The Distinct Distances Problem

Darren Zheng

April 2024

Abstract

Suppose you were given n points to place in a plane. What is the minimum size of the set of distances between each pair of points? An asymptotic lower bound was proposed by Paul Erdős in 1946, but a full proof remains elusive. In this paper, we seek to give an overview of the polynomial methods from algebraic geometry that Larry Guth and Nets Katz developed to tackle such a problem in 2010. In addition, we will describe some other fun applications of the ESGK framework they used.

Contents

1 Problem **1**

2 Background **2**

3 Elekes-Sharir-Guth-Katz Method **3**

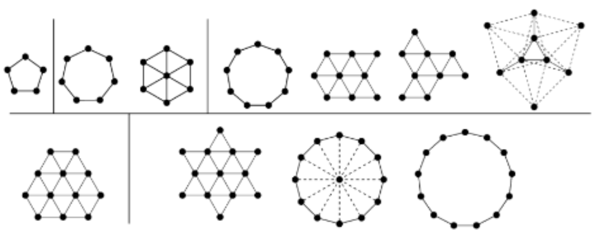
4 Further Problems **4**

1 Problem

Consider a set of n points in the plane $\mathbb{R}^2$. For each pair of points, we compute the distance between them under the usual Euclidean metric and record these points in a set S. What might be the upper and lower bounds on $|S|$? We will define $D(n)$ to be the minimum over all possible configurations of n points.

To get some intuition, we can look at the values of $D(n)$ for a few small values of n. For the case of $n = 3$, we can use the points as vertices of an equilateral triangle, so $D(3) = 1$. For $n = 4, 5, 6$, we find that a regular polygon achieves the minimum $D(4) = 2$, $D(5) = 2$, and $D(6) = 3$ respectively. Here are a few other configurations given for some values of $D(n)$.

n	1	2	3	4	5	6	7	8	9	10	11	12	13
$v(n)$	0	1	1	2	2	3	3	4	4	5	5	5	6



Sets with k distinct distances, $2 \leq k \leq 6$, and maximum number of points

Figure 1: Image from page 200 of [4]

As we can see, this minimum is not always achieved using a regular polygon; for $n = 12$, a dodecagon achieves 6 distinct distances, but the configuration in the bottom left above achieves only 5.

We can establish a naive upper bound with $D(n) = O(n)$. Given n points, we can place them equally spaced along a line of length n. It is clear that by considering the distances formed from the first point to each of the other $n - 1$ points, we get that the distance set $S = \{1, 2, \ldots, n\}$, so $|S| = n - 1$. We can get a better upper bound asymptotically by considering a particular square grid of points. First, let's state a fairly well-known number theoretic fact.

Lemma 1. *The number of integers smaller than n that can be written as the sum of two squares is $O(\frac{n}{\sqrt{\log n}})$.*

Proof. This is a result from analytic number theory and is related to the Landau-Ramanujan constant. A full proof can be found in [2]. □

Theorem 1.1 (Upper Bound on Distinct Distances). $D(n) = O(\frac{n}{\sqrt{\log n}})$

Proof. Let $P = \{(a, b) \in \mathbb{Z}^2 | 0 \leq a, b < \sqrt{n}\}$. Each distance between any pair of points from P is at most n and is by definition, the square root of a sum of two squares. Then, as an immediate corollary of Lemma 1, we get that $D(n) = O(\frac{n}{\sqrt{\log n}})$, as we've found a specific configuration that achieves achieves this. □

On the other hand, establishing a lower bound is far more difficult. Erdős conjectured that this bound above was tight in the sense that $D(n) = \Theta(\frac{n}{\sqrt{\log n}})$ ([5]). In 2010, Guth and Katz established a lower bound of $\Omega(\frac{n}{\log(n)})$, which is relatively close, only off by a factor of $\sqrt{\log n}$.

2 Background

We introduce some necessary theorems for the next section.

Definition 1. A *rigid motion* (or isometry) of the plane is a transformation that preserves distances. We call a rigid motion *proper* if it also preserves orientation.

We can see an example of such transformations below. It is not hard to note that the only proper rigid motions

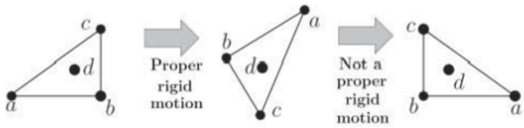


Figure 2: The second one is not proper, as it involves flipping across a line (mirror image); Traveling along abc requires a left turns pre-transformation, and a right turn post-transformation. Image taken page 100 of [1]

involve some combination of rotations and translations of the plane. We will denote a rotation about some central point c as R_c, where R_O is a rotation about the origin, and a translation by T_v, where the subscript denotes where we send the origin. Thus, it is clear that every sequence of rotations and translations can be described by a single proper rigid motion. The space of proper rigid motions forms a group, as rotations and translations are clearly invertible.

Lemma 2. *Any rigid motion which fixes* 3 *non-collinear points is the identity.*

Proof. Let the three fixed points be $a, b, c \in \mathbb{R}^2$. Then, a rigid motion that fixes a and b must also fix the unique line L through a and b. Suppose the rigid motion contains a reflection. Then, it must be across the line L. In that case, we get a contradiction, as c, a point not on L, belongs to one of the two regions formed by L and remains there post-transformation. With no reflection, it must be a proper rigid motion involving only translations or rotations. In fact, each region formed by L must be mapped onto itself.

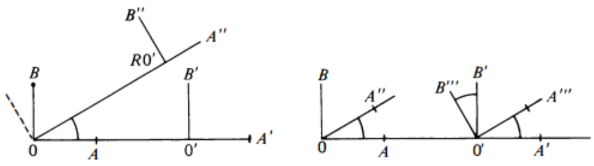
Since any arbitrary point p can uniquely be described by the combination of what side of L it lies on and its distance from a and b, our distance-preserving transformation must be unique and the identity. □

Theorem 2.1. *Every proper rigid motion can be written uniquely as $T \cdot R_c$ for some chosen point c. Stated equivalently, every proper rigid motion is some rotation about a fixed center followed by a translation.*

Proof. First, we note two useful identities.

$$R_c T_{v_1} R_c^{-1} = T \text{ and } T_{v_1} R_p T_{v_1}^{-1} = R$$

We want to show that conjugation of a translation by rotation is some translation and conjugation of a rotation by translation is some rotation. For the left-hand-side identity, T_{v_1} must send c to $c + v_1$. Then, set $T = T_{R_c(c+v_1)-c}$, the translation that sends c to $R_c(c + v_1)$. One can verify that both $R_c T_{v_1}$ and $T_{c+v_1} R_c$ have the same effect on any arbitrary triangle acb. Thus, each must be the identity and the equality is established. A very similar argument works for the second part. Consider the image below for a visual proof.



Now, the first equation above allow us to transform any ordered sequence of translations and rotations into a form where the rotations occur first, then the translations. If we choose a specific point c, we can rewrite the rotational part $R_{v_1}R_{v_2}\dots R_{v_i}$ as some proper rigid motion $T_a R_c T_a^{-1}$ using the second equation. Then, apply the first identity again to swap the order $R_c T_a^{-1} = T_b R_c$. We get the desired form of the rigid motion expressed as a rotation, then translation. This decomposition of $\tau = TR$ is sometimes called the standard form. $\square$

Theorem 2.2. *Let P contain n points in $\mathbb{R}^2$ and define $Q = \{(a,p,b,q) \in P^4 | d(a,p) = d(b,q)\}$. Then, each element of Q corresponds to a proper rigid motion. In other words, $(a,p,b,q) \in Q$ if and only if there exists a proper rigid motion τ sending the line ap to bq (the transformation sends the point a to b and point p to q).*

Proof. The backwards direction is immediate. If there is a proper rigid motion τ sending ap to bq, then $d(a,p) = d(b,q)$, so $(a,p,b,q) \in Q$.

For the forwards direction, it suffices to show that a proper rigid motion sending a to b and p to q must be unique. This really is a corollary of the previous Theorem 2.1. Suppose we had a proper rigid motions $\tau = \pi$ sending a to b and p to q. Choosing any point c, we write each in the same standard form as TR_c. Then, note that letting M be the unique translation that sends a to b, there will always be a rotation R_a about a that sends b to q after M (this requires that $d(a,b) = d(b,q)$). Thus, $\tau = \pi = R_a M$. By the identity in the proof of the previous theorem, we may rewrite this uniquely as TR_c for some translation T and rotation R_c, and so we are done. $\square$

3 Elekes-Sharir-Guth-Katz Method

The key insight by György Elekes and Micha Sharir is to reduce the problem of counting these distinct distances to counting the intersection of parabolas in $\mathbb{R}^3$. Later, Guth and Katz reduced this further to be about lines in $\mathbb{R}^3$. This general method is referred to as the ESGK reduction. Then, the polynomial methods in the previous section and some algebraic geometry are applied to finish the proof.

We can now prove our desired theorem.

Theorem 3.1 (Lower Bound on Distinct Distances). $D(n) = \Omega(\frac{n}{\log n})$

Proof. Let the set P contain n points in $\mathbb{R}^2$. Let, where $d(x,y)$ is the Euclidean metric,

$$Q = \{(a,p,b,q) \in P^4 | d(a,p) = d(b,q)\}$$

We should understand Q to be an ordered pair of directed lines, so the tuple $(a,p,b,q) \neq (b,q,a,p)$ and $(a,p,b,q) \neq (p,a,q,b)$. We even allow the quadruple $(a,a,a,a) \in Q$.

Let the distinct distances set be $S = \{r \in \mathbb{R} | \exists a,b \in P,$ such that $d(a,b) = r\}$ and set $\delta_1, \delta_2, \dots, \delta_j$ to be the elements of S (Some δ_x will be 0, since $(a,a) \in P^2$). Define

$$E_i = \{(a,p) \in P^2 | d(a,p) = \delta_i\}$$

Now, each of ordered pairs in P^2 are contained in one of these E_i. Thus, $\sum_{i=1}^{|S|} |E_i| = n^2$. Next, note there are $|E_i|^2$ elements of Q that achieve a specific distance δ_i, since each element comes from selecting two, possibly duplicate, pairs of points from E_i. Thus, we can also express $|Q| = \sum_{i=1}^{|S|} |E_i^2|$. Combining these facts, we find using Cauchy-Schwarz that

$$\begin{aligned} |Q| &= \sum_{i=1}^{|S|} |E_i^2| \\ &\geq \frac{1}{|S|}(|E_i|)^2 = \frac{n^4}{|S|} \end{aligned}$$

We've reduced this problem to showing that $|Q| = O(n^3 \log n)$. In that case, we'd be able to rearrange the above into $|S| = O(\frac{n^4}{n^3 \log(n)}) = O(\frac{n}{\log(n)})$ as desired.

To establish $|Q| = O(n^3 \log n)$, we want to use Theorem 2.2, which provides an equivalent definition of Q. Our problem becomes counting how many different quadruples correspond to each proper rigid motions, since two quadruples may correspond to the same τ. For example, if we take 3 collinear points $a, b, c \in P$ and 3 points on a parallel line $p, q, r \in P$, then (a, p, b, q), (b, q, c, r), and (a, p, c, r) all correspond to the same τ.

Let's study if τ is just a translation first. Note τ is invertible. Suppose we are given the first three points a, p, b of our quadruple, which means $\tau(a) = p$. Then, it is clear geometrically there is only one point $q = \tau(b)$ for (a, p, b, q) to correspond to our translation τ. Thus gives us that $O(n^3)$ quadruples in Q correspond to a given translation τ.

Via Theorem 2.1, we only have the remaining cases where our τ is a clockwise rotation about some point $c = (c_x, c_y) \in \mathbb{R}^2$ with an angle θ. We can parameterize this as $(c_x, c_y, \cot \frac{\alpha}{2})$. Consider the following line in $\mathbb{R}^3$ that traces out the rotation of a point $a = (a_x, a_y)$ to $b = (b_x, b_y)$; one can verify this by a tedious, but straightforward computation.

$$l_{a,b} = t(\frac{a_x + b_x}{2}, \frac{a_y + b_y}{2}, 1) + (\frac{b_y - a_y}{2}, \frac{a_x - b_x}{2}, 0)$$

If we project this line $l_{a,b}$ to $\mathbb{R}^2$ via the xy-plane, we still get a line. The slope of this line is $\frac{a_x - b_x}{b_y - a_y}$, which means this line is perpendicular to the line ab. In fact, this line also contains the midpoint of ab by looking at the y-intercept. Thus, $l_{a,b}$ is a bisector of ab.

Now, consider some quadruple $(a, p, b, q) \in Q$. Via the construction above, if $l_{a,b}$ and $l_{p,q}$ as lines in $\mathbb{R}^3$ have a non-empty intersection, then they are both the unique rotation by Theorem 2.2 that send ap to bq. Thus, a tuple corresponds to a rotation if and only if these lines in $\mathbb{R}^3$ intersect.

One now must argue that the n^2 lines of the form $l_{a,b}$ for some $a, b \in P$ is $O(n^3 \log(n))$. This turns out to not be true in general; take 2 sets of n parallel lines, such that the sets themselves are not parallel. Then, we get $O(n^4)$, which isn't sufficient. Luckily a careful analysis of our line construction shows that such a case cannot arise. This and more details are found in both chapter 9 of [1] or the original paper [8]. ☐

4 Further Problems

The Elekes-Sharir framework uses tools from algebraic geometry and polynomial partitioning to solve incidence problems, usually after restating them in a certain way. Success with this general method can be seen in the resolution of the joints problem in $\mathbb{R}^3$ and the finite field Kakeya conjecture. For the sake of space, we state them below. We encourage the reader to read the original papers, as the proofs of these thought to be impossible conjectures are remarkably short using the polynomial method.

Theorem 4.1 (Joints Problem). *Let L be a set of lines in $\mathbb{R}^3$. A joint is a point that lies at the intersection of 3 non-planar lines in L (non-planar meaning there is no plane containing all 3 lines). The maximum number of joints when $|L| = n$ is $\Theta(n^{\frac{3}{2}})$.*

Proof. [6] proves a generalization in higher dimensions. ☐

Theorem 4.2 (Finite Field Kakeya Problem). *Let $\mathbb{F}$ be the finite field of q elements. A Kakeya set K is a subset of points in $\mathbb{F}^n$, such that every line of the form $ax + y$, where $a \in \mathbb{F}$ and $y \in \mathbb{F}^n$ is contained in K. A lower bound on $|K|$ is $C_n \cdot q^n$, where C_n is a constant only dependent on n.*

Proof. [7] ☐

The original distinct distances problem posed by Erdős has still not been completely solved, as we note the gap of $\sqrt{\log(n)}$. The proof suggested above does not seem possible to improve; there exists a set of points, such that the set Q, as defined in the proof of Theorem 3.1, achieves $\Theta(\frac{n}{\log(n)})$.

Proposition 1. *Let $P = \{1, 2, 3, \ldots, \sqrt{n}\}^2$ be a square lattice of points. Let Q be the same set as defined in Theorem 3.1 for P. Then, $|Q| = \Omega(\frac{n}{\log n})$.*

Thus, markedly new proof techniques/ideas are required to close this remaining gap.

There are still many open variants of the distinct distances problem. You can ask the same distinct distances problem under other metrics, such as the L_p family, or placing points in higher dimensions $\mathbb{R}^d$. Another variant is the unit distinct distances problem: What is the maximum number of pairs of points at distance 1 given n points in $\mathbb{R}^2$? These problems remain open with relatively little progress, signaling the difficulties we face in answering even relatively simple-to-state problems in incidence theory.

Acknowledgements. I want to thank my mentor, Jae Ho Choi, for his willingness to talk about any and all math. Additionally, I want to thank the UPenn math directed reading program, and its organizers Maxine Calle and Leo Guilhoto, for organizing such a fantastic program.

References

[1] Adam Sheffer, *Polynomial Methods and Incidence Theory*, https://doi.org/10.1017/9781108959988

[2] William Judson LeVeque, *Fundamentals of number theory*, ISBN: 0201042878

[3] Adam Sheffer, *Some Plane Truths*, https://adamsheffer.wordpress.com/2013/05/25/the-elekes-sharir-framework-part-1/

[4] Peter Brass, William O. J. Moser, János Pach, *Research Problems in Discrete Geometry*, https://doi.org/10.1007/0-387-29929-7

[5] Paul Erdős, *On Sets of Distances of n Points*, https://doi.org/10.2307/2305092

[6] H. Kaplan, M. Sharir, E. Shustin, *On lines and joints*, https://doi.org/10.48550/arXiv.0906.0558

[7] Zeev Dvir, *On the size of Kakeya sets in finite fields*, https://doi.org/10.1090/S0894-0347-08-00607-3

[8] Larry Guth, Nets Hawk Katz, *On the Erdos distinct distance problem in the plane*, https://doi.org/10.48550/arXiv.1011.4105