INCIDENCE LOWER BOUNDS MINICOURSE

ALEX COHEN

Abstract. These are notes from a minicourse at the Simons School on Harmonic analysis and applications to Ramsey theory in Budapest, Hungary, June 2026. The main goal is to prove the following result from [3] (joint with Cosmin Pohoata and Dmitrii Zakharov): If $\{(p_j, \ell_j)\}_{j=1}^n$ is a family of points in the unit square along with a line through each point, then

$$\min_{j \neq k} d(p_j, \ell_k) \leq n^{-2/3+o(1)}.$$

Course Outline

Lecture 1 — Introduction.

- We pose the minimal distance problem and state our main result.
- We present the improved lower bounds of Logunov–Zakharov [14] and Hunter–Pohoata–Verstraëte–Zhang [8].
- We discuss related work of Dąbrowski–Goering–Orponen [4] on the s-Nikodym problem.
- We turn to Heilbronn's triangle problem, surveying lower bound constructions and applying our main result to prove an upper bound.

Lecture 2 — Affine rescaling, the Frostman condition, and a two-step incidence strategy.

- We recall Frostman sets and Hausdorff dimension.
- We apply affine rescaling to the minimal distance problem.
- We move from the minimal distance problem to the incidence lower bounds problem.
- We describe Roth's two-step strategy for finding incidences: first we find many incidences at a large initial scale, then we use the high–low method to show many remain at a small scale.

Lecture 3 — Initial estimate.

- We prove an initial incidence estimate under a *direction stability* condition.
- We describe a two-step strategy for proving incidence lower bounds: Find a rescaled configuration such that (1) direction stability can be used to find lots of big-scale incidences, and (2) the high-low method can be used to find lots of small-scale incidences.

Date: June 7, 2026.

Lecture 4 — Uniformity and branching functions.

- We introduce *uniform* sets and their *branching functions*, which describe the entire multiscale structure of a set by a Lipschitz function.
- We show every finite set has a large uniform subset, and extend uniformity and branching functions to the phase space of point-line pairs.
- We rephrase direction stability and high–low error using branching functions.
- We reduce the incidence lower bound theorem to a problem about Lipschitz functions.

Lecture 5 — Solving the branching function problem (bonus).

- We describe the three properties of branching functions we need: monotonicity, 1-Lipschitz, and *submodularity*.
- We design a measure on the space of candidate rectangles and show the direction-stable scales have density $1 - o(1)$ while the high-low-regular scales have positive density.
- We conclude that some rectangle satisfies both properties, letting us execute the two-step strategy and prove the incidence lower bound theorem.

Acknowledgements. Thanks to the organizers for making this lovely summer school happen, and to the attendees for their engagement and thoughtful questions. Thanks to Ben Goldman, the TA for this minicourse. Thanks to my coauthors, Cosmin Pohoata and Dmitrii Zakharov.

The author is supported by a Clay Research Fellowship. Large language models were used for light revision and for generating diagrams. All text was written by the author.

1. Lecture 1: Introduction

1.1. **Main question.** A *configuration* is a family $X = \{(p_j, \ell_j)\}_{j=1}^n$ of points in the unit square, along with a line ℓ_j through each point. The *minimal distance* of X is defined to be

$$\delta^*(X) = \min_{j \neq k} d(p_j, \ell_k),$$

see Figure 1. Let

$$\delta^*(n) = \sup\{\delta^*(X) : X \text{ is a configuration of } n \text{ point-line pairs}\}.$$

In other words, $\delta^*(n)$ is the smallest number such that in any point-line pair configuration of size n, the minimal distance is $\leq \delta^*(n)$.

Question 1.1 (Minimal distance problem)**.** *What is the asymptotic decay rate of $\delta^*(n)$ as $n \to \infty$?*

There is an easy upper bound $\delta^*(n) \lesssim n^{-1/2}$, because there is some pair of points at distance $\lesssim n^{-1/2}$.

There is an easy construction showing $\delta^*(n) \gtrsim n^{-1}$. Simply place n points on a horizontal line with even spacing, and put a vertical line through each point. It is easy to make lots of other constructions saturating this lower bound.

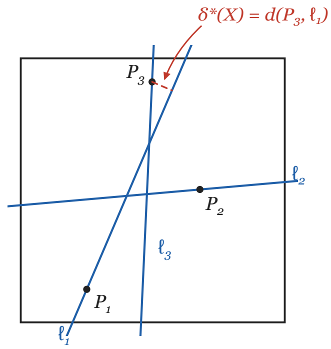

FIGURE 1.

These easy arguments show that up to constants, $\delta^*(n) \in [n^{-1}, n^{-1/2}]$. My coauthors and I wrote down Question 1.1 as a route to attack Heilbronn's triangle problem, which we will discuss later in this lecture. We quickly came up with these two easy arguments, but struggled to prove anything stronger in either direction, and we didn't find any prior literature.

We eventually proved a stronger upper bound for this question, which is the topic of these lectures.

Theorem 1.2 ([3]). *There is an upper bound $\delta^*(n) < n^{-2/3+o(1)}$. In other words, for any $\varepsilon > 0$, there is a constant C_ε such that among any configuration $\{(p_j, \ell_j)\}_{j=1}^n$, for some $j \neq k$, $d(p_j, \ell_k) \leq C_\varepsilon n^{-2/3+\varepsilon}$.*

Remark 1. We can state a finite field variant of the minimal distance problem. We say $d(p_j, \ell_k) = 0$ if $p_j \in \ell_k$, otherwise $d(p_j, \ell_k) = 1$. The analogue of Theorem 1.2 is true: If you place at least $q^{\frac{3}{2}+o(1)}$ pairs in $\mathbb{F}_q^2$, then $p_j \in \ell_k$ for some $j \neq k$. There is a simple proof using Vinh's [28] inequality. In [3] we show this estimate is sharp when $q = p^2$, p a prime.

The next several lectures focus on the proof of Theorem 1.2. In the rest of this lecture we will discuss connections to some other topics in geometric measure theory and combinatorics:

- Fractal sets and fractal dimension;
- The Furstenberg–Sárközy problem from additive combinatorics;
- The Nikodym set problem from geometric measure theory;
- Heilbronn's triangle problem from discrete geometry.

1.2. **Improved lower bounds.** In 2025, Logunov–Zakharov [14] showed $\delta^*(n) \geq cn^{-1+c}$ for some inexplicit $c > 0$. First they establish a submultiplicativity type estimate. Letting $n^*(\delta) = \max\{n : \delta^*(n) > \delta\}$, they show

$$n^*(\delta) \geq w^{-1} n^*(Cw) n^*(\delta/w^2) \qquad \text{for all } w \in [\delta^{1/2}, 1] \text{ and } C \text{ a universal constant.} \tag{1.1}$$

The easy bound implies $n^*(\delta) \gtrsim \delta^{-1}$, and they establish a logarithmic improvement: $n^*(\delta) \gtrsim \delta^{-1}\log\delta^{-1}$. It follows that

$$n^*(Cw) \geq w^{-1-c} \text{ for some } w \in (0, 1/3) \text{ and } c > 0. \tag{1.2}$$

Applying submultiplicativity many times implies

$$\begin{aligned} n^*(w^{2k}) &\geq w^{-1}n^*(Cw)n^*(w^{2k-2}) \\ &\geq w^{-j}n^*(Cw)^j n^*(w^{2k-2j}) \qquad \text{for all } j \in \{0, \ldots, k\} \\ &\geq w^{-k}n^*(Cw)^k \geq w^{-2k(1+c/2)}, \end{aligned}$$

thus $n^*(\delta) \geq \delta^{-1-c/2+o(1)}$, and $\delta^*(n) \leq n^{-1-c'}$.

To prove submultiplicativity, they describe a recipe that takes in two configurations X_1 and X_2, and builds a new configuration X_3 that resembles X_1 at small scales and resembles X_2 at large scales. The small scale pieces are made up of affine transformations of the original configuration. So their examples are self-affine fractal sets, see Figure 2.

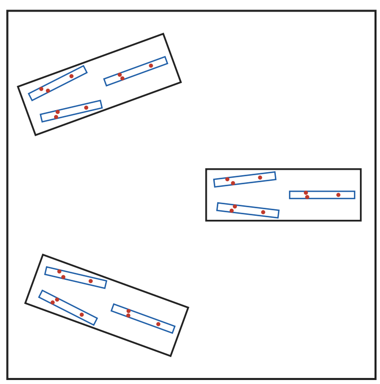

FIGURE 2. The Logunov–Zakharov [14] example is built from affine transformations; the resulting point sets are self-affine. See Section 2.3 for discussion of how the minimal distance problem behaves under affine transformations.

Hunter–Pohoata–Verstraëte–Zhang [8] further improved the lower bound with a completely different construction. The first step moves from the point-line problem to a point-parabola problem via the map

$$\Phi(x, y) = (x,\, y + x^2). \tag{1.3}$$

This map sends a line to a translate of the standard parabola $E_0 = \{(t, t^2) : t \in \mathbb{R}\}$,

$$\Phi\big(\{(t,\, b + mt) : t \in \mathbb{R}\}\big) = \Big\{\big(t,\ (t + \tfrac{m}{2})^2 + b - \tfrac{m^2}{4}\big) : t \in \mathbb{R}\Big\} = (-\frac{m}{2}, b - \frac{m^2}{4}) + E_0.$$

Thus a point-line pair (p, ℓ) maps to a point-parabola pair $(p, \Phi(\ell))$, with $\Phi(\ell)$ a translate of E_0 through p. The map Φ distorts distances in the unit square by at most constant factors, so up to constants, $\delta^*(n)$ describes the minimal distance of a point-parabola configuration.

The second step relates the point-parabola problem to the Furstenberg–Sárközy problem on square-difference-free sets, an idea going back to Szőnyi [27]. A set $A \subset \{0, \ldots, N\}$ is *square-difference-free* if for all $y \neq y' \in A$, the difference $y - y'$ is not a perfect square.

Lemma 1.3. *If $\{0, \ldots, N\}$ contains square-difference-free sets of size $\geq N^{\alpha - o(1)}$, then $\delta^*(n) \geq n^{-1/(\frac{1}{2}+\alpha)-o(1)}$.*

Proof. Given such a set A, define a point set and a family of parabolas

$$P = \Big(\tfrac{1}{\sqrt{N}}\mathbb{Z} \cap [0,1]\Big) \times \tfrac{1}{N}A, \qquad \mathbb{P} = \{\, p + E_0 : p \in P \,\}.$$

Each parabola $p + E_0$ passes through p, so $(P, \mathbb{P})$ is a configuration of $|A|\sqrt{N}$ point-parabola pairs. Consider two points $p_i = (\frac{x_i}{\sqrt{N}}, \frac{y_i}{N})$ with $x_i \in \{0, \ldots, \sqrt{N}\}$ and $y_i \in A$, for $i \in \{1, 2\}$. The vertical distance from p_1 to the parabola $p_2 + E_0$ is

$$\Big|\tfrac{y_1}{N} - \big(\tfrac{y_2}{N} + \tfrac{(x_1-x_2)^2}{N}\big)\Big| = \tfrac{1}{N}\big|\, y_1 - y_2 - (x_1 - x_2)^2 \,\big|.$$

If this vanishes then $y_1 - y_2 = (x_1 - x_2)^2$ is a perfect square, forcing $p_1 = p_2$ since A is square-difference-free; otherwise it is $\geq \frac{1}{N}$. The parabolas have slope $O(1)$ over the unit square, so Euclidean distance is comparable to vertical distance and $d(p_1, p_2 + E_0) \gtrsim \frac{1}{N}$. Hence $(P, \mathbb{P})$ is a configuration of $n = |A|\sqrt{N} \geq N^{1/2+\alpha-o(1)}$ pairs with minimal distance $\gtrsim \frac{1}{N}$, and so $\delta^*(n) \gtrsim \frac{1}{N} \geq n^{-1/(\frac{1}{2}+\alpha)-o(1)}$. □

It is easy to construct square-difference-free subsets of $\{0, \ldots, N\}$ of size $\sqrt{N}$, which recovers the easy bound $\delta^*(n) \gtrsim n^{-1}$. Ruzsa [24] constructed square-difference-free sets of size $\geq N^{0.7334}$, so by Lemma 1.3,

$$\delta^*(n) \geq n^{-0.811-o(1)}.$$

Conversely, improving the exponent in Theorem 1.2 from $\frac{2}{3}$ to $\frac{2}{3}+c$ would imply every square-difference-free subset of $\{0, \ldots, N\}$ has size $\leq N^{1-c'}$. This would resolve a long-standing question.

To summarize, the current knowledge is that

$$\delta^*(n) \in \big[\, n^{-0.811-o(1)},\ n^{-2/3+o(1)} \,\big].$$

1.3. **Related work on Nikodym sets.**

A *Nikodym set* $N \subset [0,1]^2$ is a Borel set such that for every $x \in N$, there is a line ℓ such that $\ell \cap N = \{x\}$; these are a continuous analogue of point-line pair configurations. Nikodym constructed such sets with complement having measure zero.

Dąbrowski, Goering, and Orponen define $N \subset \mathbb{R}^2$ to be an *s-Nikodym set* if for every $x \in N$ there exists an s-dimensional family of lines through x which meet N only at x. They proved the following result.

Theorem 1.4 ([4])**.** *Any s-Nikodým set $N \subset \mathbb{R}^2$ has Hausdorff dimension $\dim_H N \leq 2 - s$.*

There is work in progress by other authors showing this theorem is sharp.

Their work and ours were carried out independently and share similar insights.

1.4. **Heilbronn's triangle problem.** My collaborators and I became interested in Question 1.1 through Heilbronn's triangle problem. Given a finite set of points $P \subset [0,1]^2$, let $\Delta(P)$ be the smallest area of a triangle formed by 3 points of P (see Figure 3).

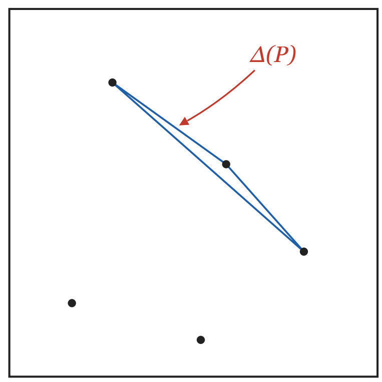

FIGURE 3.

Let

$$\Delta(n) = \sup\{\Delta(P) \,:\, P \subset [0,1]^2 \text{ has } n \text{ points}\}.$$

In other words, $\Delta(n)$ is the smallest number such that among any set of n points, there exists a triangle with area $\leq \Delta(n)$.

Question 1.5. *What is the asymptotic decay rate of $\Delta(n)$ as a function of n?*

Again, there are two easy bounds.

Easy upper bound. To prove an easy upper bound, consider splitting the unit square into $\lfloor \frac{n}{2} \rfloor - 1$ evenly spaced strips. By the pigeonhole principle, some strip has at least 3 points in it, which form a triangle with area $\leq \frac{1+o(1)}{n}$. Thus, $\Delta(n) \leq \frac{1+o(1)}{n}$.

In the other direction, one can construct a set of n points such that every triangle has area $\gtrsim \frac{1}{n^2}$.

Erdős's algebraic lower bound construction. Choose a prime p with $n \leq p \leq 2n$, and consider the p points

$$P_j = \left(\frac{j}{p}, \frac{j^2 \bmod p}{p}\right), \qquad j = 0, 1, \dots, p-1,$$

One can show no three of these points are collinear, and because they all lie on a $\frac{1}{p}$-grid, the minimal area of a triangle is at least $\frac{1}{2p^2} \geq \frac{1}{8n^2}$.

Random lower bound construction. Sample n points $p_1, \ldots, p_{2n}$ independently and uniformly from $[0,1]^2$. Let $\Delta > 0$ be a parameter; we want to count the expected number of triangles with area $\leq \Delta$. First, if p_1, p_2, p_3 are placed uniformly at random in $[0,1]^2$, we have

$$\Pr[\text{Area}(\Delta p_1 p_2 p_3) \leq \Delta] = \int_0^{\sqrt{2}} \Pr[\text{Area}(\Delta p_1 p_2 p_3) \leq \Delta \,|\, d(p_1, p_2) = r]\, p(d(p_1, p_2) = r)\, dr.$$

Fix p_1 and p_2. Using Area = Base × Height, the locus of p_3 such that $\text{Area}(\Delta p_1 p_2 p_3) \leq \Delta$ is a strip of width $\frac{\Delta}{d(p_1,p_2)}$ around the line through p_1, p_2. Thus

$$\Pr[\text{Area}(\Delta p_1 p_2 p_3) \leq \Delta \,|\, d(p_1, p_2) = r] \lesssim \frac{\Delta}{r}.$$

Next, we have

$$p(d(p_1, p_2) = r)\, dr \lesssim r dr,$$

so

$$\Pr[\text{Area}(\Delta p_1 p_2 p_3) \leq \Delta] \lesssim \int_0^2 \frac{\Delta}{r}\, r dr \lesssim \Delta.$$

Thus the expected value of the number of triangles with area $\leq \Delta$ is $\leq Cn^3\Delta$ for a universal constant C. Choose $\Delta = \frac{1}{10Cn^2}$, so the expected triangle count is $\leq \frac{n}{10}$. For each of these bad triangles, remove one vertex arbitrarily. The resulting set has $\geq \frac{n}{2}$ points, and all triangles from this set have area $\geq \frac{1}{10Cn^2}$ as needed.

Komlós, Pintz, and Szemerédi [12] improved this lower bound by a logarithmic factor to

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

Their argument is a semi-random construction. They first prove an extremal combinatorics result about independent sets in hypergraphs: under a mild assumption, a 3-uniform hypergraph on N vertices with E edges has an independent set of size logarithmically greater than the trivial bound. They make a random hypergraph by placing down $n^{1+\varepsilon}$ points at random and including all triangles with area $\leq c\frac{\log n}{n^2}$. They apply their hypergraph theorem to extract a subset of size $\gtrsim n$ with no small area triangles.

We survey the prior work on upper bounds for $\Delta(n)$.

- **1951, Roth** [19]**.** Roth proved the first $o(1/n)$ upper bound, $\Delta(n) \lesssim n^{-1}(\log\log n)^{-1/2}$, via a clever density-increment argument.
- **1971, Schmidt** [25]**.** Schmidt gave a much shorter proof of a slightly better bound, $\Delta(n) \lesssim n^{-1}(\log n)^{-1/2}$.
- **1972–1973, Roth** [20–22]**.** Roth introduced a new method to prove a polynomial upper bound $\Delta(n) \lesssim n^{-1-c}$ for some $c > 0$. He wrote two more papers sharpening his method and giving slightly better exponents. All subsequent work builds on the framework Roth introduced in these papers.
- **1981, Komlós–Pintz–Szemerédi** [11]**.** These authors sharpened a pigeonholing step in Roth's argument to prove $\Delta(n) \leq n^{-8/7+o(1)}$.

- **2023, C.–Pohoata–Zakharov** [2]**.** We reinterpreted Roth's framework in the modern context of fractal geometry and projection theory. We used a strong projection theorem of Orponen–Shmerkin–Wang [17] to improve the bound a little bit to $\Delta(n) \lesssim n^{-8/7-1/2000}$.
- **2025, C.–Pohoata–Zakharov** [3]**.** Our estimate on $\delta^*(n)$ from Theorem 1.2 immediately implies
$$\Delta(n) \lesssim_\varepsilon n^{-7/6+o(1)}.$$

Thus, the current state of knowledge is
$$\Delta(n) \in [(\log n)n^{-2}, n^{-7/6+o(1)}].$$

Our most recent bound on Heilbronn's problem follows from combining the following simple Lemma with Theorem 1.2.

Lemma 1.6. $\Delta(n) \lesssim n^{-1/2}\delta^*(\lfloor n/4 \rfloor)$.

Proof. Let $P \subset [0,1]^2$ be a set of n points. Greedily extract $\lfloor n/4 \rfloor$ disjoint pairs (p_j, p'_j) with $d(p_j, p'_j) \lesssim n^{-1/2}$. Let ℓ_j be the line through p_j, p'_j; the family $\{(p_j, \ell_j)\}_{j=1}^{\lfloor n/4 \rfloor}$ forms a point-line configuration. For some $j \neq k$, $d(p_j, \ell_k) \leq \delta^*(\lfloor n/4 \rfloor)$. The three points p_k, p'_k, p_j form a triangle with base $\lesssim n^{-1/2}$, height $\lesssim \delta^*(\lfloor n/4 \rfloor)$, and thus area $\lesssim n^{-1/2}\delta^*(\lfloor n/4 \rfloor)$. □

Remark 2. When we apply Lemma 1.6 to Heilbronn's problem, we locate triangles with a base of length $\lesssim n^{-1/2}$. By contrast, in the Erdős example or the random example, the triangles with area $n^{-2+o(1)}$ have a base of length ~ 1. So, it may be suboptimal to only look for triangles with a short base.

Remark 3. Heilbronn originally conjectured $\Delta(n) \lesssim n^{-2}$. Komlós–Pintz–Szemerédi's lower bound disproves this conjecture, but it is still possible $\Delta(n) \lesssim n^{-2+o(1)}$. I used to feel this was probably true (no good reason). Now I feel maybe it's false (also no good reason). I would encourage people to try and find examples showing $\Delta(n) \gtrsim n^{-2+c}$ for some $c > 0$.

Remark 4. Maldague–Zakharov–Wang proved a nontrivial upper bound for the minimal distance problem in $\mathbb{R}^3$, and used this to prove the first power-saving bound for Heilbronn's triangle problem in $\mathbb{R}^3$.

2. Lecture 2: Affine rescaling, the Frostman condition, and a two-step incidence strategy

2.1. Frostman sets.

Definition 2.1. Let $P \subset [0,1]^2$ be a finite set of points. We say P is (δ, α, C)-Frostman if
$$|P \cap Q| \leq C\,\mathrm{Width}(Q)^\alpha\,|P| \qquad \text{for all squares } Q \text{ with } \mathrm{Width}(Q) \in [\delta, 1]. \tag{2.1}$$
This is a discrete analogue of saying P has Hausdorff dimension $\geq \alpha$.

I like to think about the Frostman condition in terms of covering numbers. Let $P \subset [0,1]^2$, and let
$$|P|_w = \text{Minimum number of } w\text{-balls needed to cover } P.$$

Suppose P satisfies (2.1). Then

$$|P| \lesssim |P|_w \sup_{\mathrm{Width}(Q)=w} |P \cap Q| \leq C w^{\alpha} |P|_w |P|,$$

implying

$$|P|_w \gtrsim \frac{1}{C} w^{-\alpha} \qquad \text{for all } w \in [\delta, 1].$$

Moreover, if $P' \subset P$ has $|P'| \gtrsim |P|$, then the same lower bound holds. The Frostman condition says P is large when viewed at any scale.

Remark 5. A Borel measure μ is α-Frostman if $\mu(B(x,r)) \lesssim r^{\alpha}$. Frostman [5] introduced this definition to Fractal geometry, and proved that a Borel set X has positive α-dimensional Hausdorff measure if and only if X carries an α-Frostman measure. Katz and Tao codified the study of discrete, quantitative variants of Hausdorff dimension in [9].

2.2. **Point-line phase space.** We parametrize point-line pairs by 3 coordinates as follows:

$$\omega = (x, y, z) \in \mathbb{R}^3 \quad \longmapsto \quad p_\omega = (x, y), \quad \ell_\omega = (x, y) + \mathbb{R}(1, z).$$

We focus on $\omega \in [-1,1]^3$, corresponding to points in $[-1,1]^2$ and a line with slope in $[-1,1]$. We describe a point-line configuration by a set $\mathbf{X} \subset [-1,1]^3$. We let $P[\mathbf{X}]$ be the underlying set of points, $L[\mathbf{X}]$ the underlying set of lines.

2.3. **Rescaling and Frostman configurations.** An important observation about Heilbronn's triangle problem is that it behaves well under rescaling. If you have a configuration of n points and want to find a small area triangle, one strategy is to look for dense squares. If there is a square with lots of points in it, you can just look for a small triangle there, and ignore everything else.

The same strategy works for the minimal distance problem. If there is a square with lots of points, you can zoom in and just look for a small distance inside that square.

You can also zoom into tubes. If there is a w-tube with lots of lines inside of it, you can try to find a small distance inside that tube, by affine rescaling to the unit square.

More generally, if R is a rectangle, we may consider the point-line pairs where the point lies in the rectangle and the slope of the line is aligned with the slope of the rectangle. If we take an affine map from R to the unit square, then by the slope condition, distances get scaled by $\frac{1}{\mathrm{Width}(R)}$. See Figure 4. We will describe this rescaling in symbols and discuss its consequences.

Let $\mathbf{X} \subset [-1,1]^3$ be a configuration. Let R be an (approximate) rectangle in the unit square of the form

$$R = (x_0, y_0) + \{(x, y + z_0 x) : |x| \leq \mathrm{Length}(R), |y| \leq \mathrm{Width}(R)\} \qquad (x_0, y_0, z_0) \in [-1,1]^3.$$

The center of R is (x_0, y_0), and the slope is z_0. Let

$$\mathrm{Slope}(R) = z_0, \qquad \text{Slope-Set}(R) = \left[z_0 - \frac{\mathrm{Width}(R)}{\mathrm{Length}(R)}, z_0 + \frac{\mathrm{Width}(R)}{\mathrm{Length}(R)}\right],$$

this set captures slopes aligned with R. We may associate to R the phase-space set

$$\mathbf{R} = R \times \text{Slope-Set}(R) \subset [-1,1]^3.$$

If a point-line pair (p, ℓ) lands in $\mathbf{R}$, then $p \in R$, and ℓ has a slope aligned with R.

Let ψ^R be an affine map taking R to $[-1,1]^2$. This map stretches vertical distances by $\frac{1}{\mathrm{Width}(R)}$, so

$$d_{\mathrm{vertical}}(\psi^R(p), \psi^R(\ell)) = \frac{1}{\mathrm{Width}(R)} d_{\mathrm{vertical}}(p, \ell), \qquad d_{\mathrm{vertical}} \text{ denotes the vertical distance.}$$

If $\mathrm{Slope}(\ell) \in \text{Slope-Set}(R)$, then both ℓ and $\psi^R(\ell)$ have slope in $[-1,1]$. Thus, the vertical distance is approximately equal to the Euclidean distance, and

$$d(\psi^R(p), \psi^R(\ell)) \sim \frac{1}{\mathrm{Width}(R)} d(p, \ell). \tag{2.2}$$

This approximate equality shows that minimal distances behave well under affine rescaling.

2.4. **Induction on scales strategy.** Our goal is to prove

$$\delta^*(n) \leq n^{-\gamma + o(1)} \qquad \text{for } \gamma = \frac{2}{3}. \tag{2.3}$$

Eq. (2.2) suggests an induction on scales strategy. Let R be a rectangle in the unit square, and let $\mathbf{R} = R \times \text{Slope-Set}(R)$. If $\mathbf{X} \cap \mathbf{R}$ is large, we can look for a small distance pair inside $\mathbf{X} \cap \mathbf{R}$ and ignore everything else.

Let $\psi^{\mathbf{R}}$ be the affine map taking $\mathbf{R}$ to $[-1,1]^3$. This map sends

$$(p, \ell) \mapsto (\psi^R(p), \psi^R(\ell)).$$

Define the rescaled configuration

$$\mathbf{X}^{\mathbf{R}} = \psi^{\mathbf{R}}(\mathbf{X}).$$

By (2.2),

$$\delta^*(\mathbf{X} \cap \mathbf{R}) \sim \mathrm{Width}(R)\, \delta^*(\mathbf{X}^{\mathbf{R}}),$$

see Figure 4. If we are lucky we can prove the goal bound (2.3), $\delta^*(\mathbf{X}^{\mathbf{R}}) \leq |\mathbf{X}^{\mathbf{R}}|^{-\gamma + o(1)}$. This gives the desired estimate for $\mathbf{X}$ if

$$\mathrm{Width}(R)\, |\mathbf{X} \cap \mathbf{R}|^{-\gamma} \leq |\mathbf{X}|^{-\gamma},$$

or in other words, if

$$|\mathbf{X} \cap \mathbf{R}| \geq \mathrm{Width}(R)^{1/\gamma} |\mathbf{X}|. \tag{2.4}$$

So, either we get mileage out of rescaling, or $|\mathbf{X} \cap \mathbf{R}| \leq \mathrm{Width}(R)^{1/\gamma} |\mathbf{X}|$ for all the rectangles containing at least two point-line pairs. Motivated by this calculation, we define Frostman configurations as follows.

Definition 2.2. We say $\mathbf{X}$ is (δ, α, C)-Frostman if for all rectangles R with $\mathrm{Width}(R) \in [\delta, 1]$,

$$|\mathbf{X} \cap \mathbf{R}| \leq C\ \mathrm{Width}(R)^{\alpha}\, |\mathbf{X}|. \tag{2.5}$$

If Q is a square, then $\omega \in \mathbf{Q}$ if and only if $p_\omega \in Q$. So (2.5) says

$$|P[\mathbf{X}] \cap Q| \leq C\ \mathrm{Width}(Q)^{\alpha}\, |P| \qquad \text{for all squares } Q \text{ with } \mathrm{Width}(Q) \in [\delta, 1]. \tag{2.6}$$

This exactly says that $P[\mathbf{X}]$ is (δ, α, C)-Frostman.

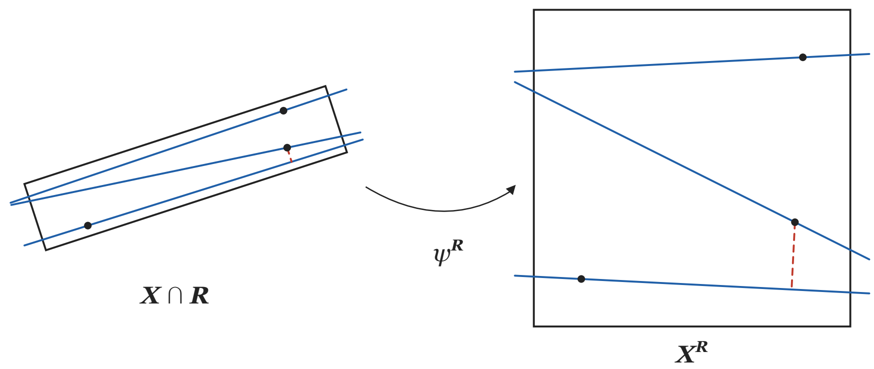

FIGURE 4. The minimal distance in $\mathbf{X} \cap \mathbf{R}$ gets scaled up by $\frac{1}{\mathrm{Width}(R)}$ after affine rescaling.

Similarly, when $R = T$ is a tube, $\omega \in \mathbf{R}$ means $\ell_\omega \subset T$. By $\ell_\omega \subset T$, we roughly mean $\ell_\omega \cap [-1, 2]^2 \subset T$. So

$$|L \cap T| \leq C \ \mathrm{Width}(T)^\alpha |L| \qquad \text{for all tubes } T \text{ with } \mathrm{Width}(T) \in [\delta, 1],$$

which says that the line set is (δ, α, C)-Frostman.

Thus, if $\mathbf{X}$ is (δ, α, C)-Frostman, then the underlying point set and the underlying line set are both α-Frostman.

2.5. **Incidence lower bound theorem.** Let $P \subset [0, 1]^2$ be a set of points. Let L be a family of lines. Define the *scale-w incidence count*

$$I(w) = I(w; P, L) = \sum_{(p,\ell) \in P \times L} \eta(d(p, \ell)/w), \qquad \eta \text{ is a bump function.} \tag{2.7}$$

For technical reasons it is convenient to take η smooth, but for these lectures, one can imagine $\eta = \mathbf{1}_{[-1,1]}$, in which case

$$I(w) = |\{(p, \ell) \in P \times L : d(p, \ell) \leq w\}|.$$

The following Theorem establishes an incidence lower bound for Frostman configurations.

Theorem 2.3. *For any $\alpha > \frac{3}{2}$, if $\mathbf{X} \subset [-1, 1]^3$ is $(\delta, \alpha, \delta^{-o(1)})$-Frostman, then*

$$I(\delta; P[\mathbf{X}], L[\mathbf{X}]) \gtrsim \delta^{1+o(1)} \, |\mathbf{X}|^2.$$

The $o(1)$ terms depend on α.

Since $|\mathbf{X}| \gtrsim \delta^{-3/2+o(1)}$, if δ is sufficiently small then $\delta^*(\mathbf{X}) \leq \delta$.

Remark 6. If P and L are placed uniformly at random, the expected number of incidences is $\sim \delta |\mathbf{X}|^2$.

Remark 7. Theorem 1.2, that $\delta^*(n) \leq n^{-2/3+o(1)}$, has an equivalent formulation: For any $\alpha > 3/2$, assuming δ is sufficiently small,

$$|\mathbf{X}| \geq \delta^{-\alpha} \implies I(\delta; P[\mathbf{X}], L[\mathbf{X}]) > |\mathbf{X}|.$$

In Theorem 2.3, the Frostman hypothesis strengthens the $|\mathbf{X}| \geq \delta^{-\alpha}$ hypothesis, and the conclusion strengthens from one extra incidence to many.

Proof of Theorem 1.2 using Theorem 2.3. Fix $\frac{3}{2} < \alpha' < \alpha$. Let $\mathbf{X} \subset [-1,1]^3$ be a configuration with $|\mathbf{X}| \geq \delta^{-\alpha}$. Assuming δ is sufficiently small, the goal is to show $\delta^*(\mathbf{X}) \lesssim \delta$.

Let $\mathbf{R} = R \times \text{Slope-Set}(R)$ be a rectangle with $\text{Width}(\mathbf{R}) \in [\delta, 1]$ maximizing the quantity

$$\text{Width}(\mathbf{R})^{-\alpha'}\, |\mathbf{X} \cap \mathbf{R}|. \tag{2.8}$$

Comparing with $R = [-1,1]^2$, we find

$$|\mathbf{X} \cap \mathbf{R}| \geq \text{Width}(R)^{\alpha'}\, |\mathbf{X}| \geq \delta^{\alpha' - \alpha}. \tag{2.9}$$

Define

$$\mathbf{X}^{\mathbf{R}} = \psi^{\mathbf{R}}(\mathbf{X} \cap \mathbf{R}), \qquad \tilde{\delta} = \frac{\delta}{\text{Width}(R)}.$$

By (2.2),

$$\delta^*(\mathbf{X}) \lesssim \text{Width}(\mathbf{R})\; \delta^*(\mathbf{X}^{\mathbf{R}}).$$

By (2.9), there is an easy bound $\delta^*(\mathbf{X}^{\mathbf{R}}) \lesssim |\mathbf{X} \cap \mathbf{R}|^{-1/2} \lesssim \delta^{\frac{\alpha - \alpha'}{2}}$, so

$$\delta^*(\mathbf{X}) \lesssim \text{Width}(\mathbf{R})\, \delta^{\frac{\alpha - \alpha'}{2}}.$$

If $\text{Width}(\mathbf{R}) \leq \delta^{1 - \frac{\alpha - \alpha'}{2}}$ we are done, so we may assume $\tilde{\delta} \leq \delta^{\frac{\alpha - \alpha'}{2}}$.

I claim $\mathbf{X}^{\mathbf{R}}$ is Frostman. If $\widetilde{\mathbf{R}}$ is a rectangle, then

$$|\mathbf{X}^{\mathbf{R}} \cap \widetilde{\mathbf{R}}| = |\mathbf{X} \cap (\psi^{\mathbf{R}})^{-1}(\widetilde{\mathbf{R}})|.$$

By the slope ~ 1 condition on $\mathbf{R}$ and our definition of Width,

$$\text{Width}((\psi^{\mathbf{R}})^{-1}(\widetilde{\mathbf{R}})) = \text{Width}(\mathbf{R})\ \text{Width}(\widetilde{\mathbf{R}}).$$

By maximality of (2.8),

$$|\mathbf{X}^{\mathbf{R}} \cap \widetilde{\mathbf{R}}| \lesssim \text{Width}(\widetilde{\mathbf{R}})^{\alpha'} |\mathbf{X}^{\mathbf{R}}|$$

as needed. By Theorem 2.3, using that $\tilde{\delta}$ is sufficiently small, we find $\delta^*(\mathbf{X}^{\mathbf{R}}) \leq \tilde{\delta}$, so

$$\delta^*(\mathbf{X}) \lesssim \text{Width}(\mathbf{R})\, \tilde{\delta} \lesssim \delta.$$

□

2.6. **High–Low method.** Let P and L be sets of points and lines in the unit square. Let $I(w) = I(w; P, L)$ be the scale-w incidence count. If P and L are placed uniformly at random, the expected incidence count is $\sim w\,|P|\,|L|$.

Theorem 2.4 (High–low inequality)**.** *Letting*

$$\mathrm{Err}(w; P, L) = \left(\frac{\sup_{Q \text{ is a } w\times w \text{ square}} |P\cap Q|}{|P|} \frac{\sup_{T \text{ is a } w\times 1 \text{ tube}} |L\cap T|}{|L|} w^{-3}\right)^{1/2}, \tag{2.10}$$

we have

$$\left|\frac{I(w)}{w\,|P|\,|L|} - \frac{I(w/2)}{(w/2)\,|P|\,|L|}\right| \lesssim \mathrm{Err}(w; P, L). \tag{2.11}$$

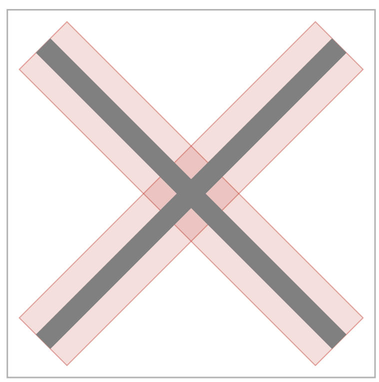

FIGURE 5. The two functions Φ_ℓ and $\Phi_{\ell'}$ are orthogonal. The gray part is negative and the red part is positive. The integral of Φ_{ℓ_1} over a transverse line is zero, because the positive and negative parts have magnitude inversely proportional to their length. So, the inner product is zero. In general the inner product is only approximately orthogonal because of boundary effects.

Proof Sketch. Encode the lines and points as

$$f_w = \sum_{\ell\in L} \tfrac{1}{w}\,\mathbf{1}_{\mathbb{T}_w(\ell)}, \qquad \mathbb{T}_w(\ell) \text{ is the } w\text{-tube around } \ell$$

$$g_{w/100} = \sum_{p\in P} \frac{1}{\mathrm{Vol}(B(w/100))} 1_{B(p,w/100)}.$$

With this definition,

$$\left|\frac{I(w)}{w\,|P|\,|L|} - \frac{I(w/2)}{(w/2)\,|P|\,|L|}\right| \sim \frac{1}{|P|\,|L|}|\langle f_w - f_{w/2}, g_{w/100}\rangle|. \tag{2.12}$$

The fact that we use $w/100$ balls to define $g_{w/100}$ rather than δ-functions means the right hand side counts a smoothed incidence count, rather than an on-the-nose incidence count. Using Cauchy–Schwarz,

$$|\langle f_w - f_{w/2}, g_{w/100}\rangle| \lesssim \|f_w - f_{w/2}\|_2 \, \|g_{w/100}\|_2. \tag{2.13}$$

To estimate $\|f_w - f_{w/2}\|_2$, we use approximate orthogonality. We have

$$f_w - f_{w/2} = \sum_{\ell \in L} \Phi_\ell, \qquad \Phi_\ell = \tfrac{1}{w}\,\mathbf{1}_{\mathbb{T}_w(\ell)} - \tfrac{1}{w/2}\,\mathbf{1}_{\mathbb{T}_{w/2}(\ell)}.$$

Morally speaking, if ℓ, ℓ' are two lines, then either ℓ, ℓ' live in a common w-tube, or $\langle \Phi_\ell, \Phi_{\ell'}\rangle$ is very small—see Figure 5. Again this is a white lie. One has to replace hard cutoffs with smooth cutoffs, and estimate the inner product more carefully.

Nevertheless, assuming the white lie,

$$\begin{aligned}\|f_w - f_{w/2}\|_2^2 &\lesssim \sum_{\ell,\ell' \in L} \|\Phi_\ell\|_2 \|\Phi_{\ell'}\|_2 \mathbf{1}_{\ell,\ell' \text{ belong to the same } w\text{-tube}} \\ &\lesssim |L|\, w^{-1} \sup_{T \text{ a } w \times 1\text{-tube}} |L \cap T|. \end{aligned} \tag{2.14}$$

Similarly,

$$\begin{aligned}\|g_{w/100}\|_2^2 &\lesssim \sum_{p,p' \in P} \left\| \frac{1}{\mathrm{Vol}(B(w/100))} 1_{B(p,w/100)} \right\|_2^2 1_{p,p' \text{ belong to the same } w/50\text{-ball}} \\ &\lesssim |P|\, w^{-2} \sup_{Q \text{ is a } w \times w \text{ square}} |P \cap Q|. \end{aligned} \tag{2.15}$$

Combining (2.12), (2.14), and (2.15) gives the result. □

Remark 8. If P is α-Frostman and L is β-Frostman, then $|P \cap Q| \lesssim \mathrm{Width}(Q)^\alpha |P|$ and $|T \cap L| \lesssim \mathrm{Width}(T)^\beta |L|$, so

$$\mathrm{Err}(w) \lesssim w^{\frac{\alpha+\beta-3}{2}}.$$

Thus, if the sum of the Frostman dimensions of P and L is > 3, then the incidence count does not change much as w varies. In particular, if $\mathbf{X}$ is α-Frostman with $\alpha > \frac{3}{2}$ and $P = P[\mathbf{X}]$, $L = L[\mathbf{X}]$, then the incidence count does not change much as w varies.

Remark 9. I first learned about the high–low method in Larry Guth's reading group, in the context of a projection theory paper by Guth–Solomon–Wang [7]. They used Fourier analysis to prove a bound like

$$B(w) \lesssim B(w/2) + (\text{High–Low error}).$$

Shortly after, while reading Roth's survey article [23], I learned that he used a very similar argument in the reverse direction,

$$B(w) \gtrsim B(w/2) - (\text{High–Low error}),$$

to prove a power-saving bound for Heilbronn's problem. As discussed in Larry Guth's notes on projection theory [6], a closely related orthogonality argument was used by Linnik [13] to prove his large sieve inequality in number theory, and by Vinh [28] to prove an incidence estimate over finite fields. Roth wrote several papers on the large sieve around the same time as his Heilbronn papers—maybe he got the idea from Linnik. I wonder where else has this orthogonality argument appeared?

Example 2.5. Let P be a family of well-spaced points in the upper half of the unit square and L a family of well-spaced lines in the lower half of the unit square. There are no incidences between P and L, even though the high–low error is small. We rule out this example in Theorem 2.3 using the fact that P and L come in incident pairs.

Example 2.6. Let P be an evenly-spaced grid of N points,

$$P = \Big\{ \big(\frac{a}{\sqrt{N}}, \frac{b}{\sqrt{N}}\big) \,:\, a, b \in \{0, \dots, N^{1/2}\} \Big\}.$$

We consider the following line set. Let

$$\Theta = \{ \frac{(c,d)}{|(c,d)|} \,:\, c, d \in \{N^{1/6}, \dots, 2N^{1/6}\} \}.$$

This is a set of $\sim N^{1/3}$ many directions. Consider the orthogonal projection

$$\pi_\theta(P) = \{ \langle p, \theta \rangle \,:\, p \in P \}, \qquad \theta \in \Theta.$$

This set consists of fractions of the form $\frac{m}{R}$, where $R \sim N^{1/2+1/6} = N^{2/3}$, and m is an integer $\lesssim R$. So, $\pi_\theta(P)$ is an arithmetic progression in $[-1, 1]$ with spacing $\sim N^{-2/3}$.

For each $\theta \in \Theta$, define the line set

$$L_\theta = \{ \pi_\theta^{-1}(x) \,:\, x \text{ lies in the middle of two points in } \pi_\theta(P) \}.$$

This is a set of $\sim N^{2/3}$ lines, each of which has distance $\sim N^{-2/3}$ from P. Define $L = \bigsqcup_\theta L_\theta$. In total, this is a family of $\sim N$ lines, all have distance $\sim N^{-2/3}$ from P.

This example is a modification of the Szemeredi–Trotter example from incidence geometry. In the Szemeredi–Trotter example, we choose the lines to be $\pi_\theta^{-1}(\pi_\theta(P))$, so they go through $\sim N^{1/3}$ points, rather than avoiding the points.

The incidence count of P, L is equal to

$$I(w; P, L) \sim \begin{cases} wN^2 & w > N^{-2/3}, \\ 0 & w < N^{-2/3}. \end{cases}$$

To compute the high–low error of P and L, we compute

$$\sup_{Q \text{ is } w \times w} |P \cap Q| \sim \max\{w^2 N, 1\},$$

$$\sup_{T \text{ is } w \times 1} |L \cap T| \sim \max\{w^2 N, wN^{2/3}, 1\}.$$

If $w > N^{-2/3}$, then both terms above are $\ll w^{\frac{3}{2}} N$, so $\mathrm{Err}(w) \ll 1$. If $w < N^{-2/3}$, then both terms above are 1, so $\mathrm{Err}(w; P, L) \sim N^{-1} w^{-3/2}$ is > 1.

There are three important ranges of w. When $w > N^{-1/2}$, the point set is doing all the work. The points are well-spaced, so any w-tube contains $\sim w|P|$ points. So for *any* line set L,

$$I(w; P, L) \sim w|P|\,|L| \qquad \text{for } w \in [N^{-1/2}, 1].$$

When $w \in [N^{-2/3}, N^{-1/2}]$, the high–low inequality is doing work. The high–low error ($\ll 1$) is much less than the normalized incidence count (~ 1), so the normalized incidence count stays at ~ 1. For any reasonably well spaced line set, the high low error is small enough that $I(w; P, L) \sim wN^2$ in this range.

At $w = N^{-2/3}$, there is a phase change: The high low error hits 1, and the incidence count drops to zero. See Figure 6 for a log-log plot of this discussion.

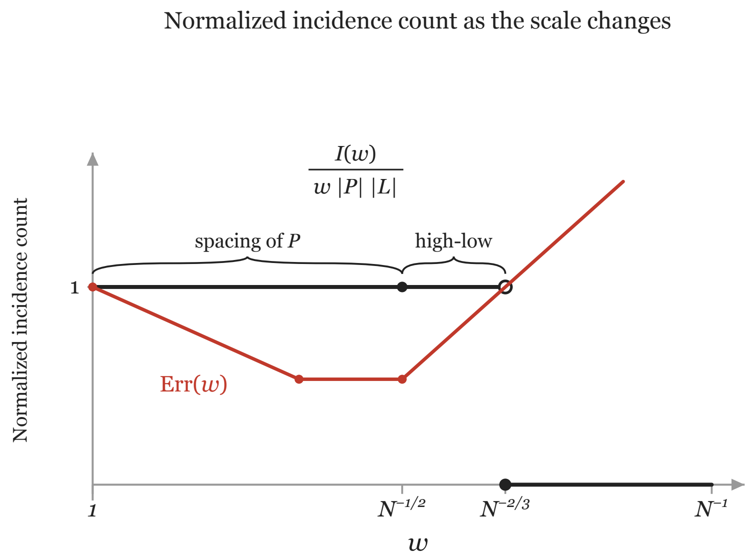

FIGURE 6. The black line records the normalized incidence count betwen P and L, and the red line records the high-low error. The incidence count starts out large because P is well spaced. It stays high between scales $N^{-1/2}$ and $N^{-2/3}$ because of the high-low inequality. At scale $N^{-2/3}$ the high-low error touches the black line, and the incidence count drops to zero.

This example shows that the high–low method is rather sharp. The incidence count hits 0 the moment the high–low error hits 1. This example explains the $n^{-2/3}$ exponent in Theorem 1.2.

3. Lecture 3 — Initial estimate

The high–low method is the gas that makes Theorem 2.3 run. The initial estimate provides the spark.

Lemma 3.1 (Initial Estimate)**.** *Let P and L be a set of points and lines in the unit square. Let $w > 0$ be a scale. Let*

$A =$ *The w-covering number of the slopes of lines passing through a typical w-square,*

$B =$ *The w-covering number of the slopes of all the lines.*

Then

$$I(w; P, L) \gtrsim \frac{A}{B}\, w\, |P|\, |L|.$$

See Figure 7.

Proof. Let $\mathbb{T}_w$ be an essentially distinct family of w-tubes covering all the lines. We assume by pigeonholing that every w-tube has roughly the same number of lines, and every w-square has roughly the same number of w-tubes through it.

Let C be the typical number of lines in a w-tube. Each point has $\sim A$ many w-tubes through it, and each w-tube has $\sim C$ lines inside of it, so each point has $\sim AC$ lines at distance $\leq w$. Thus

$$I(w; P, L) \gtrsim |P|\, AC.$$

On the other hand, there are at most w^{-1} tubes in each direction and at most B directions, so $|\mathbb{T}_w| \lesssim Bw^{-1}$, and $|L| \lesssim Bw^{-1}C$. Solving for C gives

$$I(w; P, L) \gtrsim \frac{A}{B}\, w\, |P|\, |L|$$

as desired. □

Let $\mathbf{X}$ be a point-line configuration as in Theorem 2.3. We can pick some large scale w and compute A and B for this configuration. We're happy if $\frac{A}{B} \gtrsim 1$. For example, this happens in Example 2.6. But we cannot verify this in practice. It might happen that $A \ll B$.

There are a few extreme examples. Both A and B live in the range $[1, w^{-1}]$. If $A = w^{-1}$ or if $B = 1$, then automatically $\frac{A}{B} \sim 1$.

The worst case scenario is when $A = 1$. In this case, let Q be a w-square, and consider the rescaled set

$$\mathbf{X^Q} = \text{Affine rescaling of the point-line pairs in } Q.$$

Let A_Q, B_Q be the corresponding quantities. Each line slope in this rescaled configuration corresponds to a line slope through Q in the original configuration. So $B_Q \sim 1$, and thus $\frac{A_Q}{B_Q} \sim 1$.

This is the key idea. Instead of trying to prove an initial estimate for our original point-line configuration, we will apply rescaling, and prove an initial estimate for the new configuration. The next Lemma captures this idea, see Figure 7.

Lemma 3.2. *Let $\mathbf{X}$ be a point-line configuration. Let $w > 0$ be a scale and let $k \geq 1$ be an integer. For some $j \in \{0, \ldots, k-1\}$ and some w^j-square Q, the rescaled configuration $\mathbf{X^Q}$ satisfies*

$$I(w; P[\mathbf{X^Q}], L[\mathbf{X^Q}]) \gtrsim w^{1/k} \left(w\, |P[\mathbf{X^Q}]|\, |L[\mathbf{X^Q}]| \right).$$

In practice we choose k rather large, so $w^{1/k} = w^{o(1)}$, and this is a good initial estimate.

Proof. Let

$$R_j = \text{The } w\text{-covering number of the set of slopes in } \mathbf{X^Q}, \text{ where } Q \text{ is a } w^j\text{-square.}$$

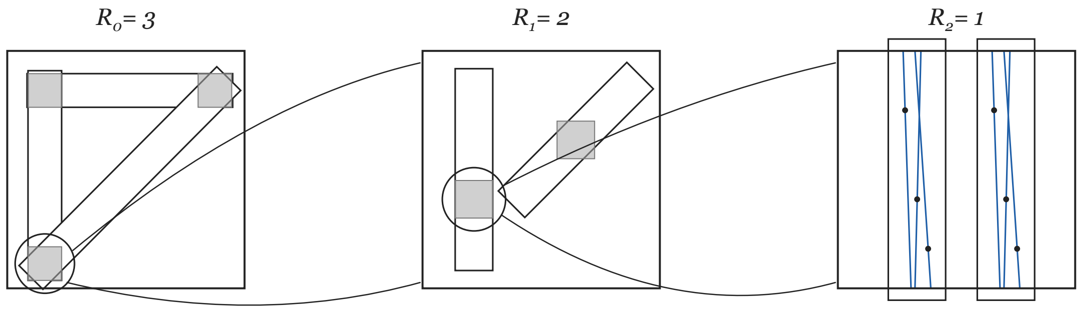

FIGURE 7. In the original configuration, $R_0 = 3$, because there are 3 different slopes of w-tubes. If we try to apply Lemma 3.1, we get $A = R_1 = 2$ (number of w-tubes through a square) and $B = R_0 = 3$, so the initial estimate is not too good. If we zoom in, we get $R_1 = 2$. The tubes in the second diagram are vertical or diagonal, because the tubes through the gray box in the first diagram are either vertical or diagonal. We zoom in again and find a configuration where all the lines point approximately vertical. Such a configuration has $A = B = 1$, so it has a good initial estimate.

We are assuming by pigeonholing that this quantity is roughly constant over all the w^j-squares Q active in covering $\mathbf{X}$.

By Lemma 3.1,

$$I(w; P[\mathbf{X}^{\mathbf{Q}}], L[\mathbf{X}^{\mathbf{Q}}]) \gtrsim \frac{R_{j+1}}{R_j} \left(w\, |P[\mathbf{X}^{\mathbf{Q}}]|\, |L[\mathbf{X}^{\mathbf{Q}}]|\right).$$

We have $M \geq R_0 \geq \cdots \geq R_k \geq 1$, so for some $0 \leq j \leq k-1$,

$$\frac{R_{j+1}}{R_j} \geq M^{-1/k}$$

as needed. □

Remark 10. In Dąbrowski–Goering–Orponen's paper [4], this Lemma provides the initial estimate. They call the condition $A/B \gtrsim 1$ *tightness*, whereas we call it *direction stability*. In our paper, this Lemma provides the initial estimate in a special case, but in the general case we need to use a more elaborate variant.

To prove Theorem 2.3 we will carefully choose a rectangle R and analyze the blowup $\mathbf{X}^{\mathbf{R}}$. The rectangle R will be pretty large, $\text{Width}(R) \sim \delta^{o(1)}$. We need to find a rectangle $\mathbf{R}$ satisfying two properties.

- **Initial estimate.** The blowup $\mathbf{X}^{\mathbf{R}}$ is *direction stable*:

$$\frac{w\text{-covering number of the slopes of lines through a } w\text{-square in } \mathbf{X}^{\mathbf{R}}}{w\text{-covering number of the slopes of all lines in } \mathbf{X}^{\mathbf{R}}} \gtrsim 1, \tag{3.1}$$

so Lemma 3.1 provides a good initial estimate.

- **Inductive step.** The blowup $\mathbf{X}^{\mathbf{R}}$ is *high-low regular*, meaning for all $v \in [\delta/\operatorname{Width}(R), w]$,

$$\operatorname{Err}(v; \mathbf{X}^{\mathbf{R}}) = \Big(\frac{\sup_{Q \text{ is } v \times v} |P[\mathbf{X}^{\mathbf{R}}] \cap Q|}{|\mathbf{X}^{\mathbf{R}}|} \frac{\sup_{T \text{ is } v \times 1} |L[\mathbf{X}^{\mathbf{R}}] \cap T|}{|\mathbf{X}^{\mathbf{R}}|} v^{-3} \Big)^{\frac{1}{2}} \ll 1, \tag{3.2}$$

so we can use the high-low inequality (Theorem 2.4) to show $|B(w_i; \mathbf{X}^{\mathbf{R}}) - B(\delta/\operatorname{Width}(R); \mathbf{X}^{\mathbf{R}})| \ll 1$.

If $\mathbf{R}$ satisfies both of these properties, we say it is *effective*.

The α-Frostman hypothesis means that $\mathbf{X}$ itself is high-low regular, but it might not be direction stable. On the other hand, we can apply Lemma 3.2 to find a square Q such that $\mathbf{X}^{\mathbf{Q}}$ is direction stable, but it might not be high-low regular. The tricky part is to find a rectangle satisfying both conditions at once.

Here is a high-level outline. The space of rectangles we can blowup into is described by two parameters, the length and the width. We describe this parameter space with $\mathsf{B} = [0, 1/10] \times [0, 1/10]$, see Figure 8. The point $(x, y) \in \mathsf{B}$ corresponds to Length $= \delta^x$, Width $= \delta^{x+y}$. We

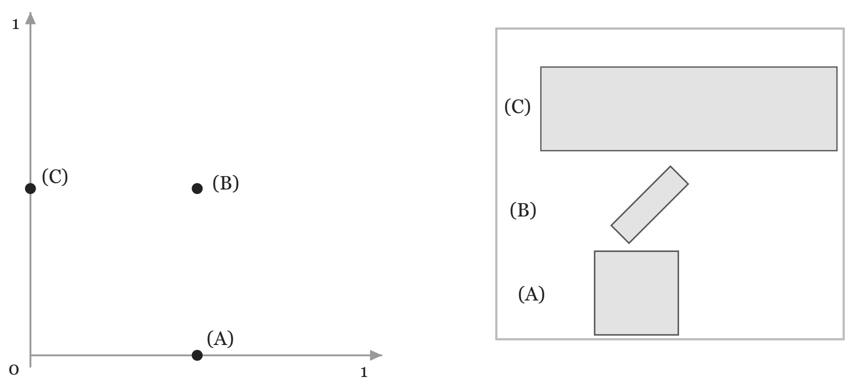

FIGURE 8. The space of rectangles is parametrized by (x, y).

carefully design a measure μ on B. The measure changes depending on what $\mathbf{X}$ looks like, it is not fixed ahead of time. Then, we show that with respect to this measure, the direction stable scales have density $1 - o(1)$ and the high-low regular scales have some fixed density $c(\alpha) > 0$, which tends to zero as $\alpha \to \frac{3}{2}$. Thus, there is some scale that is both direction stable and high-low regular. We zoom into a rectangle with these scales to prove the theorem.

Remark 11. Roth [20] introduced the idea of combining an initial estimate with an inductive step to locate lots of incidences. Recall that in Lemma 1.6, when we estimate Heilbronn's problem using the minimal distance problem, we make a line set by connecting pairs of points at distance $\sim n^{-1/2}$. Roth instead made a line set by connecting pairs of points at distance u, where $u \in [n^{-1/2}, 1]$ is a parameter. In [2], we interpret Roth's initial estimate as directly proving $A \sim w^{-1}$. This involves estimating the direction set of pairs of points in a $u \times u$-square. We recover Komlos–Pintz–Szemeredi's exponent $n^{-8/7}$ using a classical direction set estimate due to Marstrand [15], and we gave a slight improvement using a sharper recent direction set estimate due to Orponen–Shmerkin–Wang [17].

4. Lecture 4: Uniformity and branching functions

Recall our strategy from the last lecture. We are given an α-Frostman configuration $\mathbf{X}$, and we want to find a rectangle $\mathbf{R}$ such that the rescaled set

$$\mathbf{X}^{\mathbf{R}} = \psi^{\mathbf{R}}(\mathbf{X} \cap \mathbf{R}), \qquad \psi^{\mathbf{R}} \text{ is the affine map taking } \mathbf{R} \text{ to } [-1,1]^3$$

is both direction stable (3.1) and high-low regular (3.2).

Searching over the space of all rectangles is tricky, because there are a lot of rectangles. To describe a rectangle we need the length, the width, the center point, and the slope. To make the problem simpler, we do a lot of pigeonholing so that all rectangles with the same length and width behave the same under rescaling. Then there is just a 2-parameter space to search over.

The goal of this lecture is to do a lot of pigeonholing to describe the multiscale structure of $\mathbf{X}$ by a Lipschitz function

$$f(x, y, z), \qquad (x, y, z) \in [0, 1]^3,$$

called the *branching function* of $\mathbf{X}$. We can restate the problem of finding an effective rectangle as a problem about Lipschitz functions (c.f. Proposition 4.6).

4.1. **Uniformity and branching functions of points.** Let us start by looking at a set of points $P \subset [0, 1]^2$, rather than point-line pairs.

Recall that $|P|_w$ denotes the minimal number of w-balls needed to cover P. We say $P \subset \mathbb{R}^d$ is (δ, C)-*uniform* if for all $w \in [\delta, 1]$ and $x \in P$,

$$|P \cap B(x, w)| \in \Big[\frac{1}{C}\frac{|P|}{|P|_w}, C\frac{|P|}{|P|_w}\Big]. \tag{4.1}$$

If P is uniform, then the entire multi-scale structure of P is described by its covering numbers.

For example, if P is (δ, C_1)-uniform and $|P|_w \geq \frac{1}{C_2}w^{-\alpha}$, then

$$|P \cap Q| \lesssim C_1C_2 \text{ Width}(Q)^\alpha\, |P| \qquad \text{For all cubes } Q \text{ with Width}(Q) \in [\delta, 1].$$

In other words, P is $(\delta, \alpha, O(C_1C_2))$-Frostman.

Uniform sets have a tree-like structure, see Figure 9. To build a uniform set, start by choosing a sequence of numbers $R_1, R_2, \dots, R_\ell$; these will determine the covering numbers of the final set. Partition $[0, 1]^d$ into a grid $\mathcal{Q}_1$ of $\frac{1}{M}$-cubes (you can imagine M is a constant, or $M \sim \delta^{-o(1)}$), and select a subset of $\mathcal{Q}_1$ with cardinality R_1. For each cube Q selected in the first iteration, split it up into a family $\mathcal{Q}_2(Q)$ of $\frac{1}{M^2}$ cubes, and select a subset of size R_2. Continue in this way down to level ℓ, and make a point set P by selecting the center of every remaining cube. We have a described a point set P satisfying

$$|P| = R_1 \dots R_\ell, \qquad |P|_{M^{-j}} \sim R_1 \dots R_j \qquad \text{for } j \in \{0, \dots, \ell\}.$$

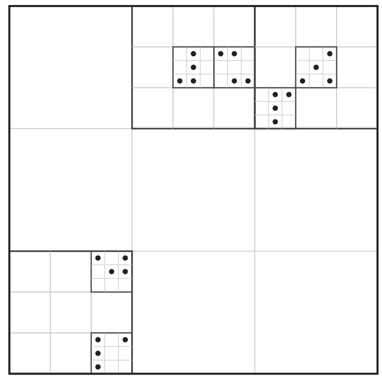

FIGURE 9. A set with a tree-like structure. Choose $M = 3$, and split the unit square into an $M \times M$ grid, and choose $R_1 = 3$ squares.. Split each of those into an $M \times M$ grid, and choose $R_2 = 2$ subsquares. Split those into a further $M \times M$ grid, and choose $R_3 = 4$ subsquares. Drop a point in each final square. Importantly, at stage k, each square gets the same numebr of children, but the children can vary from square to square.

Moreover, if $Q \in \mathcal{Q}_j$, then

$$\text{either } P \cap Q = 0, \text{ or } P \cap Q = R_{j+1} \dots R_\ell \sim \frac{|P|}{|P|_{M^{-j}}}.$$

Elsewhere in the literature uniformity is defined as sets of the above form, rather than using (4.1) (there are various technical tradeoffs).

Given a (δ, C)-uniform set P, we define the branching function to be a log-log plot of the covering numbers:

$$f(x) = -\log_\delta |P|_{\delta^x}, \qquad x \in [0,1].$$

The definition makes sense whether or not P is uniform, but it is most useful when P is uniform. For example, if P is $(\delta, \delta^{-o(1)})$-uniform, then it is $(\delta, \alpha, \delta^{-o(1)})$-Frostman if and only if $f(x) \geq \alpha x - o(1)$.

See Figure 10 for examples of uniform sets and associated branching functions.

The key technical Lemma says that every finite set admits large uniform subsets.

Lemma 4.1. *Let $P \subset [0,1]^d$ be a finite set, let $\delta \in (0, 1/3]$ be a parameter, and let $C = O(2^{d(\log \delta^{-1})^{0.6}}) \leq \delta^{-o(1)}$. There exists a (δ, C)-uniform subset $P' \subset P$ such that $|P'| \geq \frac{1}{C}|P|$.*

Proof. Let $T = \ell = \lceil \sqrt{\log_2 \delta^{-1}} \rceil$, so that $2^{-\ell T} \leq \delta$.

For $j \in \{0, \dots, \ell\}$, let $\mathcal{Q}_j$ be the grid of dyadic 2^{-jT}-cubes. When X is a set, we denote by $\mathcal{Q}_j(X)$ the cubes intersecting X.

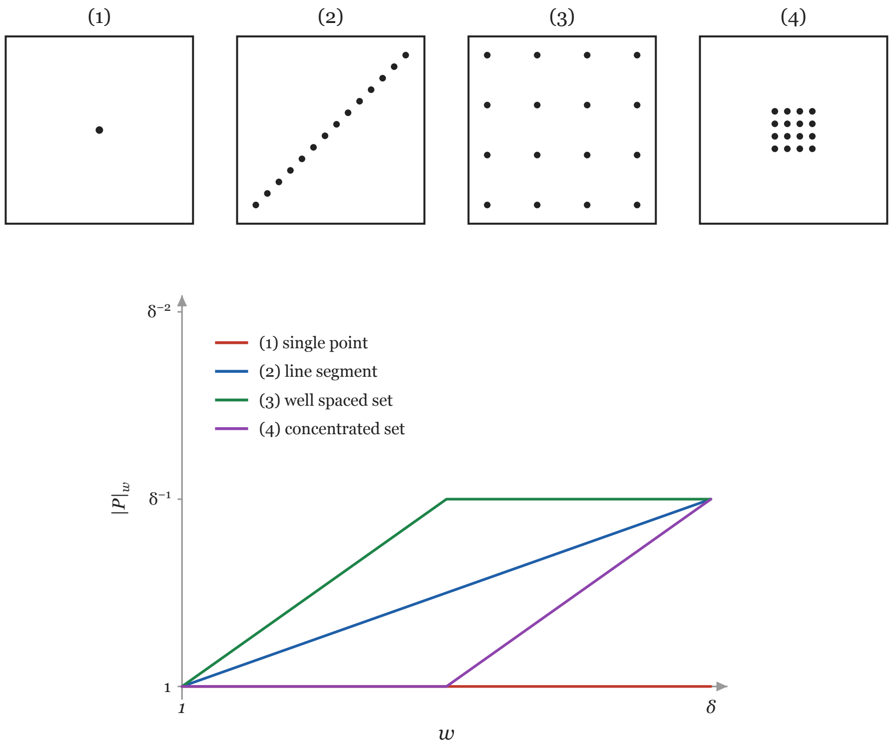

FIGURE 10. Branching functions of various sets. The last 3 sets all have δ^{-1} points, but their covering numbers at intermediate scales are different.

Start with $j = \ell$. Since there are $2^{d\ell T} \lesssim 2^d \delta^{-d}$ total cubes in $\mathcal{Q}_\ell$, by pigeonholing, there is some number $r_\ell > 0$ such that

$$\sum_{Q \in \mathcal{Q}_\ell} |P \cap Q| 1_{|P \cap Q| \in [r_\ell, 2r_\ell)} \gtrsim \frac{1}{d \log \delta^{-1}} |P|.$$

Define P_ℓ by restricting to the cubes Q above, and removing at most half the remaining points so that

$$|P_\ell| \gtrsim \frac{1}{d \log \delta^{-1}} |P|, \text{ and for all } Q \in \mathcal{Q}_\ell(P_\ell),\ |P_\ell \cap Q| = r_\ell.$$

Repeat this procedure for $j = \ell - 1$ all the way through $j = 0$. At step j, we produce a subset $P_j \subset P_{j+1}$ satisfying

$$|P_j| \gtrsim \frac{1}{d \log \delta^{-1}} |P_{j+1}|, \text{ and for all } Q \in \mathcal{Q}_j(P_j),\ |P_j \cap Q| = r_j.$$

Set $P' = P_0$. We have

$$|P'| \gtrsim (d \log \delta^{-1})^{-\ell} |P| \geq C^{-1} |P| \tag{4.2}$$

$$|P'| \gtrsim (d \log \delta^{-1})^{-\ell} |P|, \text{ and for all } j \in \{0, \ldots, \ell\} \text{ and } Q \in \mathcal{Q}_j(P_j),\ |P' \cap Q| = r_j.$$

Let $w \in [\delta, 1]$ be a scale, and consider a ball of the form $B(p, w)$ for $p \in P$. We can find two scales $\underline{w} \le w \le \overline{w}$ of the form 2^{-jT}, and boxes $Q_{\underline{w}}, Q_{\overline{w}}$, such that

$$Q_{\underline{w}} \subset B(p, w) \cap [0,1]^2 \subset Q_{\overline{w}}$$
$$\overline{w} \le 2^{2T} w, \qquad \underline{w} \ge 2^{-2T} w.$$

Thus

$$\frac{|P'|}{|P'|_{\underline{w}}} \lesssim |P' \cap B(p, w)| \lesssim \frac{|P'|}{|P'|_{\overline{w}}}.$$

For $w_1 < w_2$, there is an easy bound $|P'|_{w_1} \lesssim (w_1/w_2)^{-d} |P'|_{w_2}$. Thus

$$|P' \cap B(p, w)| \in \big[C^{-1} \frac{|P'|}{|P'|_w}, C \frac{|P'|}{|P'|_w}\big] \qquad \text{because } C \ge 2^{2dT}.$$

□

Remark 12. The earliest use of uniform sets and branching functions I know of is Bourgain's sum-product paper [1]. They have since been used extensively in fractal geometry, with an especially influential early application by Keleti–Shmerkin [10]; see also Shmerkin's ICM notes [26, Section 2]. I wonder whether there are earlier uses of branching functions—they seem like such a natural object to study.

To illustrate why uniformity is useful: while working on our first Heilbronn paper [2], we initially reasoned directly about the various squares and how sets behaved inside each one—my notebooks from then are full of drawings of squares and points. Once we learned about branching functions, we could rephrase the same strategy in terms of Lipschitz functions, and my notebooks filled up with those instead. Uniformity is even more crucial in our second paper [3].

4.2. **Uniformity and branching functions in phase space.** Recall from Section 2.3 that we consider rectangles in $\mathbb{R}^2$ of the form

$$R = (x_0, y_0) + \{(x, y + z_0 x) : x \in [-u, u], y \in [-v, v]\} \qquad (x_0, y_0, z_0) \in [-1, 1]^3.$$

We define

$$\text{Slope}(R) = z_0, \qquad \text{Slope-Set}(R) = \big[z_0 - \frac{v}{u}, z_0 + \frac{v}{u}\big].$$

A *phase-space rectangle* is a set of the form

$$\mathbf{U} = R \times [z - w, z + w], \qquad z \in \text{Slope-Set}(R),\ w \in [\delta, v/u]. \tag{4.3}$$

We say $\mathbf{U}$ has dimensions $u \times w \times v$.

Here, u is the length of the rectangle; v is the width; and w is the interval of slopes we cover. We enforce $w \le v/u$ so that the slopes in $\mathbf{U}$ are aligned with R

It is unfortunate notation that the interval for the third coordinate, w, appears second in $u \times w \times v$. But it is convenient later.

Let $\mathbf{X} \subset [-1, 1]^3$ be a configuration. We define

$$|\mathbf{X}|_{u \times w \times v} = \text{The minimal covering number of } \mathbf{X} \text{ by } u \times w \times v \text{ rectangles.}$$

For any $\omega \in [-1,1]^3$ and dimensions $u \times w \times v$, we can define a rectangle $\mathbf{U}_{u\times w\times v}(\omega)$ centered at ω.

Definition 4.2. We say a configuration $\mathbf{X} \subset [-1,1]^3$ is (δ, C)-uniform if for all $\omega \in \mathbf{X}$ and scales $u, w, v \in [\delta, 1]$, $w \leq v/u$,

$$|\mathbf{X} \cap \mathbf{U}_{u\times w\times v}(\omega)| \in [C^{-1}\frac{|\mathbf{X}|}{|\mathbf{X}|_{u\times w\times v}}, C\frac{|\mathbf{X}|}{|\mathbf{X}|_{u\times w\times v}}].$$

Lemma 4.3. *For any configuration* $\mathbf{X}$ *and* $\delta \in (0, 1/3]$*, there exists a* $(\delta, \delta^{-o(1)})$*-uniform subset* $\mathbf{X}' \subset \mathbf{X}$ *with* $|\mathbf{X}'| \geq \delta^{o(1)}|\mathbf{X}|$.

The proof of this Lemma involves associating a hypergraph to $\mathbf{X}$ and pruning the edges. It uses a very general pigeonholing Lemma about hypergraphs. See [3, Lemma 3.6]

Once we pass to uniform sets, pretty much every quantity we care about is pigeonholed. For example, if $\mathbf{X}$ is uniform, then

- The underlying point set $P[\mathbf{X}]$ is uniform: Every square contains ~ 0 points or $\sim$ const. points.
- The underlying line set $L[\mathbf{X}]$ is uniform: Every tube contains ~ 0 lines or $\sim$ const. lines.
- The rectangles are uniform. For $\delta \leq v \leq u \leq 1$, every $u \times v$ rectangle with slope in $[-1, 1]$ either intersects ~ 0 or $\sim$ const. point-line pairs. Recall that a point-line pair lies in a rectangle if the point lies in the rectangle and the slope of the line is aligned with the rectangle.
- The direction set $\mathrm{Slope}(\mathbf{X}) = \{z : (x, y, z) \in \mathbf{X}\} \subset [-1, 1]$ is uniform as a multiset. If we look in some w-sector of directions, there are ~ 0 lines pointing in that direction or $\sim$ const. lines.
- The lines through a point are uniform. For every $w \in [\delta, 1]$, and any δ-ball, the number of point-line pairs through the δ-ball and pointing in a particular w-sector is ~ 0 or $\sim$ const.

We define a branching function

$$f(x, y, z) = \log_{\delta^{-1}} |\mathbf{X}|_{\delta^x\times\delta^y\times\delta^z}, \qquad \mathrm{Domain}(f) = \{(x, y, z) : x, y \in [0, 1], z \in [0, \min(1, x+y)]\}.$$

We define a 2-parameter slice

$$f(x, y) = f(x, y, x + y).$$

- $\delta^{-f(x,y)}$ counts the number of $\delta^x \times \delta^{x+y}$ rectangles needed to cover $\mathbf{X}$. In particular, if $\mathbf{X}$ is $(\delta, \alpha, \delta^{-o(1)})$-Frostman, then $|\mathbf{X}|_{u\times uw\times w} \geq \delta^{o(1)}(uw)^{-\alpha}$, so $f(x, y) \geq \alpha(x + y)$.
- $f(x, 0)$ is the branching function of $P[\mathbf{X}]$.
- $f(0, y)$ is the branching function of $L[\mathbf{X}]$.
- $f(0, t, 0)$ is the branching function of the slope set of L.

We can use the branching function to analyze the initial estimate and inductive step.

- Let $w = \delta^t$ and apply Lemma 3.1. We get

$$A \approx \frac{|\mathbf{X}|_{w\times w\times w}}{|\mathbf{X}|_{w\times 1\times w}} = \delta^{-f(t,t,t)+f(t,0,t)},$$
$$B \approx |\mathbf{X}|_{1\times w\times 1} = \delta^{-f(0,t,0)}.$$

So,

$$B(w; P[\mathbf{X}], L[\mathbf{X}]) \gtrapprox \frac{A}{B} \gtrsim \delta^{f(0,t,0)+f(t,0,t)-f(t,t,t)}.$$

- We estimate the high–low error (2.10).

$$\sup_{Q \text{ is } w\times w} |P[\mathbf{X}] \cap Q| \sim \frac{|\mathbf{X}|}{|\mathbf{X}|_{w\times 1\times w}} \sim |\mathbf{X}|\, \delta^{f(t,0)},$$
$$\sup_{T \text{ is } w\times 1} |L \cap T| \sim \frac{|\mathbf{X}|}{|\mathbf{X}|_{1\times w\times w}} \sim |\mathbf{X}|\, \delta^{f(0,t)},$$

so the high–low error at scale w is

$$\Big(\frac{\sup_{Q \text{ is a } w\times w \text{ square}} |P\cap Q|}{|P|} \frac{\sup_{T \text{ is a } w\times 1 \text{ tube}} |L \cap T|}{|L|} w^{-3}\Big)^{1/2} \sim \delta^{\frac{1}{2}(f(t,0)+f(0,t)-3t)}.$$

By applying the high–low method many times across many scales,

$$\frac{I(\delta)}{\delta\,|P|\,|L|} \gtrapprox \frac{I(w)}{w\,|P|\,|L|} - \sup_{s\in[t,1]} \delta^{\frac{1}{2}(f(s,0)+f(0,s)-3s)}, \qquad w = \delta^t.$$

To run the incidence estimate strategy, choose a triple $(t; x_0, y_0)$. We will try to zoom into a $\delta^{x_0} \times \delta^{x_0+y_0}$ rectangle, apply the initial estimate at scale δ^t, and then run the high low method down to scale δ. The branching function of the rescaled set $\mathbf{X}^{\mathbf{R}}$ has a simple form:

$$f_{\mathbf{X}^{\mathbf{R}}}(x,y,z) = f(x,y,z;x_0,y_0) := f(x+x_0, y+y_0, z+x_0+y_0) - f(x_0,y_0,x_0+y_0).$$

We use $(\bullet; x_0, y_0)$ to denote some quantity after blowing up into a rectangle.

Definition 4.4. Let $\mathbf{X}$ be a point-line configuration, let f be the branching function, and let $(t; x_0, y_0)$ a triple.

- We say $(t; x_0, y_0)$ is *ρ-direction-stable* if

$$f(0,t,0;x_0,y_0) + f(t,0,t;x_0,y_0) - f(t,t,t;x_0,y_0) \le \rho t. \tag{4.4}$$

- We say $(t; x_0, y_0)$ is *ρ-high-low-regular* if

$$f(s,0;x_0,y_0) + f(0,s;x_0,y_0) - 3s \ge 2\rho t \qquad \text{for } s \in [t, 1-x_0-y_0]. \tag{4.5}$$

- We say $(t; x_0, y_0)$ is *ρ-effective* if it is both direction stable and high-low regular.

Proposition 4.5. *If $(t; x_0, y_0)$ is ρ-effective, then*

$$\frac{I(\delta)}{\delta\,|P|\,|L|} \ge \frac{I(\delta^{1-x_0-y_0})}{\delta^{1-x_0-y_0}\,|P|\,|L|} \gtrapprox \delta^{\rho t + x_0 + y_0}.$$

We have reduced the incidence lower bound theorem (Theorem 2.3) to a problem about branching functions.

Proposition 4.6. *Let $\alpha, \beta > 0$ satisfy $\alpha + \beta > 3$. For any $\varepsilon > 0$, the following holds for some $\rho > 0$ and for δ sufficiently small. Let $f(x, y, z)$ be a branching function of a scale-δ phase-space set, and suppose $f(x, y) \geq \alpha x + \beta y$. Then there exists a ρ-effective triple $(t; x_0, y_0)$ with $t, x_0, y_0 \in [0, \varepsilon]$.*

Recall that if $\mathbf{X}$ is $(\delta, \alpha, \delta^{-o(1)})$-Frostman, then $f(x, y) \geq \alpha x + \alpha y$.

The incidence lower bound theorem follows from combining Proposition 4.5 (two step incidence strategy) with Proposition 4.6 (location of an effective rectangle).

5. Lecture 5 (bonus): Solving the branching function problem

This lecture was not covered during the minicourse.

The goal of this lecture is to prove Proposition 4.6: If f is a phase space branching function and $f(x, y) \geq \alpha x + \beta y$, with $\alpha + \beta > 3$, then we can find an effective triple.

5.1. **Relevant properties of phase-space branching functions.**

In [3], we define the space

$$\mathcal{L} = \{\text{Branching functions } f \text{ of point-line configurations}\}.$$

Technically, $\mathcal{L}$ is the space of subsequential limits of phase-space branching functions as the scale parameter $\delta \to 0$. Can we describe the space $\mathcal{L}$?

There does not appear to be any simple way to describe $\mathcal{L}$. Lots of deep information about two-dimensional incidence geometry can be encoded as information about $\mathcal{L}$. For example, the Furstenberg set theorem of Ren–Wang [18] (following crucial work of Orponen–Shmerkin [16]) can be phrased in terms of branching functions. In some sense, describing the space $\mathcal{L}$ would amount to "solving" two-dimensional incidence geometry (up to $\delta^{-o(1)}$ errors).

To prove Proposition 4.6, we only need to use the following simple properties of $\mathcal{L}$.

- For $x', y', z' \geq 0$ we have

  $$0 \leq f(x + x', y + y', z + z') - f(x, y, z) \leq x' + y' + z' + o(1).$$

  In other words, f is monotone and Lipschitz. Monotonicity uses the fact that covering number increases as the size of the rectangles goes down. Lipschitz uses an easy bound for the covering number at a smaller scale in terms of the larger scale.
- Let $\mathbf{a}_1 = (x_1, y_1, z_1)$ and $\mathbf{a}_2 = (x_2, y_2, z_2)$ lie in the domain of f. It follows that the coordinate-wise maximum $\max\{\mathbf{a}_1, \mathbf{a}_2\}$ lies in the domain of f; assume $\min\{\mathbf{a}_1, \mathbf{a}_2\}$ lies in the domain as well. Then

  $$f(\max\{\mathbf{a}_1, \mathbf{a}_2\}) + f(\min\{\mathbf{a}_1, \mathbf{a}_2\}) \leq f(\mathbf{a}_1) + f(\mathbf{a}_2) + o(1).$$

  This property is called *submodularity*. It uses geometry of rectangles.

We describe submodularity with two examples.

Example 5.1 (Examples of submodularity)**.**

- Let P be a family of points and let $\mathbb{T}$ be a family of δ-tubes in $[0,1]^2$. In projection theory, we are interested in estimating $|T \cap P|$ for typical $T \in \mathbb{T}$. A common strategy is to use induction on scales: First estimate the number of w-balls on a w-tube, then estimate the number of δ-balls on a $\delta \times w$-tubelet,

$$\#(\delta\text{-balls on a } \delta\text{-tube}) \leq \#(w\text{-balls on a } w\text{-tube}) + \#(\delta\text{-balls on a } \delta \times w\text{-tubelet}). \tag{5.1}$$

This inequality is an example of submodularity.

We can describe this configuration by the phase-space set

$$\mathbf{X} = \{(p, \ell_p(T)) : (p,T) \in P \times \mathbb{T} \text{ and } p \in T\}, \qquad \ell_p(T) \text{ is a line in } \mathbb{T} \text{ through } p. \tag{5.2}$$

Assume $\mathbf{X}$ is pigeonholed, and let $w = \delta^t$. The three terms in (5.1) are described by the covering numbers

$$\#(\delta\text{-balls on a } \delta\text{-tube}) = \frac{|\mathbf{X}|_{\delta\times\delta\times\delta}}{|\mathbf{X}|_{1\times\delta\times\delta}} = \delta^{-\big(f(1,1,1)-f(0,1,1)\big)},$$
$$\#(w\text{-balls on a } w\text{-tube}) = \frac{|\mathbf{X}|_{w\times w\times w}}{|\mathbf{X}|_{1\times w\times w}} = \delta^{-\big(f(t,t,t)-f(0,t,t)\big)},$$
$$\#(\delta\text{-balls on a } \delta \times w\text{-tubelet}) = \frac{|\mathbf{X}|_{\delta\times\delta\times\delta}}{|\mathbf{X}|_{w\times\frac{\delta}{w}\times\delta}} = \delta^{-\big(f(1,1,1)-f(t,1-t,1)\big)}.$$

Eq. (5.1) is expressed by the inequality

$$f(0,t,t) + f(t,1-t,1) \leq f(0,1,1) + f(t,t,t).$$

By monotonicity $f(t,1-t,1) \leq f(t,1,1)$, and by submodularity $f(0,t,t)+f(t,1,1) \leq f(0,1,1) + f(t,t,t)$.

- Given a branching function f, let $d(t) = f(0,t,0)$. This is the branching function of the slope set of $L[\mathbf{X}]$.

  Let $d(t;x_0,y_0) = f(0,t,0;x_0,y_0)$ be the slope set branching function after zooming into a $\delta^{x_0} \times \delta^{y_0}$ rectangle.

  Suppose we want to measure the w-covering number of the set of slopes; let $w = w_1 w_2$. First measure the w_1-covering number of the set of slopes, say this is M_1. Then pick M_1 many tubes pointing in w_1-separated directions. Suppose that the w-covering number in each of these is M_2. Then the w-covering number of all the slopes is $\geq M_1 M_2$.

  Using branching functions, let $w = \delta^t$, $w_1 = \delta^{t_1}$, $w_2 = \delta^{t_2}$. The above estimate reads

$$d(t_1+t_2) \geq d(t_1) + d(t_2; 0, t_1).$$

In terms of f, this inequality expands to

$$f(0, t_1+t_2, 0) \geq f(0,t_1,0) + f(0,t_1+t_2,t_1) - f(0,t_1,t_1),$$

which follows from submodularity. See Figure 11.

It it useful to rewrite the direction stable condition (4.4) in terms of the direction numbers

$$d(t;x_0,y_0) = f(0,t,0;x_0,y_0)$$

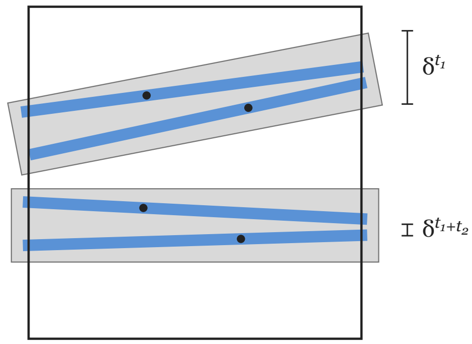

FIGURE 11. There are 2 directions at scale δ^{t_1}. Each δ^{t_1} tube has 2 directions inside of it. So there are at least 4 directions at scale $\delta^{t_1+t_2}$ (there might be more).

from Example 5.1. The triple $(t; x_0, y_0)$ is ρ-direction-stable if

$$d(t; x_0, y_0) - d(t; x_0 + t, y_0) \leq \rho t. \tag{5.3}$$

We also need to consider the dual direction function

$$d^{\vee}(t; x_0, y_0) = f(t, 0, 0; x_0, y_0)$$

which satisfies the same properties as $d(t; x, y)$ but with the roles of x and y swapped. This function has a simple meaning: $d^{\vee}(t; 0, 0)$ is the δ^t-covering number of the set of x-coordinates of $P[\mathbf{X}]$.

Using monotonicity, Lipschitz, and submodularity, we can derive the following about the direction function d (and similar properties hold for the dual direction function $d^{\vee}$).

- For $t', x', y' \geq 0$, the direction function satisfies

$$\begin{aligned} 0 \leq d(t + t'; x, y) - d(t; x, y) &\leq t' + o(1), \\ 0 \leq d(t; x, y) - d(t; x + x', y) &\leq 2x' + o(1), \\ |d(t; x, y + y') - d(t; x, y)| &\leq 2y' + o(1). \end{aligned}$$

  In particular, d is monotone with respect to t and x and Lipschitz with respect to all three parameters.
- As described in Example 5.1, the direction function satisfies

$$d(t + s; x, y) \geq d(t; x, y) + d(s; x, y + t) - o(1),$$

5.2. **Proof of Proposition 4.6.** Let $\mathsf{B} = [0, 1/10] \times [0, 1/10] \times [0, 1/10]$. This domain describes our candidate triples $(t; x_0, y_0)$. Let

$$E_{\text{DS}} = \text{Direction stable triples in } \mathsf{B}, \qquad E_{\text{HLR}} = \text{High-low regular triples in } \mathsf{B},$$

both measured with respect to a TBD tolerance ρ. The goal is to show $E_{\text{DS}} \cap E_{\text{HLR}}$ is nonempty.

The first step of our strategy is to design a special measure μ on B, which depends on the specific branching function f at hand. This measure is unbounded, but we can measure the density of E_{DS} and E_{HLR} with respect to μ. We will show

$$\text{Density}_\mu(E_{\text{DS}}) = 1 - o(1), \qquad \text{Density}_\mu(E_{\text{HLR}}) \geq c(\alpha + \beta) > 0.$$

The density lower bound on E_{HLR} depends on $\alpha + \beta$ (the Frostman parameters in Proposition 4.6), and it tends to zero as $\alpha + \beta$ tends to 3.

We describe an example of a $(\delta, \frac{3}{2}, C)$-set, which will motivate our construction of μ.

Example 5.2 (Example of a (δ, α, C)-set)**.** Let $\varphi : [-1, 1]^2 \to [-1, 1]$ be a smooth function with $|(1, \varphi(x, y)) \cdot \nabla\varphi(x, y)| \sim 1$ everywhere. Let

$$\mathbf{X} = \{(x, y, \varphi(x, y)) \,:\, (x, y) \in \delta\mathbb{Z}^2 \cap [-1, 1]^2\}.$$

The underlying point set is a δ-grid, and the line through the point (x, y) has slope $\varphi(x, y)$. We claim

$$f(x, y) = \max\{2x + y, x + 2y\} + o(1),$$

so $\mathbf{X}$ is a $(\delta, \frac{3}{2}, O(1))$-set.

The example above motivates a definition. Say (x, y) is an (α, β)-good scale for the branching function f if

$$f(x', y'; x, y) = f(x + x', y + y') - f(x, y) \geq \alpha x' + \beta y'.$$

In other words, the blowup of $\mathbf{X}$ into a $\delta^x \times \delta^{x+y} \times \delta^y$ rectangle is Frostman. In the example, (x, x) is a $(\frac{3}{2}, \frac{3}{2})$-good scale for every $x \in [0, 1/2]$.

Good scales are automatically high-low regular (4.5). They are useful because we can find other high-low regular triples near good scales. In this example, there is a whole line segment of good scales through the origin.

We will build μ to be supported near there. After zooming in, we may reduce to the special case where f has a line segment of good scales through the origin, and it is this case that we now focus on.

Case 1: There is a horizontal line of good scales. Suppose that $(x, 0)$ is an (α, β)-good scale for every $x \in [0, 1]$. In this case we don't even need to construct μ, we just repeat the argument from Lemma 3.2 to find a direction stable scale of the form $(t; x, 0)$.

Let $\rho \leq (\alpha + \beta - 3)/50$ and $t = \rho/50$. Let $m = \lfloor 0.1/t \rfloor$, and consider the telescoping sum

$$(d(t; 0, 0) - d(t; t, 0)) + \cdots + (d(t; (m-1)t, 0) - d(t; mt, 0)) = d(t; 0, 0) - d(t; mt, 0).$$

Each of the m summands on the left is nonnegative, and the right hand side is $\leq t$. So there must be some $0 \leq j \leq m-1$ so that $d(t; jt, 0) - d(t; (j+1)t, 0) \leq \frac{1}{m} t < \rho t$. The triple $(t; jt, 0)$

is direction stable with parameter ρ, and it is also high-low regular with parameter ρ because of the assumption that $(jt, 0)$ is an (α, β)-good scale. Thus $(t; jt, 0)$ is effective, as desired.

Case 2: There is a diagonal line of good scales. Suppose that (x, x) is an (α, β)-good scale for every $x \in [0, 1/2]$. In the prior case we searched for an effective triple based at a good scale, so high-low regularity came for free. The current case is harder because we need to locate a scale that is both direction stable and high-low regular. In Example 5.2, every $(\frac{3}{2}, \frac{3}{2})$-good scale (x, x) is high-low regular but not direction stable (for any t).

We consider candidate triples of the form $(t; x, x+t)$, see Figure 12. Recall that $(t; x, x+t)$

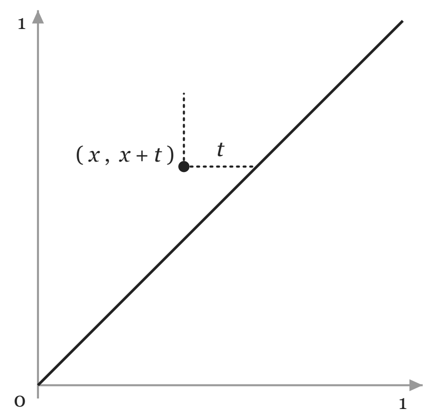

FIGURE 12. We consider candidate triples near the good line. We try to prove an initial estimate at a scale matching up with the good line.

is high-low regular with parameter ρ if

$$\frac{1}{2}[f(s, 0; x, x+t) + f(0, s; x, x+t) - 3s] \geq 2\rho t \qquad \text{for all } t \leq s \leq 1 - x - (x+t), \tag{5.4}$$

and $(t; x, x+t)$ is direction stable if

$$d(t; x, x+t) - d(t; x+t, x+t) \leq \rho t. \tag{5.5}$$

Define a measure supported on the triples $(t; x, x+t)$ with density

$$d\mu = \frac{1}{t}\, dx\, dt.$$

This measure integrates to infinity and is concentrated near $t = 0$. For each $\varepsilon > 0$, let

$$\mu_\varepsilon = \frac{1_{\{(x,t)\,:\,t \geq \varepsilon\}}}{\mu(\{(x,t)\,:\,t \geq \varepsilon\})}\mu$$

be the normalized restriction of μ to $\{t \geq \varepsilon\}$. We prove that

- For small enough ρ, E_{HLR} has positive density under the measure μ. This means that $\mu_\varepsilon(E_{\text{HLR}}) \geq c - o_\varepsilon(1)$ for some constant $c > 0$.
- For all ρ, E_{DS} has density one under the measure μ. This means that $\mu_\varepsilon(E_{\text{DS}}) \geq 1 - o_\varepsilon(1)$.

Combining these two facts proves E_{HLR} and E_{DS} have an intersection point for some choice of ρ.

We prove E_{HLR} has positive density one x-coordinate at a time. If

$$f(0,t;x,x) = s_x\, t \tag{5.6}$$

was a linear function of t for each x, then every $(x,t) \in \mathsf{B}$ would be high-low regular. Actually, this is the motivation for considering triples of the form $(t;x,x+t)$. We must have $s_x \geq \beta$ because we assumed (x,x) is an (α,β)-good point. We have

$$\begin{aligned} f(t,0;x,x+t) + f(0,t;x,x+t) - 3t &= f(t,t;x,x) - 2f(0,t;x,x) + f(0,2t;x,x) - 3t \\ &\geq (\alpha+\beta-3)t. \end{aligned}$$

where the second and third terms cancel because we assumed $f(0,t;x,x)$ is a linear function in t and we use that (x,x) is an (α,β)-good point to get $f(t,t;x,x) \geq (\alpha+\beta)t$. For $t' \geq t$ we can estimate the high-low error using the hypothesis that $(x+t,x+t)$ is a good point,

$$\begin{aligned} &[f(t',0;x,x+t) + f(0,t';x,x+t)] - [f(t,0;x,x+t) + f(0,t;x,x+t)] = \\ &\qquad = f(t'-t,0;x+t,x+t) + f(0,t+t';x,x) - f(0,2t;x,x) \geq (\alpha+\beta)(t'-t). \end{aligned}$$

This shows that under the simplifying assumption (5.6) any triple of the form $(t;x,x+t)$ is high-low regular. For general f, we can write

$$f(0,t;x,x) = s_x(t)t$$

where the slope s_x now depends on the point t. Say (x,t) is *slope minimal* if

$$s_x(t+t') \geq s_x(t) - \gamma \qquad \text{for all } t \leq t' \leq Ht$$

where γ, H are parameters. The computation above implies that if (x,t) is slope minimal then $(x,t') \in E_{\mathrm{HLR}}$ for $t \leq t' \leq (1+\eta)t$ and some $\eta > 0$. By taking successive minima of the function $s_x(t)$, we can find lots of slope minimal scales, and overall we get a lower bound on the μ-density of E_{HLR}. The main takeaway is that our choice of candidate scales along with the hypothesis that (x,x) is a good line lets us find high-low regular triples one x-coordinate at a time.

We move on to proving E_{DS} has density one. We show

$$\int t^{-1}(d(t;x,x+t) - d(t;x+t,x+t))\, d\mu_\varepsilon = o_\varepsilon(1).$$

The integral on the left hand side is $\geq \rho\, \mu_\varepsilon(E_{\mathrm{DS}}^c)$, so the estimate implies $\mu_\varepsilon(E_{\mathrm{DS}}^c) = o_\varepsilon(1)$ and E_{DS} has μ-density 1. In order to estimate this integral we have to make use of some cancellation. To do so we use the crucial direction set inequality

$$d(t+s;x,y) \geq d(t;x,y) + d(s;x,y+t).$$

With this inequality in hand, we can estimate

$$d(t;x,x+t) - d(t;x+t,x+t) \leq d(2t;x,x) - d(t;x,x) - d(t;x+t,x+t).$$

The right hand side has the advantage that it only involves direction numbers on the good line. We have

$$\begin{aligned}\int_{\mathsf{B}} t^{-1}(d(2t;x,x)-d(t;x,x)-d(t;x+t,x+t))d\mu_\varepsilon &\le \int_{\mathsf{B}} t^{-1}(d(2t;x,x)-2d(t;x,x))d\mu_\varepsilon + o_\varepsilon(1)\\ &\le o_\varepsilon(1).\end{aligned}$$

The first inequality comes from integrating out the x-coordinate, and the second inequality comes from integrating out the t-coordinate and using change of variables.

General good lines. The argument we described in Case 2 works whenever there is a line of good scales $\{(x,mx) : 0 \le x \le 1/10\}$ with slope $m \in (0,1]$. In order to deal with good lines that have slope > 1 we use point-line duality. The dual notion for the slope set of $L[\mathbf{X}]$ is the set of x-coordinates of $P[\mathbf{X}]$. By replacing lines with points and slopes with x-coordinates, the argument works in the same way.

References

[1] J. Bourgain, *The discretized sum-product and projection theorems*, J. Anal. Math. **112** (2010), 193–236.

[2] A. Cohen, C. Pohoata, and D. Zakharov, *A new upper bound for the Heilbronn triangle problem*, preprint, 2023, available at `arXiv:2305.18253`.

[3] A. Cohen, C. Pohoata, and D. Zakharov, *Lower bounds for incidences*, Invent. Math. **240** (2025), no. 3, 1045–1118, DOI 10.1007/s00222-025-01331-2.

[4] D. Dąbrowski, M. Goering, and T. Orponen, *On the dimension of s-Nikodým sets*, preprint, 2024, available at `arXiv:2407.00306`.

[5] O. Frostman, *Potentiel d'équilibre et capacité des ensembles avec quelques applications à la théorie des fonctions*, Meddel. Lunds Univ. Mat. Sem. **3** (1935), 1–118 (French).

[6] L. Guth, *Projection theory*, Lecture notes for MIT 18.156, Spring 2025, MIT OpenCourseWare, `https://ocw.mit.edu/courses/18-156-projection-theory-spring-2025/`.

[7] L. Guth, N. Solomon and H. Wang, Incidence estimates for well spaced tubes, Geom. Funct. Anal. **29** (2019), no. 6, 1844–1863; MR4034922

[8] Z. Hunter, C. Pohoata, J. Verstraëte, and S. Zhang, *Large point-line matchings and small Nikodym sets*, preprint, 2026, available at `arXiv:2601.19879`.

[9] N. H. Katz and T. Tao, *Some connections between Falconer's distance set conjecture and sets of Furstenburg type*, New York J. Math. **7** (2001), 149–187.

[10] T. Keleti and P. Shmerkin, *New bounds on the dimensions of planar distance sets*, Geom. Funct. Anal. **29** (2019), no. 6, 1886–1948.

[11] J. Komlós, J. Pintz, and E. Szemerédi, *On Heilbronn's triangle problem*, J. London Math. Soc. (2) **24** (1981), no. 3, 385–396.

[12] J. Komlós, J. Pintz, and E. Szemerédi, *A lower bound for Heilbronn's problem*, J. London Math. Soc. (2) **25** (1982), no. 1, 13–24.

[13] Yu. V. Linnik, *The large sieve*, Dokl. Akad. Nauk SSSR **30** (1941), 292–294 (Russian).

[14] A. Logunov and D. Zakharov, *A fractal-like configuration of point-line pairs for the minimal distance problem*, preprint, 2025, available at `arXiv:2511.10509`.

[15] J. M. Marstrand, *Some fundamental geometrical properties of plane sets of fractional dimensions*, Proc. London Math. Soc. (3) **4** (1954), 257–302.

[16] T. Orponen and P. Shmerkin, *Projections, Furstenberg sets, and the ABC sum-product problem*, preprint, 2023, available at `arXiv:2301.10199`.

[17] T. Orponen, P. Shmerkin, and H. Wang, *Kaufman and Falconer estimates for radial projections and a continuum version of Beck's theorem*, Geom. Funct. Anal. **34** (2024), no. 1, 164–201.

[18] K. Ren and H. Wang, *Furstenberg sets estimate in the plane*, preprint, 2023, available at `arXiv:2308.08819`.

[19] K. F. Roth, *On a problem of Heilbronn*, J. London Math. Soc. **26** (1951), 198–204.

[20] K. F. Roth, *On a problem of Heilbronn. II*, Proc. London Math. Soc. (3) **25** (1972), 193–212.

[21] K. F. Roth, *On a problem of Heilbronn. III*, Proc. London Math. Soc. (3) **25** (1972), 543–549.

[22] K. F. Roth, *Estimation of the area of the smallest triangle obtained by selecting three out of n points in a disc of unit area*, in: Analytic number theory (Proc. Sympos. Pure Math., Vol. XXIV, St. Louis Univ., 1972), Amer. Math. Soc., Providence, R.I., 1973, pp. 251–262.

[23] K. F. Roth, *Developments in Heilbronn's triangle problem*, Adv. Math. **22** (1976), no. 3, 364–385, DOI 10.1016/0001-8708(76)90100-6.

[24] I. Z. Ruzsa, *Difference sets without squares*, Period. Math. Hungar. **15** (1984), no. 3, 205–209.

[25] W. M. Schmidt, *On a problem of Heilbronn*, J. London Math. Soc. (2) **4** (1971/72), 545–550.

[26] P. Shmerkin, *Slices and distances: on two problems of Furstenberg and Falconer*, in: International Congress of Mathematicians 2022, EMS Press, Berlin, 2023, pp. 3266–3290.

[27] T. Szőnyi, *Note on the existence of large minimal blocking sets in Galois planes*, Combinatorica **12** (1992), no. 2, 227–235.

[28] L. A. Vinh, *The Szemerédi–Trotter type theorem and the sum-product estimate in finite fields*, European J. Combin. **32** (2011), no. 8, 1177–1181, DOI 10.1016/j.ejc.2011.06.008.

Courant Institute, New York University. New York, NY, USA.

Email address: `alexcohen@nyu.edu`