcal{P} \setminus \{v, u\}$ are contained in the intersections of $\mathcal{C}_v$ and $\mathcal{C}_u$.

We now derive an improved bound by using incidences. This bound was originally derived by Moser [9] in 1952, by using a different argument.

Claim 5.4. $d(n) = \Omega(n^{2/3})$.

Proof. Consider an n point set $\mathcal{P}$ that determines d distinct distances, and denote these distances as $D = \{\delta_1, \ldots, \delta_d\}$. Let $\mathcal{C}$ denote the set of nd circles with a center in $\mathcal{P}$ and a radius in D. The claim is proved by double counting $I(\mathcal{P}, \mathcal{C})$.

For every point $v \in \mathcal{P}$, the points of $\mathcal{P} \setminus \{v\}$ are contained in the d circles of $\mathcal{C}$ centered at v. Thus, $I(\mathcal{P}, \mathcal{C}) = n(n-1)$.

Next, let $\mathcal{C}_i$ denote the subset of circles of $\mathcal{C}$ with radius δ_i. By Theorem 4.1, we have $I(\mathcal{P}, \mathcal{C}_i) = O(n^{4/3})$ (notice that Theorem 4.1 is valid as long as all of the circles have the same radii). Thus, we have

$$I(\mathcal{P}, \mathcal{C}) = \sum_{i=1}^{d} I(\mathcal{P}, \mathcal{C}_i) = O(dn^{4/3}).$$

Combining our two bounds for $I(\mathcal{P}, \mathcal{C})$ immediately implies the assertion of the claim. □

A simpler proof of Claim 5.4 is as follows. Each of the $\binom{n}{2}$ pairs of points determines a distance. By Theorem 4.1 any distance occurs $O(n^{4/3})$ times, so to cover $\Theta(n^2)$ pairs there must be $\Omega(n^{2/3})$ distinct distances. We presented the longer proof since it sheds more light about how to use incidences.

Both proofs show that the distinct distances problem can in some sense be reduced to the unit distances problem. An upper bound of $u(n) = O(n^{1+c/\log\log n})$ would yield an almost tight bound for $d(n)$.

In 2010, Guth and Katz [6] proved the almost tight lower bound $d(n) = \Omega(n/\log n)$. Unlike the proof that we saw, this is a "deep" result that requires combining tools from several different fields. One of the peaks of this course will be a proof of this result. Even though the distinct distances problem is solved (up to a gap of $\sqrt{\log n}$), some interesting variants of it are still wide open. A couple of examples:

- The problem is still open in $\mathbb{R}^d$ for any $d \geq 3$. Erdős constructed a set of n points in $\mathbb{R}^d$ that determines $\Theta(n^{2/d})$ distinct distances, and conjectured that no set determines a smaller number. So far no one managed to apply the polynomial method even for the case of $\mathbb{R}^3$.

- In $\mathbb{R}^2$, characterizing the n point sets that determine $O(n/\log n)$ distinct distances seems to be a very difficult problem. The past several decades yielded many conjectures but hardly any results concerning this.

For a list of many other related open problems, see [11].

6 A problem about unit area triangles

Before concluding this chapter we briefly mention another problem, which can be considered as one of the many generalizations of the unit distances problem. The question is: What is the maximum number of unit area triangles that have their vertices in a set of n points in $\mathbb{R}^2$?

Consider two points $p, q \in \mathbb{R}^2$ at a distance of d from each other. A key observation is that p and q form a unit area triangle with a third point r if and only if r is on one of the two lines that are parallel to the segment pq and at a distance of $2/d$ from this segment (e.g., see Figure 3(a)). Thus, by taking two parallel lines at a distance of 2 from each other, and placing $n/2$ points at unit intervals on each, we obtain $\Theta(n^2)$ unit triangles (e.g., see Figure 3(b)). Erdős and Purdy [5] showed that that a

$\sqrt{\log n} \times n/\sqrt{\log n}$ section of the integer lattice determines $\Omega(n^2 \log\log n)$ triangles of the same area.

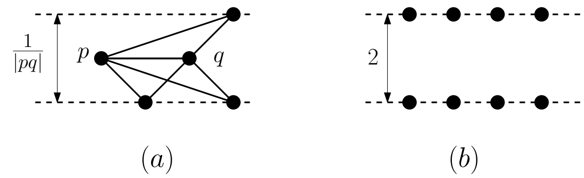

Figure 3: (a) The points that for a unit triangle with p and q are on two parallel lines. (b) A configuration with $\Theta(n^2)$ unit triangles.

Claim 6.1 (Pach and Sharir [12]). *Every planar set of n points determines $O(n^{7/3})$ unit triangles.*

Proof. Consider a set $\mathcal{P}$ of n points. For a point $p \in \mathcal{P}$ we bound the number of unit triangles that are determined by p and two other points of $\mathcal{P}$. For any $q \in \mathcal{P} \setminus \{p\}$, we denote by d_{pq} the distance between p and q, and by ℓ_{pq}, ℓ'_{pq} the lines that are parallel to the segment pq and at a distance of $2/d_{pq}$ from it. We set $\mathcal{L}_p = \{\ell_{pq}, \ell'_{pq} : p, q \in \mathcal{P}\}$. Notice that any line $\ell_{p,q}$ can originate from at most two points $q \in \mathcal{P} \setminus \{p\}$. Thus, $n - 1 \leq |\mathcal{L}_p| \leq 2n - 2$. The number of unit triangles that involve p is at least $I(\mathcal{P}, \mathcal{L}_p)/2$. By Theorem 1.1, we have $I(\mathcal{P}, \mathcal{L}_p) = O(n^{4/3})$. The assertion of the claim is obtained by summing this bound over every $p \in \mathcal{P}$. □

Recently, Raz and Sharir [10] improved this bound to $O(n^{20/9})$ by considering incidences with two-dimensional surfaces in $\mathbb{R}^4$.

References

[1] M. Ajtai, V. Chvátal, M. M. Newborn, and E. Szemerédi, Crossing-free subgraphs, *Annals Discrete Math.* **12**:9–12, 1982.

[2] B. C. Berndt and R. A. Rankin, *Ramanujan: Letters and Commentary*, Amer. Math. Soc., Providence, RI, 1995.

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

[4] P. Erdős, On some of my favourite theorems, *Combinatorics, Paul Erdős is Eighty,* Vol. 2 (D. Miklós et al., eds.), Bolyai Society Mathematical Studies 2, Budapest, 1996, 97–132.

[5] P. Erdős and G. Purdy, Some extremal problems in geometry, *J. Combinat. Theory* **10** (1971), 246–252.

[6] L. Guth and N.H. Katz, On the Erdős distinct distances problem in the plane, *Annals Math.* **181** (2015), 155–190.

[7] E. Landau, Über die Einteilung der positiven ganzen Zahlen in vier Klassen nach der Mindeszahl der zu ihrer additiven Zusammensetzung erforderlichen Quadrate, *Arch. Math. Phys.* **13** (1908), 305–312.

[8] F. T. Leighton, *Complexity issues in VLSI: optimal layouts for the shuffle-exchange graph and other networks*, MIT press, 1983.

[9] L. Moser, On the different distances determined by n points, *American Mathematical Monthly* **59** (1952), 85–91.

[10] O. E. Raz and M. Sharir, Unit-area triangles in the plane: Theme and variations, *Proc. 31st ACM Symp. on Computational Geometry* (2015), to appear.

[11] Adam Sheffer, Distinct Distances: Open Problems and Current Bounds, arXiv:1406.1949.

[12] J. Pach and M. Sharir, Repeated angles in the plane and related problems, *J. Combinat. Theory Ser. A* **59** (1992), 12–22.

[13] J. Spencer, E. Szemerédi, and W. T. Trotter, Unit distances in the Euclidean plane, *Graph Theory and Combinatorics* (ed. B. Bollobás), Academic Press, 1984.

[14] L. Székely, Crossing numbers and hard Erdős problems in discrete geometry, *Combinatorics, Probability and Computing* **6** (1997), 353–358.

[15] E. Szemerédi and W. T. Trotter, Extremal problems in discrete geometry, *Combinatorica* **3** (1983), 381–392.