Few distinct distances with forbidden 4-point patterns

Talk by Travis Dillon

Notes by Sanjana Das

October 4, 2024

§1 Introduction

This is based on a paper by Terence Tao from this year.

§1.1 Distinct distances

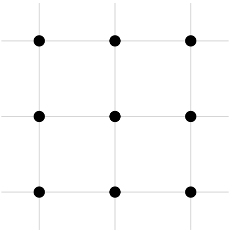
For some context, the starting point is the distinct distances problem.

> **Question 1.1.** What's the minimum number of distinct distances that n points in $\mathbb{R}^2$ can span?

For example, if you drop the n points randomly, you'd expect that all distances between them are different, so you'd get $\binom{n}{2}$ distances. But maybe you could instead put them all on a line (with equal spacing); then you'd have $n-1$ distinct distances, which is much better (our goal is to minimize the number of distances).

In fact, you can do a bit better than this:

> **Theorem 1.2** (Erdős 1946)
> The $\sqrt{n} \times \sqrt{n}$ square grid has $O(n/\sqrt{\log n})$ distinct distances.



(This comes down to some number-theoretic fact about how many integers can be expressed as a sum of two squares.)

Erdős conjectured that this was the right answer. He proved some lower bound, and there was a series of improvements; but since we're not talking about this exact problem, we'll skip to the punchline.

Theorem 1.3 (Guth–Katz)
Every set of n points determines $\Omega(n/\log n)$ distinct distances.

So this problem has been solved up to a factor of $\sqrt{\log n}$.

§1.2 The problem

Question 1.4 (Erdős). Suppose we have a *local* condition that every small subset of points has many distinct distances. Does this force there to be *globally* many distinct distances?

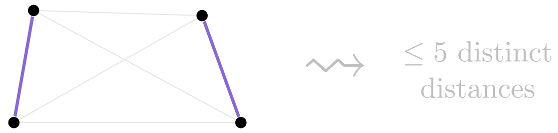
In particular, as a special case of this problem, Erdős had the following conjecture:

Conjecture 1.5 (Erdős) — If every set of 4 points spans at least 5 distances, then the entire set (of n points) spans $\Omega(n^2)$ distances.

Why the number 5? If we replace 4 with k and 5 with r, we get the following definition.

Definition 1.6. We define $\phi(n, k, r)$ as the minimum number of distinct distances spanned by n points in $\mathbb{R}^2$ with the property that every k points span at least r distances.

We can notice right away that $\phi(n, 4, 6) = \binom{n}{2}$ — if we pick any two pairs of points, then the local condition tells you that their distances have to be different (because they form a set of four points).



So the reason for the numbers 4 and 5 is that it's the smallest weakening — if you weaken 6 to 5 then you no longer guarantee that all $\binom{n}{2}$ distances are distinct, but maybe you can still guarantee a quadratic number of distances.

This was a conjecture Erdős brought up several times; the paper we'll discuss essentially says that it's false.

Theorem 1.7 (Tao 2024+)
We have $\phi(n, 4, 5) = O(n^2/\sqrt{\log n})$.

In particular, $\phi(n, 4, 5)$ is asymptotically less than n^2.

There's also work on this problem for various other values of k and r (particularly by Fox–Pach–Suk 2018). For example, it's known that

$$\phi(n, 7, 20) = \Omega(n^2).$$

This is the same type of weakening — if you had $(7, 21)$ then you'd get that all $\binom{n}{2}$ distances are different, and here weakening 21 to 20 still gives you something quadratic. Similarly, it's also known that

$$\phi(n, 8, 26) = \Omega(n^2),$$

which is a one-further decrease compared to $\binom{8}{2} = 28$.

> **Remark 1.8.** Both of these bounds come purely from 'coloring' considerations (you can think of the points as forming a graph, with edges colored by their distances) — for example, the bound on $\phi(n, 8, 26)$ comes from the observation that if you have much fewer than n^2 distances, then you can find some distance repeated 4 times, giving you 8 points with $\binom{8}{2} - 3$ distances.

We're going to discuss the main ideas of Theorem 1.7 (and maybe see some details, depending on time).

§2 Forbidden patterns

There's two big ideas to start with. The first idea is that we're going to look for our set of points by picking n points from a $n \times n$ grid $[n]^2$. This grid already has only $n^2/\sqrt{\log n}$ distinct distances; so if we refine it, we're certainly not going to make any more distances.

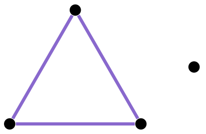
The second idea is that somehow, we want to turn the condition that any four points have at least five distinct distances into a 'forbidden patterns' problem. Both of these ideas come from previous papers — Tao cites several, but a lot of the calculation and work comes from a paper by Dumitrescu (2018).

Proposition 2.1 (Dumitrescu 2018)

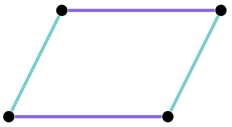
Any set of 4 points with at most 4 distinct distances falls into one of eight specific patterns $\Pi_1, \ldots, \Pi_8$.

We won't write down all the patterns, but we'll write down a few examples. (The proof is to think about what has to happen if you have 4 points but don't have 5 distances.)

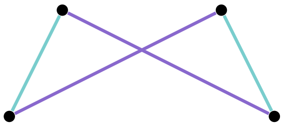
(Π_1) The points form an equilateral triangle with one extra point.



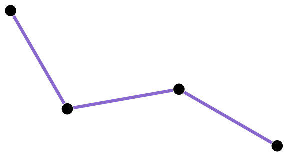
(Π_2) The points form a parallelogram.



(Π_3) The points form an isosceles trapezoid.



(Π_4) The points form a 'path' (where segments along the path have equal length).



The important thing about these patterns is how often they appear in the grid.

Theorem 2.2 (Dumitrescu 2018)

- The number of instances of Π_1 in the grid $[n]^2$ is 0.
- The number of instances of Π_2 is $O(n^6)$.
- The number of instances of $\Pi_3, \ldots, \Pi_8$ are each $O(n^5)$.

We'll prove this in the end if we have time.

§3 First attempts

§3.1 A random construction

Now that we have a 'forbidden patterns' description of the problem, we can make a first attempt at proving Theorem 1.7 — choose a subset $X \subseteq [n]^2$ *randomly*, by including each point independently with probability $1/n$ (since we want a subset of size n).

Then what happens? By whatever concentration method you prefer, we have $|X| = \Theta(n)$ with high probability. Similarly, the number of instances of each Π_i will be concentrated around its mean; this means we'll have $O(n)$ instances of $\Pi_3, \ldots, \Pi_8$, and $O(n^2)$ instances of Π_2.

This is a problem, because we want to get rid of all instances of the Π_i. But we've gotten pretty close: If we now take a further refinement of X where we include each point with probability ε independently (where ε is a small constant), then we'll have $|X| = \Theta(\varepsilon n)$, but the number of instances of Π_i will be $O(\varepsilon^4 n)$ for each $3 \le i \le 8$ (because for a particular instances of Π_i to be kept, we need to keep all four of its points). So if we choose ε to be small enough, then we can delete a point from each of these patterns, giving the following intermediate result.

Proposition 3.1

There is a $\Theta(n)$-sized subset of $[n]^2$ that avoids $\Pi_1, \Pi_3, \Pi_4, \ldots, \Pi_8$.

So we're very close; we've avoided everything except Π_2. But the problem is that this construction *is* going to contain parallelograms.

(This is all from Dumitrescu's paper.)

§3.2 Avoiding parallelograms

As a second attempt, we'll focus directly on parallelograms, and forget about all the other patterns.

Question 3.2. How can we get a $\Theta(n)$-sized subset of $[n]^2$ that doesn't have any parallelograms?

The idea is that instead of doing something probabilistic, we'll do something algebraic. We fix a prime $p \in [n, 2n]$ and consider the parabola over $\mathbb{F}_p$ given by $\{(x, x^2) \mid x \in \mathbb{F}_p\}$. Then we move this parabola to our grid $[n]^2$, and define Y as its intersection with the grid — i.e., we define

$$Y = \{(x, y) \in [n]^2 \mid y \equiv x^2 \pmod{p}\}.$$

Fact 3.3 — The set Y does not contain Π_2.

(The intuition is that Y is a parabola mod p, and an actual parabola doesn't have any parallelograms.)

Proof. Assume for contradiction that Y does contain a parallelogram. Suppose that the first three points of this parallelogram are (x, x^2), $(x+a, (x+a)^2)$, and $(x+b, (x+b)^2)$ (with all coordinates mod p); then the fourth must be

$$(x + a + b, (x + a)^2 + (x + b)^2 - x^2).$$

And for this fourth point to lie on the parabola, we need

$$(x + a)^2 + (x + b)^2 - x^2 \equiv (x + a + b)^2 \pmod{p}.$$

If we do the arithmetic, this simplifies to $2ab \equiv 0 \pmod{p}$ (everything else cancels out). And this is only possible if a or b is 0 mod p; this would mean that our points wouldn't be distinct (since all points on the parabola have distinct x-coordinates). So this is a contradiction. □

This is great — we've avoided parallelograms — but there's a couple of questions. First, what is $|Y|$? You can use some number theory to show that $|Y| = \Theta(n)$, so this is fine. But the bigger problem is what happens to the other patterns $\Pi_3, \dots, \Pi_8$ — this was a nice aside showing that we could avoid Π_2 alone, but to prove Theorem 1.7, we need to deal with all the other patterns too.

The paper that mentioned this construction suggested that you might want to throw some hard number theory at this to handle the other patterns. But it turns out that you don't need to do this at all.

§4 Combining the two attempts

So far, we've talked about the relevant history up to Tao's paper. Tao's contribution is to say, why not do both — what if we take the parabola and randomize it in some way? So instead of trying to deal with $\Pi_3, \dots, \Pi_8$ using number theory, we introduce randomness into the parabola. More specifically, we take a *random affine transformation* of Y. And the hope is that maybe this gives enough randomness to get the same properties we had when choosing things *completely* randomly (which successfully avoided $\Pi_3, \dots, \Pi_8$).

The way this works is that we choose $a, b, c, d \in \mathbb{F}_p$ uniformly among all quadruples satisfying

$$\det \begin{bmatrix} a & c \\ b & d \end{bmatrix} \not\equiv 0 \pmod{p}$$

(essentially, we're choosing two linearly independent vectors in $\mathbb{F}_p^2$), and we also choose $e \in \mathbb{F}_p$ uniformly at random. Then we define

$$X = \{(x, y) \in [n]^2 \mid (ax + by)^2 \equiv cx + dy + e \pmod{p}\}.$$

This essentially corresponds to taking a random nondegenerate affine transformation of our parabola; the determinant condition ensures that the transformation is nondegenerate.

First, here's a fact we'll use several times.

Fact 4.1 — If $a, b, c, d \in \mathbb{F}_p$ are chosen uniformly at random (over *all* possibilities), then

$$\det \begin{bmatrix} a & c \\ b & d \end{bmatrix} \neq 0$$

with high probability (specifically, with probability $1 - O(1/p)$).

The reason this is useful is that it allows us to go back and forth between the two ways of choosing a, b, c, and d — we can assume they're genuinely uniform when that's convenient, and assume they're uniform conditioned on having nonzero determinant when that's convenient. (Explicitly, we could imagine generating the distribution with the determinant condition by first choosing a, b, c, and d completely at random, and then conditioning on them having nonzero determinant; and Fact 4.1 means that the conditioning doesn't have much effect.)

Lemma 4.2

We have $|X| = \Theta(n^2/p) = \Theta(n)$ with high probability.

Proof. Suppose we fix (x, y) and consider $\mathbb{P}[(x, y) \in X]$. Then we can look at our equation

$$(ax + by)^2 \equiv cx + dy + e \pmod{p} \tag{4.1}$$

and consider what happens when we plug in x and y; if we fix a, b, c, and d (as well as x and y), there's exactly one choice of e that makes this equation true. So we get

$$\mathbb{P}[(x, y) \in X] = \frac{1}{p},$$

just by considering the randomness over e.

Now suppose that we fix *two* distinct points (x, y) and (x', y') and consider

$$\mathbb{P}[(x, y) \in X \text{ and } (x', y') \in X].$$

We can again imagine fixing a and b; then for (4.1) to hold, we need the two dot products

$$(c, d, e) \cdot (x, y, 1) \quad \text{and} \quad (c, d, e) \cdot (x', y', 1)$$

to take on particular values, namely $(ax + by)^2$ and $(ax' + by')^2$. And the two vectors $(x, y, 1)$ and $(x', y', 1)$ are linearly independent, so these two dot products are independent (over the randomness in choosing c, d, and e). And each of the dot products hits the desired value with probability $1/p$, so

$$\mathbb{P}[(x, y) \in X \text{ and } (x', y') \in X] = \frac{1}{p} \cdot \frac{1}{p} = \frac{1}{p^2}.$$

This means the events $(x, y) \in X$ are pairwise independent; then Chebyshev implies that $|X| = \Theta(\mathbb{E}[|X|]) = \Theta(n)$ with high probability. □

Remark 4.3. In this proof, we're cheating a bit because c and d aren't actually uniform given a and b (due to the determinant condition); but as mentioned before, we can pretend that they are by using Fact 4.1 to transfer between the actual distribution of a, b, c, and d and the uniform one.

Lemma 4.4

The set X avoids Π_1 and Π_2.

Proof. It avoids Π_1 because the grid itself has no equilateral triangles. And it avoids Π_2 because it's an affine transformation of Y, so if X had a parallelogram then so would Y. □

Now it remains to account for the other patterns. For this, we'll need the following fact, which says that the probability *any* four points appear in X is comparable to what it would be if X were fully random.

Lemma 4.5

Given any four (distinct) points $p_1, \ldots, p_4 \in \mathbb{F}_p^2$, we have $\mathbb{P}[p_1, \ldots, p_4 \in X] = O(p^{-4})$.

Proof. First, the distribution of X is invariant under any fixed affine transformation — if we first perform a fixed affine transformation and then perform the random construction, then we get the same result as if we performed the random construction directly (since the affine transformation in the construction gets composed with this one). This means we can translate and scale so that

$$p_1 = (0,0),\ p_2 = (0,1),\ \text{and}\ p_3 = (1,0).$$

(We can't do this if the points are collinear; but if they are collinear then $\mathbb{P}[p_1, \ldots, p_4 \in X] = 0$, since you can't have three collinear points on a parabola. So we can ignore this case.)

Now let $p_4 = (s,t)$. Then to see what it means for these four points to be on our parabola, we can plug in their (x,y)-values into (4.1) and get a system of equations. When we plug in p_1, we get a bunch of 0's, and we're left with $0 \equiv e$. For p_2, we're left with $b^2 \equiv d$; and for p_3, we're left with $a^2 \equiv c$. Finally, p_4 gives

$$(as + bt)^2 \equiv cs + dt + e.$$

And we want to know, when are these equations simultaneously satisfied?

If we plug in the first three equations into the fourth and rearrange, we get

$$a^2(s^2 - s) + b^2(t^2 - t) + 2stab \equiv 0.$$

The values of s and t are fixed (they come from our particular point p_4), so this is a quadratic in a and b. And it can't vanish (that would require either s or t to be 0, in which case three of $p_1, \ldots, p_4$ would be collinear). So there's at most $2p$ solutions for (a,b).

And once we know what a and b are, the remaining variables c, d, and e are all determined (by the first three equations). So then

$$\mathbb{P}[p_1, \ldots, p_4 \in X] \leq \frac{2p}{p^5} = O(p^{-4}). \qquad \square$$

Now this means that for each $3 \leq i \leq 8$, we have

$$\mathbb{E}[\#\Pi_i] = O\left(\frac{n^5}{p^4}\right) = O(n)$$

(we had $O(n^5)$ copies of Π_i to start with, and each survives with probability p^{-4}). And this is exactly what we wanted — it's the same situation as we had with the truly random construction, where we have a set of size n and there's $O(n)$ copies of these patterns. And we can refine in the same way as before — we keep each point with probability ε to get a set that *almost* avoids all the patterns, and then we can kill all the patterns by deleting a point from each.

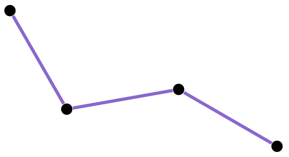
So we've used the same idea from the first attempt of taking a random subset. But instead of taking a *purely* random subset, we started with something more structured to avoid parallelograms; this made things a bit more complicated, but they still work out.

§5 Counting patterns in the grid

Finally, we'll talk a bit about the one gap we've left, which is Theorem 2.2 (counting the patterns in $[n]^2$).

§5.1 Counting paths of length 3

Suppose we want to count paths of length 4 (this is the pattern Π_4).



For this, we'll use the following number-theoretic fact.

Proposition 5.1
For any fixed $x \in [n]^2$ and distance δ, we have $\#\{y \mid d(x,y) = \delta\} = O_\varepsilon(n^\varepsilon)$ for every $\varepsilon > 0$.

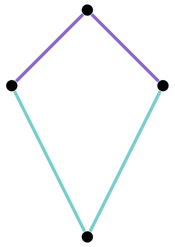
Then to count paths, we can call our four points x_1, x_2, x_3, x_4, so we want to count

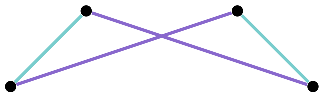
$$\#\{(x_1, x_2, x_3, x_4) \mid d(x_1, x_2) = d(x_2, x_3) = d(x_3, x_4)\}.$$

There's n^2 choices for each of x_1 and x_2 (we can choose them however we want); then that determines a distance, and at that point there's n^ε choices for x_3, and then n^ε choices for x_4. So we get $n^{4+2\varepsilon}$ paths.

§5.2 Counting kites

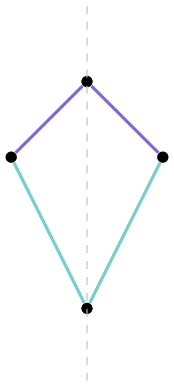
Most of the patterns follow behavior similar to this — you can look at some fixed thing and draw circles and use basic geometry, and you end up getting a bound like this. But we said $O(n^5)$, which is notably much bigger than this bound. That's because of a few exceptions — kites and isosceles trapezoids, where it turns out there actually are n^5.





For example, for kites, there's already n^5 axis-aligned kites.

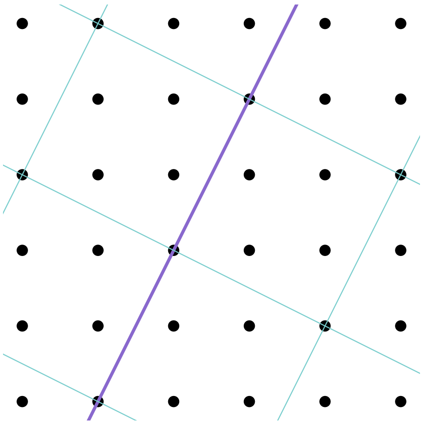
The way the authors deal with these patterns in the paper is that each of these has some axis of symmetry; and we take the line that's the axis of symmetry, and stratify based on the number of integer points on that line. (We'll focus on kites.)



Proposition 5.2

Given a line with at most j points, the number of kites with that axis of symmetry is $O(j^4)$.

Intuitively, the idea is that you can consider the grid rotated by this slope; lines in the perpendicular direction will also have roughly j points, so this will be a $j \times j$ grid.



And all points on the kite will need to be in this $j \times j$ grid. So there's j ways to choose each of the two points on the axis and j^2 ways to choose the third point, and this determines the fourth.

Once we have this, we need some bound on the number of lines with j points.

Proposition 5.3

The number of lines which intersect $[n]^2$ in at least j poins is $O(n^4/j^3)$.

One way to get this is from Szemerédi–Trotter. (Proving Szemerédi–Trotter for grids is also not that bad.)

Then we can combine these bounds. The naive calculation would be to sum over all $j = 1, 2, \ldots, n$, but this wouldn't work — it would end up giving you

$$\sum_{j=1}^{n} \frac{n^4}{j} \cdot j^4 = n^4 \sum_{j=1}^{n} j \asymp n^6,$$

which is bad. But this is wasteful because we've counted things a bunch of times. To be less wasteful, we can instead block our values of j into groups that scale exponentially — we consider blocks $j \in [2^i, 2^{i+1})$. After this stratification, we get

$$\sum_{i=1}^{\log n} \frac{n^4}{2^{3i}} \cdot 2^{4i} = n^4 \sum_{i=1}^{\log n} 2^i \asymp n^5,$$

which is what we wanted.