Heilbronn's triangle problem and projection theory

Talk by Alex Cohen and Dima Zakharov

Notes by Sanjana Das

October 13, 2023

This is joint work with Cosmin Pohoata.

§1 Heilbronn's triangle problem

Heilbronn's triangle problem is about finding small triangles in arbitrary sets of points — imagine that we start with a unit square, and we place a set of points $\mathcal{P}$ in it. Among all of these points, we consider the three of them which form the triangle with the smallest area; we let $\Delta(\mathcal{P})$ be the area of this triangle. And we define

$$\Delta(n) = \max_{|\mathcal{P}|=n} \Delta(\mathcal{P}).$$

In other words, $\Delta(n)$ is the smallest quantity such that no matter how you place down n points in the unit square, you can always find a triangle with area at most $\Delta(n)$.

Question 1.1 (Heilbronn)**.** What is the asymptotic decay rate of $\Delta(n)$?

The authors end up bringing in tools from fractal geometry and projection theory to study this problem.

§1.1 Lower bounds

We're mainly interested in proving *upper* bounds for $\Delta(n)$, where we want to find small triangles in arbitrary sets of points. Before we get to this, we'll talk briefly about *lower* bounds, where we want to construct sets of points such that all triangles have not-too-small area.

First, Erdős proved that $\Delta(n) \gtrsim 1/n^2$. His construction was to take

$$\mathcal{P} = \left\{ \left(\frac{j}{p}, \frac{j^2 \bmod p}{p} \right) \;\middle|\; 0 \le j \le p-1 \right\}$$

for each prime p. (No three points are collinear, and all their coordinates are integer multiples of $1/p$, so all their triangle areas are integer multiples of $1/2p^2$.)

In the 1980s, Komlos, Pintz, and Szemerédi improved this to

$$\Delta(n) \ge \frac{\log n}{n^2}.$$

They used a semi-random construction (which involved placing down points randomly, and showing that you can choose a good subset of these points).

§2 Basic observations

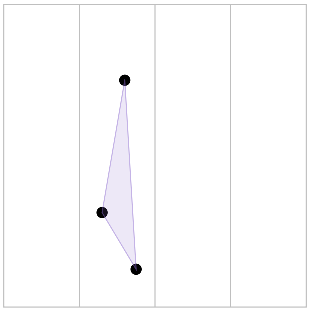
We'll first make some basic observations to situate ourselves.

§2.1 A simple upper bound

First, we have the following easy upper bound.

> **Fact 2.1** — We have $\Delta(n) \leq 5/n$.

Proof. Take the unit square and split it into strips of width roughly $5/n$. By pigeonhole, one of these strips must have five points in it; suppose we take three of those points and make a triangle.



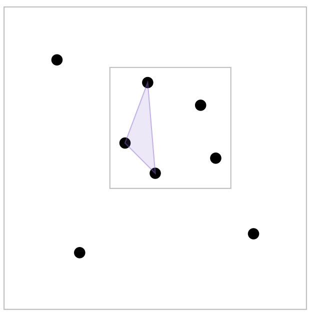
Then the area of this triangle is at most the area of the strip, which is $5/n$. □

So the first thing we might ask is whether $\Delta(n)$ decays asymptotically faster than this simple bound.

> **Question 2.2.** Is $\Delta(n) = o(1/n)$?

§2.2 Scaling

The second observation has to do with how the set of points should be distributed in the square. Intuitively, if you want to choose a set of points with no small triangles, then you'd want the points to be as spread out as possible, because if they're clumped up in any region, then you can probably find a smaller triangle in that region.



One way to make this precise is that if we have a higher density of points inside some smaller square, then we can run the same strip argument inside that small square to find a small triangle there. So if all you want to do is improve on the easy bound in Fact 2.1, then you can assume the points are roughly uniformly distributed in the square.

(And if you come up with a better way to find triangles than the strip argument, then you can run that on the smaller square as well.)

So this is the observation of *scaling* — if we have a high density of points in some region, then we can zoom in on that region and try to find a triangle there.

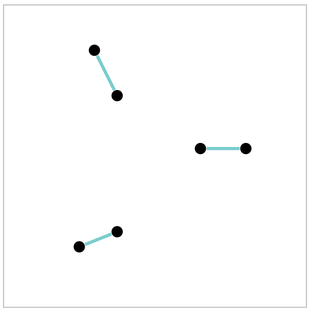
§2.3 An incidences problem

The third observation is a reformulation of the problem of finding small triangles in terms of *incidences*.

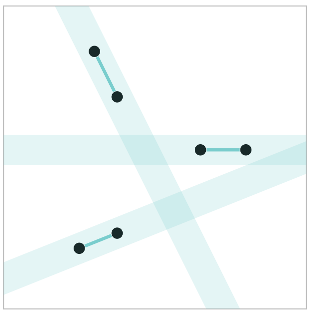
We're *not* going to look at the n^3 triangles formed by all triples of points and measure all of their areas — there's just too many of them. Inststead, we'll use the formula

$$\text{Area} = \frac{1}{2} \cdot \text{Base} \cdot \text{Height}.$$

We're first going to pick some pairs of points to be potential bases (such that each of these pairs has a short distance — specifically, distance at most u).



Then we look at the line spanned by such a pair; for each potential third point, the area of the corresponding triangle depends on its distance to that line. So we take all these pairs and draw little tubes with width w around the lines they span. If we can find another point in one of these tubes, then we get a triangle of area at most $\frac{1}{2}uw$.



All previous work on this problem has used this idea — we select pairs at distance at most u, form strips of width w around the lines they span, and try to find a third point in some strip.

§3 History and main results

§3.1 Previous upper bounds

Now we'll review the prior work on upper bounds for this problem.

- The trivial bound is $\Delta \lesssim n^{-1}$ (as in Fact 2.1).
- We *can* do asymptotically better — Roth (1951) proved that

  $$\Delta \lesssim n^{-1}(\log \log n)^{-1/2}.$$

  The proof is by a density increment argument, similar to Roth's work on arithmetic progressions. This bound decays faster than n^{-1}, but only slightly.
- Schmidt (1972) proved that

  $$\Delta \lesssim n^{-1}(\log n)^{-1/2}.$$

  He didn't just look at *one* value of u, but *many* values; the $\log n$ comes from synthesizing information on different scales.
- Roth (1972–73) then obtained a huge improvement — he proved a power-saving bound $\Delta \lesssim n^{-1.1}$.
- Komlos, Pintz, and Szemerédi (1981) then proved $\Delta \lesssim n^{-8/7}$. (This is kind of a natural bound that comes out of Roth's work.)
- The authors improved the exponent from 8/7 to something slightly larger (by a constant).

§3.2 Roth's setup

Now we'll describe the argument Roth used to prove a power-saving bound.

We start with a set of points $\mathcal{P} \subseteq [0,1]^2$. As mentioned before, we first create a set of lines connecting pairs of points that are close to each other — we define $\mathcal{L}$ as the set of lines spanned by points with $|x-y| = u$. (Roth used only one value of u.)

Definition 3.1. We define the number of incidences at scale w as

$$I(w) = \#\{(p,\ell) \in \mathcal{P} \times \mathcal{L} \mid d(p,\ell) \leq w\}.$$

In other words, we're taking all our lines, forming tubes around them of width w, and summing up (over all lines) the number of points in these tubes.

In this language, our goal is to show that $I(w_f) > 2\,|\mathcal{L}|$ for some small w_f. This is because if we can show this, then the *average* number of points in a tube is 2, so some tube has 3 points (including the two used to define the tube); that corresponds to a small triangle.

(Some typical values for the parameters are $u \sim n^{-1/3}$ and $w_f \sim n^{-3/4}$; if you run the argument with these choices, then you get $\Delta \lesssim n^{-13/12}$, which is a power-saving improvement.)

So finding small triangles corresponds to finding lower bounds for the number of incidences between tubes and points. And Roth invented a two-step method for proving lower bounds for incidences.

We pick two scales $w_i \sim n^{-0.1}$ and $w_f \sim n^{-3/4}$ (so the initial scale is much larger than the final scale). The first step is an *initial estimate* where we show that there's lots of incidences at the thick initial scale, i.e.,

$$I(w_i) \gtrsim w_i\,|\mathcal{P}|\,|\mathcal{L}|\,.$$

(Note that the right-hand side is the number of incidences we'd expect if the points were distributed randomly — we have $|\mathcal{L}|$ tubes, and each has area roughly w_i so is expected to contain roughly $w_i\,|\mathcal{P}|$ points.)

We can't use this alone to find any small triangles at all — using Area $= \frac{1}{2} \cdot \text{base} \cdot \text{height}$ at the initial scale would give us triangles of area much greater than $1/n$. The core of Roth's argument is the second step, an *inductive step* where we show that the normalized number of incidences doesn't change very much as the width changes — more precisely, that

$$\left| \frac{I(w_i)}{w_i \, |\mathcal{P}| \, |\mathcal{L}|} - \frac{I(w_f)}{w_f \, |\mathcal{P}| \, |\mathcal{L}|} \right| \ll 1.$$

The key ingredient for this is showing that the points and lines aren't too concentrated in smaller regions.

§3.3 Main result

Now we'll talk about why 8/7 (as proved by KPS) is a natural estimate. Roth's first paper had some inductive step, and all the following papers iteratively refined this step; and 8/7 is the best bound you can prove using the initial estimate from Roth's paper (i.e., the bound you get if you have the best possible inductive step).

The authors' improvement comes from keeping the inductive step the same, but improving the initial estimate a bit. Their main theorem is the following.

Theorem 3.2
We have $\Delta(n) \lesssim n^{-8/7-1/2000}$.

At a high level, they analyze discrete point sets using fractal geometry. There's a notion of *fractal dimension*, which measures how concentrated or spread out a set of points is; they do different things depending on how spread out the points are, using theorems from fractal geometry.

In this connection, Roth's inductive step is related to the high-low method in Fourier analysis, projection theory, and related fields. Guth–Solomon–Wang (2017) developed this to prove *upper* bounds for incidences (while Roth proved lower bounds).

This isn't where the improvement comes from; it's the same as what Roth did. The step where the improvement comes from is in finding new ways to prove initial estimates, using *direction set estimates* — this is different from what Roth does.

§4 The inductive step

When you first see this argument, the inductive step is the most surprising part — how in the world can you relate the number of incidences on different scales? So we'll start by discussing this step.

This step works in a fairly general setting, where $\mathcal{P}$ is an arbitrary set of points and $\mathcal{L}$ is an arbitrary set of lines (it doesn't have to be generated by pairs of points in $\mathcal{P}$). Recall that we defined the number of incidences at scale w as

$$I(w) = \#\{(p, \ell) \mid d(p, \ell) \le w\}.$$

The idea of the inductive step is the following observation: if $\mathcal{P}$ and $\mathcal{L}$ are not too concentrated on scale w, then the number of incidences should change smoothly at this scale. More precisely, imagine that we want to compare scale w with scale $10w$; we'd like to say that these two counts are similar to each other. We'd expect the number of incidences to be roughly proportional to $w \, |\mathcal{P}| \, |\mathcal{L}|$, so to compare the two scales we need to normalize by this — so we want to say that if $\mathcal{P}$ and $\mathcal{L}$ aren't too concentrated, then

$$\left| \frac{I(w)}{w \, |\mathcal{P}| \, |\mathcal{L}|} - \frac{I(10w)}{10w \, |\mathcal{P}| \, |\mathcal{L}|} \right|$$

is small. (The eventual goal is to relate scales w_i and w_f; we can do this by summing over dyadic scales.)

First, what does it mean that $\mathcal{P}$ and $\mathcal{L}$ aren't too concentrated? For $\mathcal{P}$, not being too concentrated means that no small square has too many points inside it. There's something similar for $\mathcal{L}$ — we don't want a bunch of lines to be concentrated in some rectangle. This condition fits quite nicely with our earlier observation about scaling — if this condition is *not* satisfied, then hopefully we can win by reducing to the subsquare where there are too many points of $\mathcal{P}$ (and we can do something similar if $\mathcal{L}$ is concentrated).

Now we'll try to explain how we can prove something about our scary expression. For convenience, we'll ignore the terms of $|\mathcal{P}|$ and $|\mathcal{L}|$; so our goal is to understand

$$\frac{I(w)}{w} - \frac{I(10w)}{10w}.$$

In $I(w)$, we get an incidence if a point p lies in the strip of width w around a line ℓ. So we define $T_\ell(w)$ as the indicator function of the strip of width w around ℓ; then we can formally write

$$\frac{I(w)}{w} - \frac{I(10w)}{10w} = \left\langle \mathbf{1}_{\mathcal{P}}, \sum_{\ell\in\mathcal{L}} \frac{T_\ell(w)}{w} - \frac{T_\ell(10w)}{10w} \right\rangle. \tag{4.1}$$

So the quantity we care about is a scalar product of some functions (corresponding to the relevant indicator functions). And when we see scalar products, we'd like to apply Cauchy–Schwarz to say something like

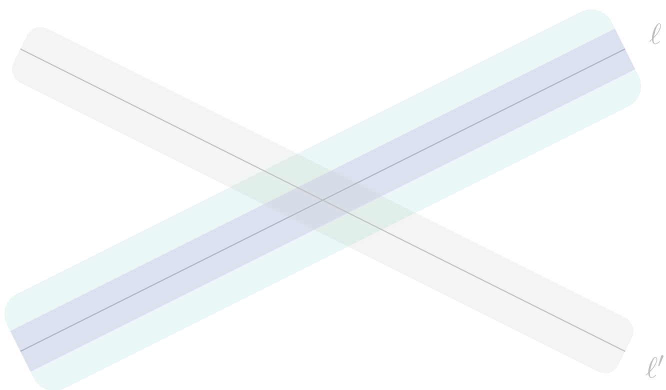
$$\left|\frac{I(w)}{w} - \frac{I(10w)}{10w}\right| \leq \|\mathbf{1}_{\mathcal{P}}\|_2 \cdot \left\|\sum_{\ell\in\mathcal{L}} \frac{T_\ell(w)}{w} - \frac{T_\ell(10w)}{10w}\right\|_2.$$

However, we need to be a bit careful — $\mathbf{1}_{\mathcal{P}}$ is a bunch of points, so its L^2 norm is infinite, which means we need to smooth $\mathbf{1}_{\mathcal{P}}$ a bit before doing Cauchy–Schwarz. For this, we imagine drawing a small ball around each of the points in $\mathcal{P}$; then instead of $\mathbf{1}_{\mathcal{P}}$, we have $\mathbf{1}_P * \eta$ (where η is some function corresponding to the small balls). And controlling $\|\mathbf{1}_P * \eta\|_2$ is easy — it reduces to looking at how many points you have in a square. So we can control the first term.

The second term looks scarier, but thankfully we can actually deal with it pretty easily — it turns out that the summands (in the sum over lines ℓ) are pairwise orthogonal. In other words, if we take two lines ℓ and ℓ' and draw strips of width w and $10w$ around them, then when we take the scalar product

$$\left\langle \frac{T_\ell(w)}{w} - \frac{T_\ell(10w)}{10w}, \frac{T_{\ell'}(w)}{w} - \frac{T_{\ell'}(10w)}{10w} \right\rangle$$

and expand things out, everything cancels and we get 0 (as long as the lines are not parallel). (In fact, the function on the left is orthogonal to $T_{\ell'}(w')$ for any w'.)



However, there's the issue that right now we're taking an integral of an infinite strip $T_\ell(w)$, and this integral is infinite. To fix this, we cut the integral off (by multiplying by some smooth bump around the origin). This causes the orthogonality to break by a little bit, but it doesn't break by too much if the angle between the lines isn't small (this is why we need the lines to not be too concentrated), and after some manipulations, we can get good control on

$$\left\| \sum_{\ell \in \mathcal{L}} \frac{T_\ell(w)}{w} - \frac{T_\ell(10w)}{10w} \right\|_2 .$$

If you do this kind of computation for arbitrary $\mathcal{P}$ and $\mathcal{L}$, then the eventual upper bound you get on the difference (between incidences on different scales) is sharp — you can construct examples showing that you can't improve this. This is a significant obstruction to improving Roth's bounds; and KPS pretty much pinned down the correct upper bound in this inductive step, so there's no room for improvement in this part of the argument.

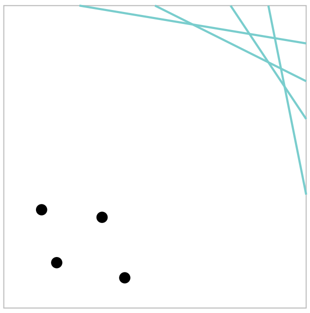
§5 The initial step

Now we'll talk about the other part of the proof — the inductive step allows us to compare different scales, so now all we need to do is to get a good initial exponent on some large scale.

In our situation, $\mathcal{P}$ and $\mathcal{L}$ are not arbitrary — $\mathcal{P}$ is our set of points, and

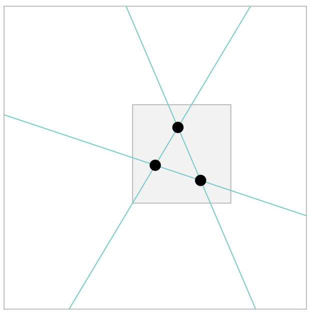
$$\mathcal{L} = \{\ell \mid \ell \text{ is spanned by } x, y \in \mathcal{P} \text{ with } d(x, y) \le u\}.$$

The fact that $\mathcal{P}$ and $\mathcal{L}$ are very important. Otherwise we could imagine a situation where all the points are in one corner, and all the lines are in the opposite corner; if the picture looked like this, then you'd have no chance of proving any lower bound.

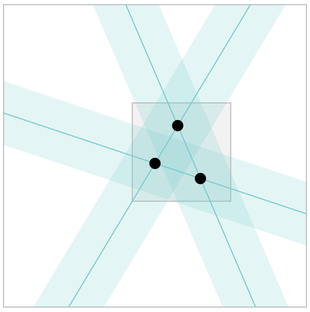


So we need to rule out a scenario like this; the point of the inductive step is that it allows us to work with a larger scale for the initial estimate, where it's easier to rule this out.

Let's suppose that $w_i = n^{-0.1}$. Our point set $\mathcal{P}$ lives somewhere inside the unit square; we'll select $u \times u$ squares which cover $\mathcal{P}$, and focus on one square $\mathcal{Q}$ at a time. The nice thing is that all pairs of points inside $\mathcal{Q}$ already have pairwise distance at most u (ignoring constant factors); so we know $\mathcal{L}$ contains all lines determined by this set of points, which we call $\mathcal{L}_{\mathcal{Q}}$.



Our goal is to find many incidences between $\mathcal{L}_Q$ and $\mathcal{P}$ at scale w_i. The way this picture looks is that we have a small square, and we draw wide tubes around each line originating from that square.



The key observation about this picture is that the directions of the lines in $\mathcal{L}_Q$ tell us how many incidences we should get — if the directions of the lines in $\mathcal{L}_Q$ are kind of uniformly spread out on the unit circle $\mathbb{S}^1$, then we have a bunch of lines coming out of our small square that look like they're all over the place; then picking a tube from this collection looks kind of the same as picking a uniform random tube in the plane, so we get the correct number of incidences with $\mathcal{P}$. (This is informal, but there's a simple lemma which says that this is indeed the case in some precise way.)

This leads us to try to understand the direction set of $\mathcal{L}_Q$ — we want to show that the directions of $\mathcal{L}_Q$ are indeed all over the place. And this is where ideas from projection theory come into play.

Let's look closer at $\mathcal{P} \cap \mathcal{Q}$ (the set of points inside our square $\mathcal{Q}$). This is some finite set of points, and we'll think of it as a fractal and look at its dimension. What does this mean? To define the dimension of a fractal set, you imagine covering your set with squares at a certain scale and counting how many squares you need to cover it; and you look at the asymptotics of this number as the scale varies. And you can do the same procedure for a discrete set.

And the idea is that if its dimension is large, i.e., $\dim(\mathcal{P} \cap \mathcal{Q}) > 1$ (which you should think of as saying that $\mathcal{P} \cap \mathcal{Q}$ doesn't lie on a curve), then its direction set $\operatorname{dir}(\mathcal{P} \cap \mathcal{Q}) \subseteq \mathbb{S}^1$ should have full dimension. To make this intuition precise, we use a classical theorem, which gives a precise statement of this form.

Theorem 5.1 (Marstrand 1954)
Let $X \subseteq \mathbb{R}^2$ have Hausdorff dimension $\dim(X) > 1$. Then $\dim(\operatorname{dir}(X)) = 1$.

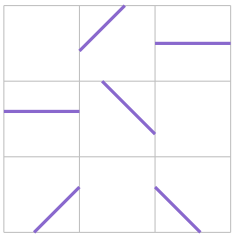
(Note that $\operatorname{dir}(X)$ is a subset of $\mathbb{S}^1$, so 1 is the maximum possible dimension it could have.)

Using a discrete version of this theorem gives an initial estimate when $\mathcal{P} \cap \mathcal{Q}$ is more than one-dimensional. Otherwise, we again use the scaling observation. So the initial estimate has two cases:

- If $\mathcal{P} \cap \mathcal{Q}$ is one-dimensional, then it is concentrated on small squares, and we can use the scaling argument to get small triangles.
- Otherwise, we can use this projection theory argument.

Combining this with the inductive step gets $\Delta(n) \lesssim n^{-8/7}$, recovering the exponent of KPS.

How do we improve this? The obstruction in this approach is the following situation: suppose we have our unit square and we split it into $u \times u$ squares, and $\mathcal{P}$ intersects most of these $u \times u$ squares, but inside each one it's kind of one-dimensional (i.e., it lives inside some curve or one-dimensional fractal). This is kind of the bottleneck of the approach, because then we can't get any estimates below scale u.



To fix this, we need a projection theory result for sets of dimension *less* than 1.

Theorem 5.2 (OSW 2022)
If $X \subseteq \mathbb{R}^2$, then as long as X is not contained in any line, we have

$$\dim(\operatorname{dir}(X)) \geq \min\{1, \dim(X)\}.$$

We always have $\dim(\operatorname{dir}(X)) \leq 1$ because $\operatorname{dir}(X)$ lives in $\mathbb{S}^1$; and this theorem says that $\dim(\operatorname{dir}(X))$ can be as small as the dimension of the original set, but it can't be smaller.

Then using a discrete version of this result, the authors can get initial estimates for smaller u in the above picture; and this eventually leads to an upper bound of $\Delta(n) \lesssim n^{-8/7-1/2000}$.

Remark 5.3. The limit of this kind of argument would be the bound $\Delta(n) \lesssim n^{-7/6}$, coming from a set which is 2-dimensional; then the obstacle would lie in the inductive step. Meanwhile, the limit of the incidence approach in general would be $\Delta(n) \lesssim n^{-3/2}$ (this comes from picking points at distances $n^{-1/2}$ from each other; then the best thing you can hope for is to find an extra point in a strip of width n^{-1}, which gives you $n^{-3/2}$). This is still very far from the lower bound of around n^{-2}.