The Kakeya needle problem, projective geometry, and fractal dimensions

Alan Chang

Mathcamp 2020 Week 4

Contents

1 Introduction

1.1 Course blurb

Let's go through the three topics in the course title.

1. The Kakeya needle problem asks the following question: Suppose you have a unit line segment (a "needle") in the plane and you'd like to rotate it 180 degrees, so that it points in the opposite direction. What is the area of the smallest region you can do this in? This problem can be solved with elementary geometric techniques, and the answer may not be what you expect!

2. The real projective plane is like the Euclidean plane, except parallel lines intersect at "points at infinity." In this space, there is a magic trick. If you wave your wand and say the magic words ("point-line duality"), then all the points will transform into lines and vice versa. Furthermore, any theorem about points and lines that was true will still remain true!

3. We can generalize the notion of "dimension" to talk about s-dimensional sets for any nonnegative real number s. This allows us to better understand sets such as fractals. For example, the Koch snowflake (look it up!) has infinite length and zero area, and turns out to be neither 1-dimensional nor 2-dimensional. It is actually $\log_3 4$-dimensional!

For the first few days, we will discuss these three topics independently of each other. Then we will see the surprising connections they have to each other, as well as to the Kakeya conjecture, a famous unsolved problem in analysis.

2 Day 1: The Kakeya needle problem

2.1 The Kakeya needle problem

Videos for fun:

- Numberphile: https://www.youtube.com/watch?v=j-dce6QmVAQ
- Mathologer: https://www.youtube.com/watch?v=IM-n9c-ARHU

(The comments on the Mathologer video seem to be more favorable than the ones on the Numberphile video.)

Kakeya needle problem: Suppose you have a unit line segment, and you want to rotate it 180 degrees in the plane in as small area as possible. How would you do this?

Here are some possibilities:

1. Rotate the needle about its center. You get a circle of radius 1/2:

2. Do the entire rotation inside an equilateral triangle of height 1. Rotate the needle at one of its endpoints from one side of the triangle to another, and then translate along the side of the triangle.

3. Do a "three-point turn," like how you might make a U-turn. The resulting shape is called a deltoid (not to be confused with the muscle). It is like a triangle but with curved sides.

You can check by calculating the areas that the deltoid is better than the triangle is better than the circle. Calculating the area of a deltoid seems complicated... but it's not needed, because it's not optimal.

Theorem 2.1 (Answer to the Kakeya needle problem)**.** *For any $\varepsilon > 0$, there is a way to rotate a unit line segment 180 degrees in area less than ε.*

(Historical note: This was proved by Besicovitch in 1919.)

Another way to state Theorem 2.1 is that "a unit line segment can be rotated 180 degrees in arbitrarily small area." (In general, when we say a quantity is "arbitrarily small," we mean that for all $\varepsilon > 0$, there is a way to make that quantity less than epsilon.)

The rest of today will be to prove Theorem 2.1. The key idea (keep this in mind as you read the following sections) will be to cut the triangle into smaller triangles and translate the triangles so they overlap a lot.

First some reductions. It is clearly enough to prove the following

Theorem 2.2. *For any $\varepsilon > 0$, there is a way to rotate a unit line segment 60 degrees in area less than ε.*

2.2 Translations are almost "free"

If our line segment is horizontal, then we can easily translate it horizontally while covering zero area. This is because a line has zero area.

What is less trivial, is that we can translate it in other directions while covering arbitrarily small area.

Theorem 2.3 (Pál join)**.** *Suppose we want to translate a line segment from one position to another. Then for any ε, we can do this translation in area less than ε.*

Proof. Take the initial and final positions of the needle and extend them into lines. We now have two parallel lines. From the initial position, rotate the needle slightly. This covers some area. Now translate the needle until it hits the other parallel line. This covers zero area. Then rotate the needle so it lies inside the line, and then translate it to its final position. □

The trick used in Theorem 2.3 is called a "Pál join." It allows us to ignore translations and "teleport" the line segment for "free" as long as we do not rotate it in the process.

2.3 Sliding 2 triangles

Let $\triangle ABC$ be any triangle. Let its area be denoted $[\triangle ABC]$. We can draw the median from vertex C to base AB to divide the triangle into two triangles (of equal area). Now we can "slide" the two pieces towards each other so that they overlap.

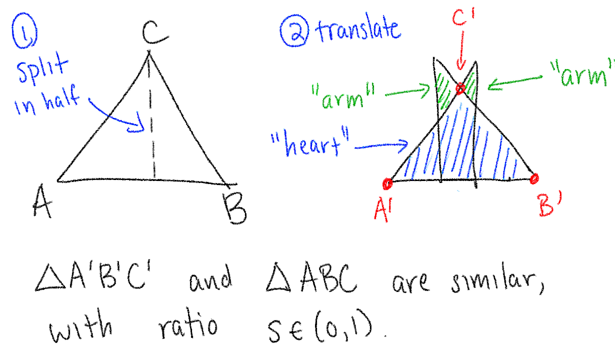


Figure 1: Sliding 2 triangles

The union of the two overlapping triangles is a complicated figure, but notice that there is a triangle that is similar to $\triangle ABC$. Let's call this the "heart." There are also two (pointy) triangles sticking out of the heart. Let's call these the "arms." See Figure 1.

The more we translate, the smaller the heart gets, and the larger the arms get. Suppose $0 < s < 1$ is the similarity ratio between the the heart and $\triangle ABC$. (That is, it is the ratio of lengths of the corresponding sides.) Then the area of the heart is $s^2[\triangle ABC]$.

Also, by elementary geometry, the total area of the two arms is $2(1-s)^2[\triangle ABC]$. One way to see this is to first consider the top half of the two arms. We can combine them together to form a triangle that is similar to $\triangle ABC$, with similarity ratio $1-s$. Thus, the total area of the top half of the two arms is $(1-s)^2[\triangle ABC]$. The bottom half has the same area as the top half.

Here is a summary of what we have shown so far.

Theorem 2.4 (Basic step)**.** *Let $0 < s < 1$. Let $\triangle ABC$ be any triangle. By doing the process above, we can obtain a figure whose with one heart and two arms. The heart is similar to $\triangle ABC$, with similarity ratio s. Furthermore,*

$$\text{area of heart} = s^2 \cdot (\text{area of } \triangle ABC) \tag{2.1}$$

$$\text{total area of arms} = 2 \cdot (1-s)^2 \cdot (\text{area of } \triangle ABC). \tag{2.2}$$

2.4 Sliding 4 triangles

Now we're going to iterate this. In the words of Elias Stein, "we will use this process to generate our monster, which will have a tiny heart and many arms." (It is also called a "Perron tree.")

Let's first see how to do this with 4 triangles. Start with a triangle $\triangle ABC$, and a similarity ratio $0 < s < 1$.

Divide the base AB into 4 segments of equal lengths, and group the 4 triangles into 2 pairs. (Steps 1 and 2 of Figure 2.) Now for each pair, do the basic construction, so for each

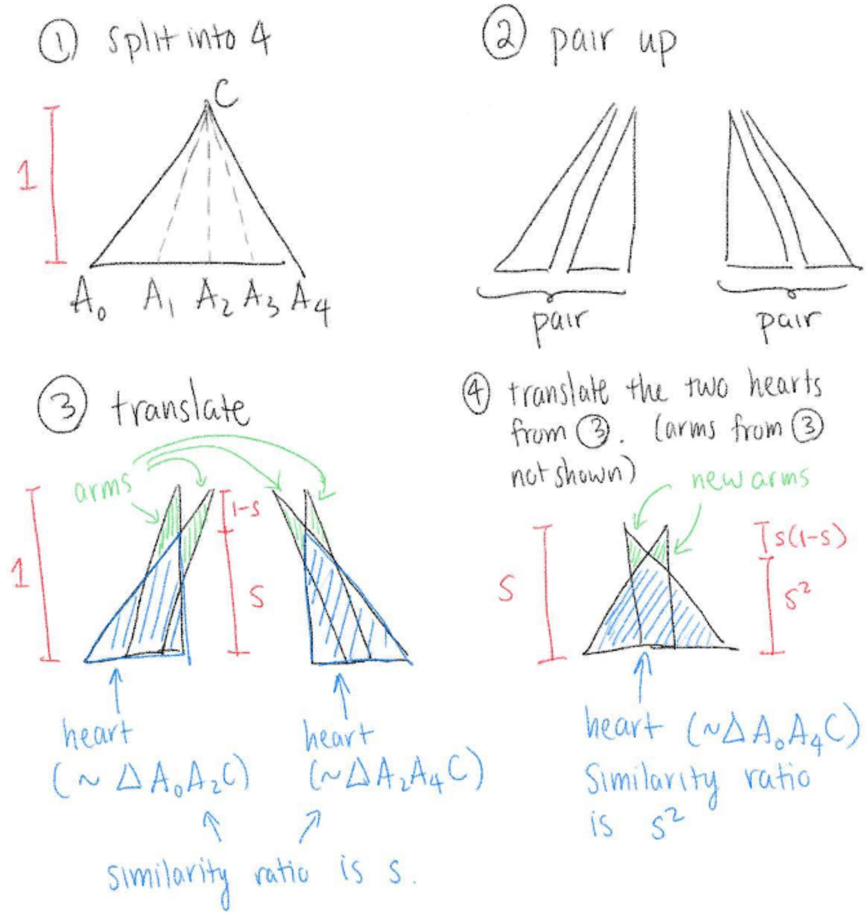


Figure 2: Sliding 4 triangles

(a) (b)

Figure 3: Stage 1 arms in blue. Stage 2 arms in green.

pair, make them overlap, so they have a heart (with similarity ratio s) as well as two arms. We will call these stage 1 hearts and stage 1 arms. (Step 3 of Figure 2.) We call these two figures "stage 1 monsters."

At this point, note that the 2 stage 1 hearts fit together, and could be combined to create a triangle similar to $\triangle ABC$ with ratio s. But we're not going to combine them. Instead, we repeat the basic process and overlap the two hearts. From the two hearts, we create a single new heart ("stage 2 heart") and two new arms ("stage 2 arms"). (Step 4 of Figure 2.)

This is the end. Our final figure (a "stage 2 monster") consists of a stage 2 heart, 4 stage 1 arms, and 2 stage 2 arms. See Figure 3.

The total area of the stage 1 arms is $2 \cdot (1-s)^2[\triangle ABC]$, which is the same as the RHS of (2.2). The argument is very similar to the case of the two triangles in Section 2.3.

Theorem 2.5 (Sliding 4 triangles)**.** *Let $0 < s < 1$. Let $\triangle ABC$ be any triangle. By doing the process above, we can obtain a figure whose with a stage 2 heart, 4 stage 1 arms, and 2 stage 2 arms. The stage 2 heart is similar to $\triangle ABC$, with similarity ratio s^2. Furthermore,*

$$\text{area of stage 2 heart} = s^4 \cdot (\text{area of } \triangle ABC) \tag{2.3}$$

$$\text{total area of stage 1 arms} = 2 \cdot (1-s)^2 \cdot (\text{area of } \triangle ABC) \tag{2.4}$$

$$\text{total area of stage 2 arms} = 2 \cdot s^2 \cdot (1-s)^2 \cdot (\text{area of } \triangle ABC). \tag{2.5}$$

(Here we're adding up the area of the each individual arm. The arms may overlap, so the total region covered by the arms may have smaller area.)

2.5 Sliding 2^n triangles

In general, we can start with any triangle $\triangle ABC$, a similarity ratio s, and a natural number n. We're going to divide the triangle into 2^n parts. (So the process described above corresponded to $n = 2$.)

First we group the 2^n triangles into 2^{n-1} pairs. We translate these to form stage 1 monsters, consisting of stage 1 hearts and stage 1 arms. Now we pair up stage 1 monsters. We combine two stage 1 monsters to form a stage 2 monster, and so on.

Theorem 2.6 (Sliding 2^n triangles)**.** *Let $0 < s < 1$. Let $\triangle ABC$ be any triangle. By doing the process above, we can obtain a figure whose with a stage n heart and for $k = 1, \dots, n$, it has 2^{n-k+1} stage k arms. The stage n heart is similar to $\triangle ABC$, with similarity ratio s^n. Furthermore,*

$$\text{area of stage } n \text{ heart} = s^{2n} \cdot (\text{area of } \triangle ABC) \tag{2.6}$$

$$\text{total area of stage } k \text{ arms} = 2 \cdot s^{2(k-1)} \cdot (1-s)^2 \cdot (\text{area of } \triangle ABC) \tag{2.7}$$

(Here we're adding up the area of the each individual arm. The arms may overlap, so the total region covered by the arms may have smaller area.)

By the geometric series formula, it follows that

$$\begin{aligned}
\frac{\text{area of stage } n \text{ monster}}{\text{area of } \triangle ABC} &\leq s^{2n} + \sum_{k=1}^{n} 2 \cdot s^{2(k-1)} \cdot (1-s)^2 && (2.8)\\
&\leq s^{2n} + \sum_{k=1}^{\infty} 2 \cdot s^{2(k-1)} \cdot (1-s)^2 && (2.9)\\
&= s^{2n} + 2 \cdot \frac{1}{1-s^2} \cdot (1-s)^2 && (2.10)\\
&= s^{2n} + 2 \cdot \frac{1-s}{1+s} && (2.11)\\
&\leq s^{2n} + 2(1-s) && (2.12)
\end{aligned}$$

2.6 The conclusion of the proof

Now we are ready to prove Theorem 2.2. We want to show that we can rotate a needle 60 degrees in arbitrarily small area.

The idea is that if $\varepsilon > 0$ is given, we can choose the parameters s and n so that $s^{2n} + 2(1-s) < \varepsilon/2$. Then we can add Pál joins so that the total area of the Pál joins is $< \varepsilon/2$. The following explains this in slightly more detail.

Let $\varepsilon > 0$ be given. First choose the similarity ratio s very close to 1, so that $2(1-s) \leq \varepsilon/4$. Then choose n very large, so that $s^{2n} \leq \varepsilon/4$. Now do the construction described above with s and n and an equilateral triangle $\triangle ABC$ of height 1. It follows that the resulting monster we get has area

$$\leq (s^{2n} + 2(1-s))[\triangle ABC] \leq (\frac{\varepsilon}{4} + \frac{\varepsilon}{4}) \cdot 1 = \frac{\varepsilon}{2}. \tag{2.13}$$

Now use Pál joins to connect the parallel line segments together. We can make the Pál joins so that the total area added is $\leq \varepsilon/2$. Thus, our region has area $\leq \varepsilon$, which completes the proof.

This concludes our solution to the Kakeya needle problem.

3 Day 1 exercises

Exercise 3.1. ( ) Not really an "exercise" but this is important. Go through the argument on your own and make sure you understand all the steps! One good way to see if you understand the argument is if you can write out all the steps yourself.

Exercise 3.2. ( ) Let's study the Kakeya needle problem in three dimensions. For any $\varepsilon > 0$ show that there exists a set $K \subset \mathbb{R}^3$ with volume $\leq \varepsilon$, and with the following property:

For any initial and final directions, it is possible to start with the needle inside K pointing in the initial direction, and move it, completely within K, so that it points in the final direction. (The orientation of the needle is all that matters. The actual position is not important.)

4 Day 2: Projective geometry

References: Jürgen Richter–Gebert, Perspectives on Projective Geometry

Projective geometry is useful in computer vision and computer graphics, drawing (perpsectives, vanishing geometry), algebraic geometry.

Later we'll be able to use projective geometry to give us a better understanding the sliding triangles construction from Day 1.

Basic idea:

1. $\mathbb{R}^2$ is the Euclidean plane or "affine plane."
2. $\mathbb{P}^2$ is the projective plane. It consists of $\mathbb{R}^2$ together with "points at infinity," where parallel lines of $\mathbb{R}^2$ intersect.

Note: We call $\mathbb{P}^2$ the "real projective plane," or just "projective plane" for short. Sometimes this is denoted $\mathbb{RP}^2$ or $\mathbb{P}^2(\mathbb{R})$ to emphasize the "real" part, and to distinguish it from other kinds of projective planes. We will not use other kinds of projective planes, but see **??** if you are interested.

4.1 Point-line duality in the Euclidean plane

We're going to do something very weird with notation. We'll denote points in $\mathbb{R}^2$ by $p(a, b)$, where "p" stands for "point." Normally this point would simply be denoted (a, b) without the "p."

We will also have the following notation for lines in $\mathbb{R}^2$:

$$\ell(a, b) = \{p(x, y) : ax + by + 1 = 0\}. \tag{4.1}$$

Figure 4 shows the geometric relationship between $p(a, b)$ and $\ell(a, b)$. We say that $p(a, b)$ and $\ell(a, b)$ are *dual* to each other.

Note the symmetry between (a, b) and (x, y) in the equation $ax + by + 1$. This implies that

$$p(x, y) \in \ell(a, b) \qquad \Longleftrightarrow \qquad \ell(x, y) \ni p(a, b) \tag{4.2}$$

This suggests that there is a symmetry between the roles played by points and lines in $\mathbb{R}^2$.

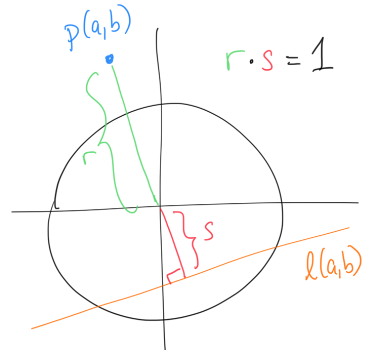


Figure 4: The geometric relationship between $p(a,b)$ and $\ell(a,b)$

For example, (4.2) implies that

$$\begin{gathered}\text{two lines } \ell(a,b) \text{ and } \ell(c,d) \text{ intersect at } p(x,y)\\ \Longleftrightarrow \\ \text{the line going through the two points } p(a,b) \text{ and } p(c,d) \text{ is } \ell(x,y).\end{gathered} \tag{4.3}$$

(There are some issues with (4.3) which we will come back to later.)

If $p(x,y) \in \ell(a,b)$, we say that $p(x,y)$ and $\ell(a,b)$ are *incident* to each other.

Suppose we have theorem about points and lines in the plane that only deals with incidences between points and lines, i.e., no reference to lengths, angles, etc. Then we can form a dual theorem by interchanging points and lines. (Actually, certain ratios of lengths are allowed. These are called *cross ratios*. We will not need them.)

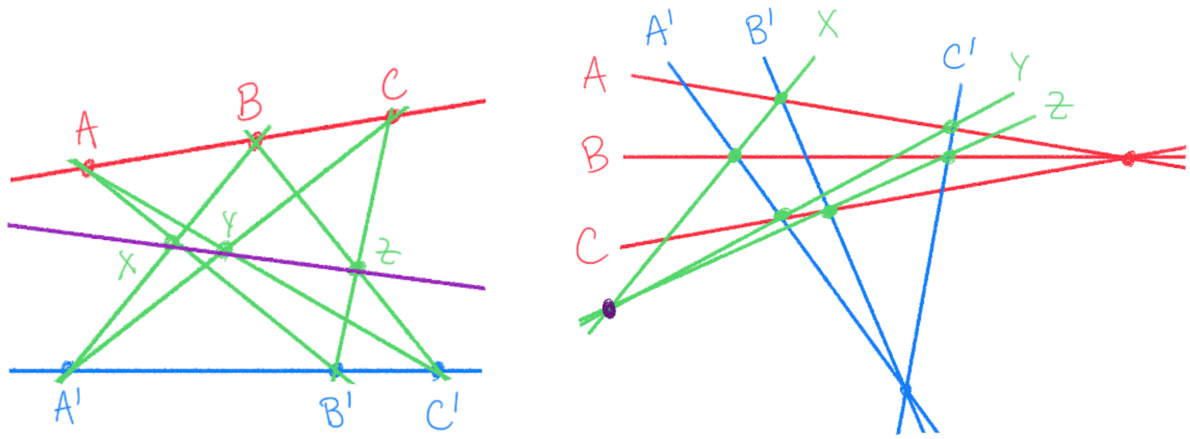
For example, *collinearity* is a property that can be stated in terms of incidences: Three points are collinear if there is a line that all three points are incident to. The dual notion is *concurrency*. Three lines are concurrent if there is a point that all three lines are incident to.

Pappus's theorem is the following theorem about collinear points.

Theorem 4.1 (Pappus's theorem)**.** *(Note: This theorem is not stated with the $p(a,b)$, $\ell(a,b)$ notation.) Suppose that points A, B, C are collinear, and that points A', B', C' are collinear. Let X be the intersection of AB' and A'B, and define Y and Z similarly. (See Figure 5a.) Then the points X, Y, Z are collinear.*

By dualizing Pappus's theorem, we get the following theorem about concurrent lines.

Theorem 4.2 (Dual of Pappus's theorem)**.** *(Note: This theorem is not stated with the $p(a,b)$, $\ell(a,b)$ notation.) Suppose that lines A, B, C are concurrent, and that points A', B', C' are collinear. Let X be the line through of $A \cap B'$ and $A' \cap B$, and define Y and Z similarly. (See Figure 5b.) Then the lines X, Y, Z are concurrent.*



(a) Pappus's theorem

(b) Dual of Pappus's theorem

Figure 5

There is are some issues with the duality between points and lines we described so far.

1. We would like this duality to be a bijection:

$$\{\text{points in } \mathbb{R}^2\} \quad \longleftrightarrow \quad \{\text{lines in } \mathbb{R}^2\} \tag{4.4}$$

   However, there is no way to represent lines through the origin in the form $\ell(a, b)$.

2. Consider again the statement (4.3). Observe that the lines $\ell(a, b)$ and $\ell(c, d)$ are parallel if and only if line through the points $p(a, b)$ and $p(c, d)$ also passes through the origin. In this case, the point $p(x, y)$ in the first half of (4.3) does not exist, and there is no way to represent the line in the second half of (4.3).

3. $p(0, 0)$ makes perfect sense. It is the origin. But its dual $\ell(0, 0)$ makes no sense. (Recall (4.1).)

To fix these kinds of issues, we need to add more points to $\mathbb{R}^2$.

4.2 The real projective plane $\mathbb{P}^2$

Consider the following dual statements.

1. For any two points in $\mathbb{R}^2$, there exists a line which is incident to both. (True!)

2. For any two lines in $\mathbb{R}^2$, there exists a point which is incident to both. (False! Parallel lines don't intersect.)

We are going to define a new space, called the *real projective plane* and denoted $\mathbb{P}^2$, which is better for duality, in that the following are true:

1. For any two points in $\mathbb{P}^2$, there exists a line which is incident to both.

2. For any two lines in $\mathbb{P}^2$, there exists a point which is incident to both.

(Note that some other texts may use notation like $\mathbb{RP}^2$ or $\mathbb{P}^2_{\mathbb{R}}$ to emphasize that it is the *real* projective plane. There are other projective planes.)

To define $\mathbb{P}^2$, we use $\mathbb{R}^2$ as a starting point.

1. Points in $\mathbb{P}^2$: The points in $\mathbb{P}^2$ are the points in $\mathbb{R}^2$, plus some additional points, called *points at infinity.*

2. Lines in $\mathbb{P}^2$: For every line in $\mathbb{R}^2$, add a single point at infinity to it. Do this in a way so that any two parallel lines have the same point at infinity, while any two non-parallel lines have different points at infinity.

   These lines are lines in $\mathbb{P}^2$. Also, $\mathbb{P}^2$ has an additional line, called the *line at infinity*, which is the set of all points at infinity.

So there are many points at infinity. Each line of $\mathbb{R}^2$, gains an additional point at infinity. Which infinite point it gains depends on its slope (or direction) in $\mathbb{R}^2$.

We can see the following are indeed true.

1. For any two points in $\mathbb{P}^2$, there exists a line which is incident to both.

2. For any two lines in $\mathbb{P}^2$, there exists a point which is incident to both.

4.3 Homogeneous coordinates for $\mathbb{P}^2$

Here is one way to denote points in $\mathbb{P}^2$. We start with three real numbers, and write them with brackets and colons: $p[x : y : z]$ denotes a point in $\mathbb{P}^2$.

It seems like we just defined $\mathbb{R}^3$, not $\mathbb{P}^2$. So we make the following rules.

1. We do not allow x, y, z to be zero all at the same time

2. We say that $p[x : y : z] = p[x' : y' : z']$ if there exists a $\lambda \neq 0$ such that $(x, y, z) = \lambda(x', y', z')$.

The second rule is why these are called *homogeneous coordinates.*

A line in $\mathbb{P}^2$ in homogeneous coordinates is

$$\ell[a : b : c] = \{p[x : y : z] \in \mathbb{P}^2 : ax + by + cz = 0\}. \tag{4.5}$$

Observations:

1. If $z \neq 0$, then we can always rescale to make the third coordinate 1: $p[x : y : z] = p[\frac{x}{z} : \frac{y}{z} : 1]$. We identify $p[x : y : 1]$ with the point in the Euclidean plane $p(x, y)$. That is, we have a bijection

$$\begin{array}{rcc} \mathbb{R}^2 & \to & \{p[x : y : z] \in \mathbb{P}^2 : z \neq 0\} \\ p(x, y) & \mapsto & p[x : y : 1] \end{array} \tag{4.6}$$

So, there's no harm in thinking of $\mathbb{R}^2$ as a subset of $\mathbb{P}^2$, which we will do. We sometimes refer to $\mathbb{R}^2$ as the "affine plane," to distinguish it from the "projective plane."

2. The remaining points $\mathbb{P}^2 \setminus \mathbb{R}^2$ are the points at infinity. Any point in $\mathbb{P}^2 \setminus \mathbb{R}^2$ is of the form $p[x : y : 0]$. Recall that a point at infinity corresponds to a direction or slope. We can think of $p[x : y : 0]$ as the direction of the vector (x, y), or as the slope y/x. In other words, every line in $\mathbb{R}^2$ with slope y/x also contains the point $p[x : y : 0]$.

   Here's another way to see this: For $z \neq 0$, $p[x : y : z]$ corresponds to the affine point $p(\frac{x}{z}, \frac{y}{z})$. If we keep x and y fixed and send $z \to 0$, then the point $p(\frac{x}{z}, \frac{y}{z})$ moves to infinity in the direction of the vector (x, y).

3. $\ell[a : b : c]$ really is a line. The equation $ax + by + cz = 0$ might look like the equation for a plane, but recall that we're using homogeneous coordinates. In fact,

$$\ell[a : b : c] \cap \mathbb{R}^2 = \{p(x, y) : ax + by + c = 0\} \tag{4.7}$$

4. If you pick a, b, c in a certain way (Exercise 5.3), you get the line at infinity.

Example 4.3. Consider the lines $\ell([2 : 1 : 0])$ and $\ell([2 : 1 : 1])$ in $\mathbb{P}^2$. We have

$$\ell([2 : 1 : 0]) = \{p[x : y : z] \in \mathbb{P}^2 : 2x + y = 0\} \tag{4.8}$$
$$\ell([2 : 1 : 1]) = \{p[x : y : z] \in \mathbb{P}^2 : 2x + y + z = 0\}. \tag{4.9}$$

To consider the affine part of these two lines, set $z = 1$:

$$\ell([2 : 1 : 0]) \cap \mathbb{R}^2 = \{p(x, y) \in \mathbb{R}^2 : 2x + y = 0\} \tag{4.10}$$
$$\ell([2 : 1 : 1]) \cap \mathbb{R}^2 = \{p(x, y) \in \mathbb{R}^2 : 2x + y + 1 = 0\}. \tag{4.11}$$

so these are two parallel lines with slope -2.

Note that both lines also contain the point at infinity $p[1 : -2 : 0]$. This is where the two lines intersect in $\mathbb{P}^2$.

4.4 The projective sphere

(We didn't cover this section. It's not needed for the remaining classes.)

Here is another way to think about $\mathbb{P}^2$. Given any point $p[x:y:z] \in \mathbb{P}^2$, we can assume WLOG that $x^2+y^2+z^2=1$, i.e., the point (x,y,z) (as a point in $\mathbb{R}^3$ lies in the unit sphere $\mathbb{S}^2$. But there are two ways to do this, because $p[x:y:z] = p[-x:-y:-z]$. These two points are *antipodal*, meaning they are diametrically opposite of each other on the sphere.

A line $\ell[a:b:c]$ is the set of points with $ax+by+cz=0$, which in $\mathbb{R}^3$ is a plane though the origin. It intersects $\mathbb{S}^2$ in a great circle.

So we can think of $\mathbb{P}^2$ in the following way:

1. A point of $\mathbb{P}^2$ is a pair of antipodal points on $\mathbb{S}^2$.

2. A line of $\mathbb{P}^2$ is a great circle on $\mathbb{S}^2$.

The nice thing about using the sphere to understand $\mathbb{P}^2$ is that is very symmetric. There is no need to talk about points at infinity.

The equator of the sphere $\{z=0\} \cap \mathbb{S}^2$ is the line at infinity. If we remove the equator, we can project the remaining points to the plane $\{z=1\} \subset \mathbb{R}^3$ via gnomonic projection. This is how the upper hemisphere is like $\mathbb{R}^2$. And the lower hemisphere as well.

The dual of $p[x:y:z]$ is the intersection of $\mathbb{S}^2$ with the plane that is orthogonal to the vector (x,y,z).

4.5 Projections

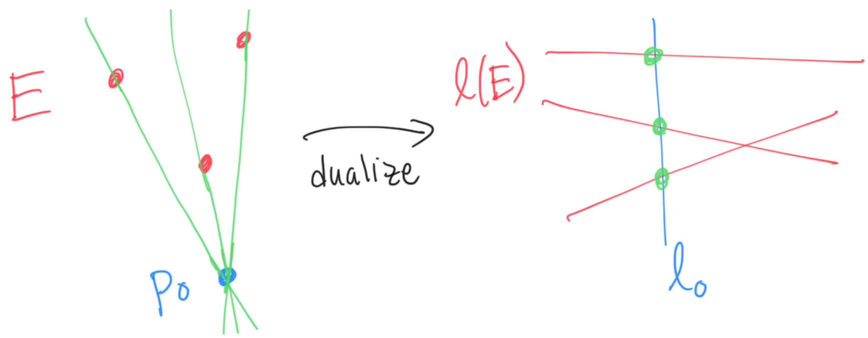
Given a set of points $E \subset \mathbb{P}^2$, we can define the *projection* of E from a point $p_0 \in \mathbb{P}^2$ to be the set of lines through p_0 that intersect E. (Normally, people talk about projection from a point onto a line, but we do not need the "onto a line" part.) See the left halves of Figure 6a and Figure 6b for some examples.

For $E \subset \mathbb{P}^2$, define

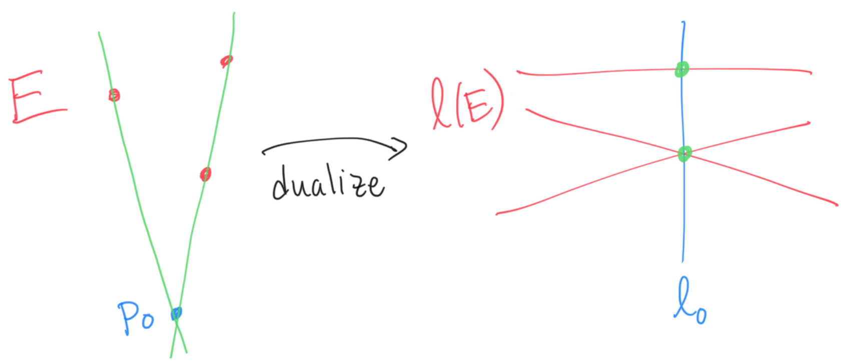
$$\ell(E) = \{\ell[a:b:c] : p[a:b:c] \in E\} \tag{4.12}$$

It turns out that the dual notion of projection of E from p_0 is the intersection of $\ell(E)$ with the line l_0, where ℓ_0, is ℓ_0 is the dual of p_0. See the right halves of Figure 6a and Figure 6b.

Informal discussion: Suppose we're trying to show that the union of the lines in $\ell(E)$ covers very small area. For example, maybe we are interested in Besicovitch sets. Having small area means "most" lines should intersect $\ell(E)$ in a "small" set. But by dualizing, this means that the projections of E from "most" points is "small." So keep this in mind: for a set of lines to cover small area, its dual (a set of points) should have small projections.



(a)



(b)

Figure 6: In both figures, the point p_0 is dual to ℓ_0. The green lines on the left are dual to the green points on the right. (The objects $\ell(E)$ and ℓ_0 are not accurately depicted.)

4.6 For the rest of the week

Projective geometry is a very interesting subject, with many applications. Since we're focusing on applications to the Kakeya problem, here are the important things for the rest of the week.

1. We are only going to apply point-line duality in $\mathbb{R}^2$. But it will be very useful to have some intuitive understanding of points at infinity. We won't need homogeneous coordinates.

2. In Section 4.5, we saw that "the projection from a point" dualizes to "the intersection with a line." This will be very important.

5 Day 2 exercises

Exercise 5.1. ( ) Using the geometric relationship between $p(a, b)$ and $\ell(a, b)$ (i.e., Figure 4), give a geometric proof of (4.2).

Exercise 5.2. () What values of a, b, c make $\ell[a : b : c]$ equal to the line at infinity?

Exercise 5.3. () Let p_0 be a point at infinity. What is the line dual p_0? Describe it geometrically, and compare it with Figure 4.

Exercise 5.4. () (Some linear algebra would help for this problem.)

We can use 3×3 matrices to describe some transformations on $\mathbb{P}^2$. For example consider the matrix

$$A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{5.1}$$

and note that

$$\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 2x \\ 2y \\ z \end{pmatrix} \tag{5.2}$$

So A can be viewed as a function $\mathbb{P}^2 \to \mathbb{P}^2$ given by $p[x : y : z] \mapsto p[2x : 2y : z]$. By setting $z = 1$, we see that the effect on the Euclidean plane is dilation by 2 with respect to the origin. (Also, the points at infinity don't move.)

Here are some geometric transformations $\mathbb{R}^2 \to \mathbb{R}^2$ that can be represented by 3×3 matrices.

1. Can you find a matrix that represents translation by vector (a, b)? (Note that translation in $\mathbb{R}^2$ is not a linear transformation. So it cannot be represented by a 2×2 matrix.)

2. Can you find a matrix that represents rotation by angle θ around the point (a, b)?

3. Can you find two matrices that represent the same transformation $\mathbb{R}^2 \to \mathbb{R}^2$?

Exercise 5.5. () Any invertible 3×3 matrix defines a map $\mathbb{P}^2 \to \mathbb{P}^2$. Such maps are called "projective transformations." Note that differen matrices can correspond to the same projective transformation. Which matrix group is isomorphic to the group of projective transformations?

Exercise 5.6. () We can define the n-dimensional projective space over any field $\mathbb{F}$, denoted $\mathbb{P}^n(\mathbb{F})$, as follows: Start with $\mathbb{F}^{n+1} \setminus \{0\}$. Write the elements in the form $[x_1 : \cdots : x_{n+1}]$. And add the rule that

$$[x_1 : \cdots : x_{n+1}] = [y_1 : \cdots : y_{n+1}] \quad \text{if } \exists \lambda \neq 0 \text{ such that } (x_1, \ldots, x_{n+1}) = \lambda(y_1, \ldots, y_{n+1}). \tag{5.3}$$

Note that what we did in Section 4.3 corresponds to $n = 2$ and $\mathbb{F} = \mathbb{R}$.

1. We know $\mathbb{R}^2$ and $\mathbb{C}$ are the same as topological spaces. (By "same" I mean homeomorphic.) Are $\mathbb{P}^2(\mathbb{R})$ and $\mathbb{P}^1(\mathbb{C})$ the same?

2. Let $\mathbb{F}_q$ be the finite field with q elements. How many points does $\mathbb{P}^n(\mathbb{F}_q)$ have?

6 Day 3: Dimensions of fractals

6.1 Motivating examples

Before we introduce the notions of box dimension and Hausdorff dimension, let's provide some motivating examples. (This section is not rigorous.)

1. If you have a line segment, you can split it into two equal parts. Then you get 2 copies of the original line segment, rescaled by 1/2.

2. If you have a square, you can split it to get 4 copies of the original square rescaled by 1/2.

3. If you have a cube, you can split it to get 8 copies of the original cube rescaled by 1/2.

4. (If you have a point, you get 1 copy of the original point rescaled by 1/2.)

From these four examples we should "expect" the following to hold for the dimension of a set: If you have a self-similar set $E \subset \mathbb{R}^n$, and you can split it to get m copies of E, all rescaled by $r \in (0, 1)$, then the dimension of E should satisfy $m = (1/r)^{\text{dimension of } E}$, i.e.,

$$\text{dimension of } E \;= \frac{\log m}{\log(1/r)} = \log_{1/r} m \tag{6.1}$$

(In Section 6.8, we'll see that this usually holds.) Let's apply this to analyze some fractals.

1. Consider the Sierpiński triangle. You can split it into $m = 3$ copies, all of which are rescaled by $r = 1/3$. So the dimension is $\log_2 3$.

2. Consider the Koch snowflake curve. (Here I don't mean the inside of the snowflake. What I have There are 4 copies rescaled by 1/3. So dimension $\log_3 4$.

3. Middle thirds Cantor set. There are 2 copies rescaled by 1/3. So dimension $\log_2 3$.

This kind of argument is nice, but it only works with self-similar sets. We will try to introduce something that works for all sets.

6.2 Box dimension

For a set $E \subset \mathbb{R}^n$ and $\delta > 0$, let $N_\delta(E)$ be the minimal number of axis-parallel cubes (i.e. "boxes") of side-length δ needed to cover the set.

1. If E is a point, then $N_\delta(E) = 1$.
2. If E is a unit line segment, then $N_\delta(E) = \lceil \frac{1}{\delta} \rceil \approx \frac{1}{\delta}$.
3. If E is a unit square, then $N_\delta(E) = \lceil \frac{1}{\delta} \rceil^2 \approx \frac{1}{\delta^2}$.

The *box dimension* (or *box-counting dimension* or *Minkowski dimension*) of E (denoted $\dim_B E$) is, roughly speaking, the exponent of $1/\delta$ in $N_\delta(E)$. More precisely,

$$\dim_B E = \lim_{\delta \to 0} \frac{\log N_\delta(E)}{\log(1/\delta)} \tag{6.2}$$

With this definition, we see that a point, unit line segment, and unit square have box dimensions 0, 1, 2, respectively.

If the limit in (6.2) doesn't exist for some set E, then $\dim_B E$ is not defined. This does occur, but let's not worry about this. In most nice cases, the limit will exist. Also if E is an unbounded set, then $N_\delta(E) = \infty$. So let's stick to bounded sets.

Also one remark before we look at some examples. Recall that $\dim_B E$ is, roughly speaking, the exponent of $1/\delta$ in $N_\delta(E)$. This means that we don't have to be very precise when calculating $N_\delta(E)$, as the following shows:

Theorem 6.1. *Suppose there exist constants $A, B > 0$ and $s \geq 0$ such that the following is true:*

$$A(1/\delta)^s \leq N_\delta(E) \leq B(1/\delta)^s \qquad \text{for all } \delta > 0. \tag{6.3}$$

Then $\dim_B E = s$.

Proof. Start with (6.3) and take logarithms to get:

$$\log A + s\log(1/\delta) \leq \log N_\delta(E) \leq \log B + s\log(1/\delta) \tag{6.4}$$

so

$$\frac{\log A}{\log(1/\delta)} + s \leq \frac{\log N_\delta(E)}{\log(1/\delta)} \leq \frac{\log B}{\log(1/\delta)} + s \tag{6.5}$$

Now apply the squeeze theorem to conclude $\dim_B E = s$. □

Theorem 6.1 shows that constant multiplicative factors are not important. Let's introduce some useful and standard notation to hide unimportant constant factors.

Definition 6.2. Let f and g be to nonnegative functions. First,

$$f(\delta) \lesssim g(\delta) \qquad \text{means} \qquad \exists A > 0 \text{ s.t. } \forall \delta, f(\delta) \le A\, g(\delta). \tag{6.6}$$

Next, $f(\delta) \gtrsim g(\delta)$ means $g(\delta) \lesssim f(\delta)$

Finally, $f(\delta) \approx g(\delta)$ means $f(\delta) \lesssim g(\delta)$ and $f(\delta) \gtrsim g(\delta)$. In other words,

$$f(\delta) \approx g(\delta) \qquad \text{means} \qquad \exists A, B > 0 \text{ s.t. } \forall \delta > 0, A\, g(\delta) \le f(\delta) \le B\, g(\delta). \tag{6.7}$$

(The notation in Definition 6.2 is closely related to big O notation. Also, in analytic number theory, people tend to use $\ll$ instead of $\lesssim$.)

Now we can restate Theorem 6.1 as: If $N_\delta(E) \approx (1/\delta)^s$, then $\dim_B E = s$.

6.3 Example: The middle third Cantor set

Consider the middle third Cantor set. Start with a unit line segment to get a set E_0. Remove the middle third to get E_1, the union of two intervals of length $1/3$. Repeat this way to get $E_2, E_3, \ldots$. Note that E_n consists of 2^n intervals of length 3^{-n}. The middle third Cantor set is $E = \bigcap_{n=0}^{\infty} E_n$.

We already saw in Section 6.1 that E "should" have dimension $\log_2 3$. Let's prove it for the box dimension.

Theorem 6.3. *The middle third Cantor set has box dimension* $\log_2 3$.

Proof. Let's calculate $N_\delta(E)$. To motivate the more general argument to follow, suppose first that $\delta = 3^{-n}$. That is, we want to count how many intervals of size 3^{-n} we need to cover E. Since $E \subset E_n$, we have $N_\delta(E) \le N_\delta(E_n)$. Also note that $N_\delta(E_n) = 2^n$. To get a lower bound for $N_\delta(E)$, note that each "box" (i.e., interval) of size δ can intersect at most one interval of E_n. Thus, $N_\delta(E) \ge 2^n$. So just showed

$$N_{3^{-n}}(E) = 2^n \tag{6.8}$$

Now we consider general values of δ. Let an n be the integer satisfying $3^{-n} \le \delta \le 3^{-n+1}$. Then by the same reasoning as before, $N_\delta(E) \le N_\delta(E_n) \le 2^n$. The argument for the lower bound changes slightly. Each interval of size δ can intersect at most one interval of E_{n-1}. There are 2^{n-1} intervals. So $N_\delta(E_n) \ge 2^{n-1}$. We just showed

$$\frac{1}{2} \cdot 2^n \le N_\delta(E) \le 2^n \tag{6.9}$$

Also by how n was chosen, we have

$$\frac{1}{3} \cdot 3^n \le 1/\delta \le 3^n \tag{6.10}$$

By the right half of (6.9) and the left half of (6.10), we have

$$N_\delta(E) \le 2^n = (3^n)^{\log_3 2} \le (3/\delta)^{\log_3 2} = 3^{\log_3 2}(1/\delta)^{\log_3 2} = 2(1/\delta)^{\log_3 2} \tag{6.11}$$

Similarly, by the other two halves,

$$N_\delta(E) \ge \frac{1}{2} \ge \frac{1}{2}(1/\delta)^{\log_3 2}. \tag{6.12}$$

So we have shown

$$\frac{1}{2} \cdot (1/\delta)^{\log_3 2} \le N_\delta(E) \le 2 \cdot (1/\delta)^{\log_3 2}. \tag{6.13}$$

The theorem now follows from Theorem 6.1. □

Protip 6.4. Using the asymptotic notation of Definition 6.2, (6.9) and (6.10) could be rewritten as

$$N_\delta(E) \approx 2^n \tag{6.14}$$

$$1/\delta \approx 3^n \tag{6.15}$$

Then the remaining calculations in the proof above could be summarized the following chain:

$$N_\delta(E) \approx 2^n = (3^n)^{\log_2 3} \approx (1/\delta)^{\log_2 3} \tag{6.16}$$

The nice thing about this approach is that you don't need to deal with all the constant multiplicative factors that we don't care about anyways. Doing calculations in this way may seem confusing at first, but after some practice, it will become natural. This is an essential part of the training to become an Analysis Master™.

6.4 Box dimension as a way to measure "size" of a set

Let's consider sets in the plane.

Theorem 6.5. *If a bounded set $E \subset \mathbb{R}^2$ has positive area, then* $\dim_B E = 2$.

Proof. Suppose E has area $A > 0$. Let $\delta > 0$. Suppose we try to cover E by squares of side δ. Each square has area δ^2. So we need at least A/δ^2 many squares to cover A, i.e., $N_\delta(E) \ge A/\delta^2$. This implies $\dim_B E = 2$. □

The theorem works in higher dimensions as well:

Theorem 6.6. *If a bounded set $E \subset \mathbb{R}^n$ has positive n-dimensional volume, then* $\dim_B E = n$.

Proof. Proof is the same as Theorem 6.5. □

The contrapositive of Theorem 6.6 is: If a bounded set $E \subset \mathbb{R}^n$ has box dimension $< n$, then it has zero n-dimensional volume. This shows that box dimension is a useful notion of "size" of a set in $\mathbb{R}^n$ for sets that have zero volume.

6.5 Flexibility in the definition of $N_\delta(E)$

(We didn't cover this section. It's not needed for the remaining classes.)

For $E \subset \mathbb{R}^2$, let $N_\delta^\circ(E)$ be the number of disks of radius δ needed to cover E.

1. If we have a cover of E by squares of side δ, then we can cover each square by a disk of radius δ. This implies that $N_\delta^\circ(E) \le N_\delta(E)$.

2. If we have a cover of E by disks of radii δ, then we can cover each disk by 4 squares of side δ. This implies that $N_\delta(E) \le 4N_\delta^\circ(E)$.

Using notation the $\approx$ notation from before

$$N_\delta(E) \approx N_\delta^\circ(E). \tag{6.17}$$

So we could have also defined the box dimension as

$$\lim_{\delta\to 0} \frac{\log N_\delta^\circ(E)}{\log(1/\delta)} = \lim_{\delta\to 0} \frac{\log N_\delta(E)}{\log(1/\delta)} \tag{6.18}$$

which means we could have also defined $\dim_B E$ as the LHS of (6.18). This argument works in higher dimensions. (We need different constants in (6.17), but that doesn't matter.)

6.6 Lebesgue measure zero and Hausdorff dimension

(We didn't cover this section. It's not needed for the remaining classes.)

"Measure zero" is a basic concept from measure theory. We can define it without having to define measure. (Measures are hard to define, and we don't need them.)

Definition 6.7. Let $n \in \mathbb{N}$. A set $E \subset \mathbb{R}^n$ has *Lebesgue measure zero* (or *n-dimensional Lebesgue measure zero*, if we want to be precise) if the following statement is true:

$$\forall \varepsilon > 0, \exists \text{ balls } \{B(x_i, r_i)\}_{i=1}^\infty \text{ such that } E \subset \bigcup_i B(x_i, r_i) \text{ and } \sum_i r_i^n < \varepsilon. \tag{6.19}$$

Roughly speaking, E has Lebesgue measure zero if it can be covered by balls which have arbitrarily small total volume. In fact, when I say a set has "zero volume," what I really mean is that it has Lebesgue measure zero.

If we change the exponent n at the end of (6.19) to a different number, we get a closely related notion:

Definition 6.8. Let $n \in \mathbb{N}$. Let $s \in \mathbb{R}_{\ge 0}$. A set $E \subset \mathbb{R}^n$ has *$\mathcal{H}^s$-measure zero* (or *s-dimensional Hausdorff measure zero*) if the following statement is true:

$$\forall \varepsilon > 0, \exists \text{ balls } \{B(x_i, r_i)\}_{i=1}^\infty \text{ such that } E \subset \bigcup_i B(x_i, r_i) \text{ and } \sum_i r_i^s < \varepsilon. \tag{6.20}$$

(Note that in $\mathbb{R}^n$, Lebesgue measure zero and $\mathcal{H}^n$-measure zero are the same thing.)

In both of these definitions, the goal is to find an "efficient" cover of E by balls. Note that if we're trying to make $\sum_i r_i^s$ very small, then each r_i should be very small. The smaller the exponent s is, the harder our task is.

Basic properties:

Example 6.9. Let $E \subset \mathbb{R}^2$ be a unit line segment.

First let's show that E has $\mathcal{H}^s$-measure zero when $s > 1$. For any $M \in \mathbb{N}$, we can cover E by M balls of radius $1/M$. (Well, we could have done it with $M/2$ balls, but constant factors won't matter.) For these balls

$$\sum_{i=1}^{M} r_i^s = \sum_{i=1}^{M} (1/M)^s = M(1/M)^s = M^{1-s} \tag{6.21}$$

Since $s > 1$, we have $\lim_{M\to\infty} M^{1-s} = 0$. This completes the proof that E has $\mathcal{H}^s$-measure zero when $s > 1$.

Next, let's show that E does not have $\mathcal{H}^s$-measure zero when $s \leq 1$. We need to show that any covering of E with balls is inefficient, i.e., $\sum_i r_i^s$ is large. Let $\{B(x_i, r_i)\}_i$ be any covering of E. Since each ball can only cover a length $2r_i$ of E, and E has length 1, we must have $\sum_i (2r_i) \geq 1$.

We may assume WLOG that $r_i \leq 1$ for all i. (Think about why.) Then since $s \leq 1$, $r_i^s \geq r_i$, so $\sum r_i^s \geq \sum r_i \geq \frac{1}{2}$. So E does not have $\mathcal{H}^s$-measure zero when $s \leq 1$.

The example suggests the following way to define a dimension:

Definition 6.10. Let $E \subset \mathbb{R}^n$. Then the *Hausdorff dimension* of E is

$$\dim_H E = \inf\{s \geq 0 : E \text{ has } \mathcal{H}^s\text{-measure zero}\}. \tag{6.22}$$

Some basic properties:

Theorem 6.11. *Let $s < t$. If $E \subset \mathbb{R}^n$ has $\mathcal{H}^t$-measure zero, then it has $\mathcal{H}^s$-measure zero.*

Proof. Kind of similar to the $r_i^s \geq r_i$ argument at the end of Example 6.9. □

Theorem 6.12. *If $E \subset \mathbb{R}^n$ has positive volume, then it does not have Lebesgue measure zero.*

Proof. Kind of similar to Theorem 6.5. □

6.7 Box vs Hausdorff dimension

(We didn't cover this section. It's not needed for the remaining classes.)

In box dimension, we look at one scale at a time. We cover by boxes (or balls) of the same size.

In Hausdorff dimension, we look at multiple scales at the same time. We're not restricted to using balls of the same size. It is easier to get an "efficient" cover of balls if we can use different sizes. As a result, $\dim_H E \le \dim_B E$. (This requires proof, but intuitively it's true.)

6.8 Dimension of self-similar sets

(We didn't cover this section. It's not needed for the remaining classes.)

Here we explain that informal discussion in Section 6.1 does in fact work in most cases. In particular, the formula (6.1) can be used to compute the dimension of many self-similar sets.

Let $0 < r < 1$ and $m \in \mathbb{N}$. Suppose E is a self similar set made up of m copies of itself scaled by r. Suppose something called the "open set condition" is also satisfied. (Most nice fractals you can think of will satisfy this condition.) Then $\dim_H E = \dim_B E = \frac{\log m}{\log(1/r)}$.

Here is the precise statement. The open set condition is (6.24), below. It roughly says that the copies in the self similar set do not overlap too much.

Theorem 6.13. *Let $0 < r < 1$ and $m \in \mathbb{N}$. For $i = 1, \ldots, m$, let $\phi_i : \mathbb{R}^n \to \mathbb{R}^n$ be a similarity transformation with scaling factor r. Then*

$$\exists \text{ a unique non-empty compact set } E \text{ such that } \bigcup_{i=1}^{m} \phi_i(E) = E. \tag{6.23}$$

Suppose furthermore that

$$\exists \text{ a non-empty bounded open set } V \subset \mathbb{R}^n \text{ such that } \bigcup_{i=1}^{m} \phi_i(V) \subset V \tag{6.24}$$

Then $\dim_H E = \dim_B E = \frac{\log m}{\log(1/r)}$.

Proof. We won't prove the theorem. One reference is Kenneth Falconer's *Fractal Geometry*, Theorem 9.3 □

The case of the middle third Cantor set corresponds to $n = 1$, $m = 2$, $r = 1/3$, and the two similarity transformations are $\phi_1(x) = \frac{1}{3}x$ and $\phi_2(x) = \frac{1}{3}x + \frac{2}{3}$.

Fun fact 6.14. A collection of maps ϕ_i like the one in Theorem 6.13 is called an *iterated function system.* These are useful in the study of dynamical systems.

6.9 Some random stuff

https://en.wikipedia.org/wiki/List_of_fractals_by_Hausdorff_dimension

https://www.youtube.com/watch?v=gB9n2gHsHN4

7 Day 3 exercises

Exercise 7.1. () Think of other self-simliar sets and use the informal discussion in Section 6.1 to determine what their "dimensions" should be. (Most likely, these numbers will be correct, because of Theorem 6.13.)

Exercise 7.2. () Here are other kinds of Cantor sets. Let $0 < p < 1$. Start with the unit interval. Remove an interval of length $1 - 2p$ from the middle, so that you have two intervals of length p on either side. Repeat this so that you have 4 intervals of length p^2, and so on. What is the box dimension of this set? (The middle third Cantor set corresponds to $p = 1/3$.)

Exercise 7.3. () Let $S^1 \subset \mathbb{R}^2$ denote the unit circle: $S^1 = \{(x, y) \in \mathbb{R}^2 : x^2 + y^2 = 1\}$. Find upper and lower bounds on $N_\delta(S^1)$ and use these to determine $\dim_B S^1$.

Exercise 7.4. () Suppose you have a self-similar set $E \subset \mathbb{R}^n$, and you can split it to get m copies of E, all rescaled by $r \in (0, 1)$. Show that for all $\delta > 0$, $N_{r\delta}(E) \leq mN_\delta(E)$. This is one step in the proof of Theorem 6.13.

Exercise 7.5. () Can you find an example of a bounded set E for which $\dim_B E$ is undefined? What you need to do is make $\frac{\log N_\delta(E)}{\log(1/\delta)}$ oscillate as $\delta \to 0$.

Exercise 7.6. () Can you figure out the similarity transformations used in Sierpiński triangle? Or the Koch snowflake? Or your favorite self-similar set? (By similarity transformation, I mean the functions ϕ_i in Theorem 6.13.)

8 Day 4: Putting things together

8.1 Besicovitch sets and the Kakeya conjecture

In Section 2, we solved the Kakeya needle problem: We showed that we can rotate a needle in arbitrarily small area. What's left? Zero area? (To define "zero area" rigorously, see Definition 6.7. We won't need this rigorous definition.)

It turns out a continuous rotation in zero area is not possible, by an intermediate value theorem-type argument; see `https://terrytao.wordpress.com/2008/12/31/a-remark-on-the-kakeya-`
However, if we allow the needle to "teleport from one position to another uncountably many times," then it is possible.

Theorem 8.1 (Besicovitch)**.** *There exists a set in in the plane with zero area, and which contains a unit line segment pointing in every direction.*

Here's one way to prove Theorem 8.1: The idea is, roughly speaking, to take a limit of the sliding triangle construction above as $n \to \infty$. To do this rigorously, we need to consider the metric space of compact subsets of $\mathbb{R}^2$ with the Hausdorff metric. (I will not go into the details.) We'll see another way using point-line duality.

Sets in $\mathbb{R}^n$ which have a unit line segment in every direction are called *Besicovitch sets*. Theorem 8.1 states that in $\mathbb{R}^2$, such sets can have zero area. In fact, this implies that in higher dimensions, Besicovitch sets can also have zero n-dimensional volume. See Exercise 9.2.

As we described yesterday, the box dimension is a way of comparing sizes of sets that have zero volume. The Kakeya conjecture is, roughly speaking, the statement that Besicovitch sets in $\mathbb{R}^n$ cannot be much smaller than zero volume.

Conjecture 8.2 (Kakeya conjecture)**.** *Every Besicovitch set in $\mathbb{R}^n$ has box dimension and Hausdorff dimension equal to n.*

(The way I stated the conjecture is kind of redundant since the Hausdorff dimension is less than or equal to the box dimension.)

Conjecture 8.2 is known to be true in 2 dimensions. But we don't know for dimensions ≥ 3.

Conjecture 8.2 has connections to questions in many different fields, including harmonic analysis, PDEs, analytic number theory, additive combinatorics, etc. If you solve it, you will be given offers to join the math department at many famous universities. And probably win the Fields Medal.

8.2 Point-line duality, revisited

(Warning: the notation used here is different from the notation in Section 4.)

Given $(a, b) \in \mathbb{R}^2$, define $\ell(a, b) = \{(x, y) : y = ax + b\}$. Note this is different from the duality from Section 4. The reason for doing this is that we're interested in lines of many different directions, and the parameter a is precisely the direction. However, this is not symmetric between (a, b) and (x, y):

$$(a, b) \in \ell(x, y) \not\Leftrightarrow (x, y) \in \ell(a, b) \tag{8.1}$$

To address this lack of symmetry, we're going to consider two copies of $\mathbb{R}^2$, which we'll refer to as the *primal* and *dual*. We will denote them by $\mathbb{R}^2_{\text{primal}}$ and $\mathbb{R}^2_{\text{dual}}$. Given a point in $\mathbb{R}^2_{\text{primal}}$, we define its dual line in $\mathbb{R}^2_{\text{dual}}$ as above:

$$(a, b) \in \mathbb{R}^2_{\text{primal}} \qquad \longrightarrow \qquad \ell(a, b) = \{(x, y) : y = ax + b\} \subset \mathbb{R}^2_{\text{dual}}. \tag{8.2}$$

If we have a point $(x, y) \in \mathbb{R}^2_{\text{dual}}$, then its "primal line" should also be defined by $y = ax + b$. But now, x and y are fixed, and this is a line in the ab-plane. That is,

$$\tilde{\ell}(x, y) = \{(a, b) : b = (-x)a + y\} \subset \mathbb{R}^2_{\text{primal}} \qquad \longleftarrow \qquad (x, y) \in \mathbb{R}^2_{\text{dual}}. \tag{8.3}$$

(I'm trying to keep the primal on the left and the dual on the right.) So the point (x, y) "un-dualizes" to become the line in the ab-plane with slope $-x$ and b-intercept y.)

With these definitions, it is true that

$$(a, b) \in \tilde{\ell}(x, y) \iff (x, y) \in \ell(a, b). \tag{8.4}$$

Now let's consider vertical lines in $\mathbb{R}^2_{\text{dual}}$. The primal of the line $\ell_t = \{(x, y) : x = t\}$ is a point; let's call it p_t. What is p_t? If we look at (8.2), we see that $\ell(a, b)$ on produces non-vertical lines. So p_t must be a point at infinity.

In particular, p_t is the point at infinity $[1 : -t : 0]$, that is, the one corresponding to the direction vector $(1, -t)$ or slope $-t$. There are several ways to see this.

1. The point $(t, 0) \in \ell_t$ is a statement in $\mathbb{R}^2_{\text{dual}}$. The primal statement is $\tilde{\ell}(t, 0) \ni p_t$. We know $\tilde{\ell}(t, 0)$ is a line with slope $-t$ and that p_t is a point at infinity. So p_t must be the point at infinity contained in all lines with slope $-t$. That is, it is the point at infinity of direction $(1, -t)$.

2. Consider $(a, b) = (r, -tr) \in \mathbb{R}^2_{\text{primal}}$, where r is very large. The dual to this is the line $y = rx - tr$, or $x = t + \frac{y}{r}$. If we send $r \to \infty$, this becomes $x = t$, as desired.

See Figure 7 for a summary of this section.

$\mathbb{R}^2_{\text{primal}}$	$\mathbb{R}^2_{\text{dual}}$
(a, b)	$\ell(a, b) = \{(x, y) : y = ax + b\}$
$\tilde{\ell}(x, y) = \{(a, b) : b = (-x)a + y\}$	(x, y)
$p_t =$ point at infinity in direction $(1, -t)$	vertical line $\ell_t = \{(x, y) : x = t\}$

Figure 7: Summary of the correspondence between $\mathbb{R}^2_{\text{primal}}$ and $\mathbb{R}^2_{\text{dual}}$

8.3 The area of a set of line segments

Given $E \subset \mathbb{R}^2_{\text{primal}}$, we define $L(E) = \bigcup_{(a,b)\in E} \ell(a,b)$. Let S be the vertical strip $\{(x,y) : 0 \leq x \leq 1\}$. We are interested in the area of $L(E) \cap S$. By splitting into cross sections,

$$\text{area of } L(E) \cap S = \int_0^1 (\text{length of } L(E) \cap \ell_t)\, dt \tag{8.5}$$

(Recall $\ell_t = \{x = t\}$.) In particular, if for all $t \in [0,1]$, the length of $L(E) \cap \ell_t$ is $\leq \varepsilon$, then the area of of $L(E) \cap S$ will be $\leq \varepsilon$.

The object $L(E) \cap \ell_t$ is something in $\mathbb{R}^2_{\text{dual}}$. What is the corresponding thing in the primal $\mathbb{R}^2_{\text{primal}}$? We know that the primal of $L(E)$ is E itself. As discussed in Section 4.5, the intersection $L(E) \cap \ell_t$ un-dualizes to become the projection of E from the point p_t. Since p_t is a point at infinity, we are interested in how many lines of slope $-t$ intersect E. We can just think of this as an orthogonal projection.

Remark 8.3. Actually, the intersection $L(E) \cap \ell_t$ is not congruent to the orthogonal projection in direction $(1, -t)$. If you work out the relationship with some equations, you will see that they are similar to each other, with a scaling factor of $\sqrt{1+t^2}$. Since we only care about $t \in [0,1]$, this scaling factor is between 1 and 2, so it doesn't affect anything.

8.4 Sliding triangles, revisited

(We're going to rotate our sliding triangles setup 90 degrees, so that triangles slide vertically instead of horizontally.)

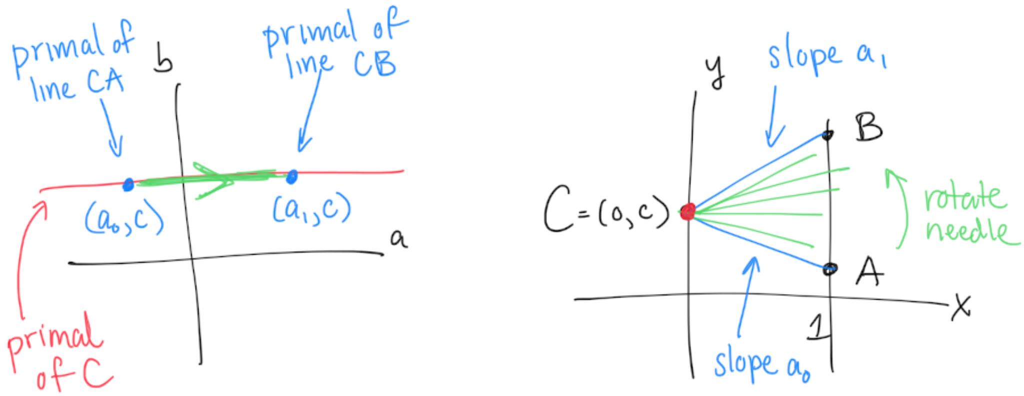
Suppose we have a triangle ABC, with C on the y-axis, and A and B on the vertical line $x = 1$ with A below B. Let a_0 be the slope of CA and let a_1 be the slope of CB. (So $a_0 < a_1$.) And suppose the point C is $(0, c)$.

Rotation of the needle from CA to CB can be thought of as something happening in the dual. In the primal, it is moving along the line segment from (a_0, c) to (a_1, c). This makes sense, because the primal of the point C is the horizontal line with b-coordinate equal to c. See Figure 8.

Sanity check: Intersection of ABC with vertical line ℓ_t ($t \in [0,1]$) corresponds to projecting the line segment in direction $(1, -t)$. For example when $t = 0$, this is projecting horizontally, which only gives one point. As we increase t to 1 we get larger and larger projections. See Figure 9.

Now let's cut triangle ABC in half. The median from C to AB has slope $\frac{a_0+a_1}{2}$. So in the primal, we're just dividing our line segment in half. The left half of the segment (from a_0 to $\frac{a_0+a_1}{2}$) corresponds to the lower half of the triangle ABC. See Figure 10.

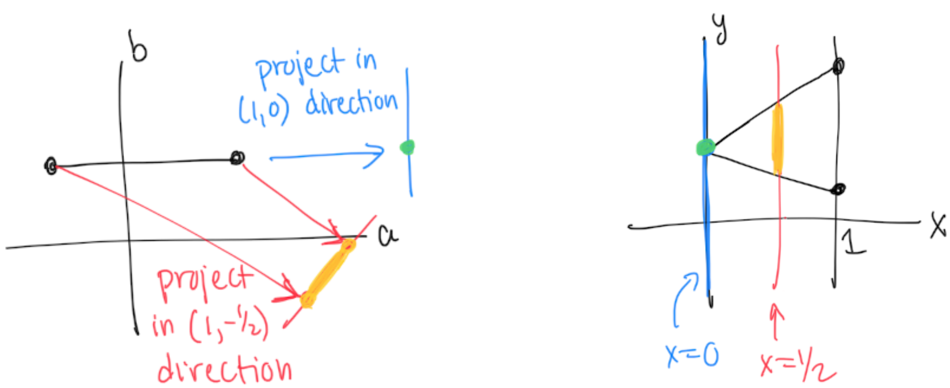
Now we vertically translate one of the two halves. Let's move the left-half upwards. That



(a) primal

(b) dual

Figure 8: The green line segment in the primal corresponds to rotating the needle in the dual.



(a) primal

(b) dual

Figure 9: Projection in the $(1, -t)$ direction corresponds to intersection with the vertical line $x = t$. This is illustrated for $t = 0$ and $t = 1/2$.

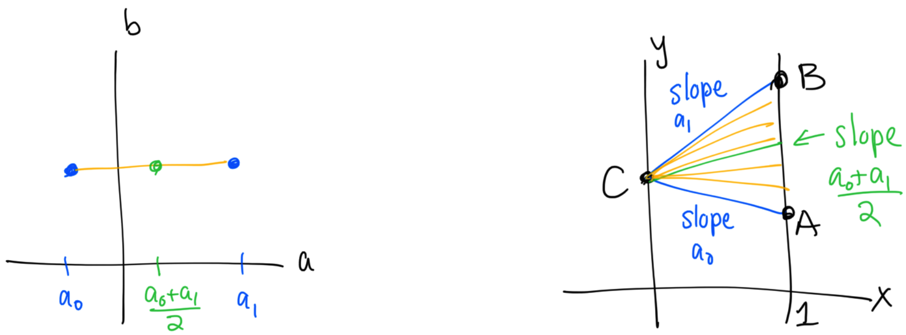
preserves all the slopes, but it changes the y-intercept. So in the primal, that is translating the left half of the line segment upwards. See Figure 11.

The vertical discontinuity in the line segment corresponds to "teleportation."

Note that the projections in some directions that we care about have gotten smaller thanks to overlap. This corresponds to the vertical slices in the dual having smaller total length thanks to overlap.

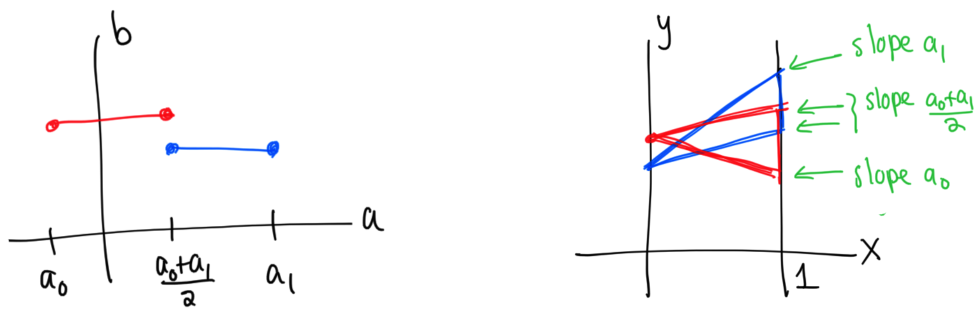
Another sanity check. Let's compare some projections with some vertical slices. See Figure 12.

The idea for sliding m triangles is the same. To cut ABC into m triangles, just divide our line segment into m pieces. Then we can translate each of the m pieces vertically. And the



(a) primal (b) dual

Figure 10: Dividing the line segment in half in the primal corresponds to dividing the triangle in half in the dual.



(a) primal (b) dual

Figure 11: Translating the red segment upwards in the primal corresponds to translating the red triangle up in the dual.

goal is for the projection to be small for any direction between $(1, 0)$ (horizontal direction) and $(1, -1)$ (45 degrees direction).

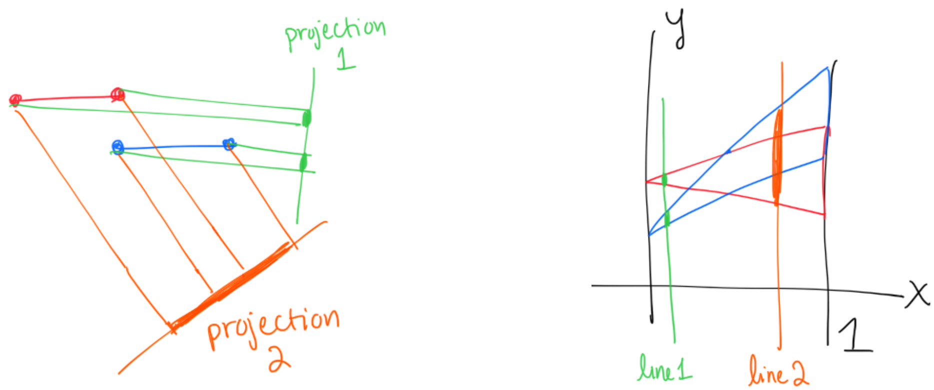
Back in Section 2, we showed the following.

Theorem 8.4. *Start with an equilateral triangle* ABC *of height* 1*, with* AB *vertical. For any* $\varepsilon > 0$*, we can divide this triangle (by drawing lines from* C *to* AB*) and translate each piece vertically so that the resulting figure has area* $\leq \varepsilon$*.*

By duality, Theorem 8.4 is essentially the same as the following theorem.

Theorem 8.5. *Start with the line segment connecting* $(-1, 0)$ *and* $(1, 0)$*. For any* $\varepsilon > 0$*, we can divide this segment into subsegments and translate each piece vertically so that for any* $t \in [0, 1]$*, the orthogonal projection of our segments in direction* $(1, -t)$ *has length* $\leq \varepsilon$*.*

Here's why I said "essentially the same" instead of "the same": The line segment in Theorem 8.5 corresponds to a right triangle in the dual instead of an equilateral triangle,



(a) primal (b) dual

Figure 12: The projections on the left correspond to the vertical slices on the right. Notice that the green projection is made up of two small intervals, while the orange projection is a single interval. The same is true for the corresponding slices on the right.

but that's not important. Also, Theorem 8.5 implies Theorem 8.4, but the converse is not quite true. A set has small area if all of its vertical cross sections have small length. But the converse is not true.)

Theorem 8.5 is true, but we're not going to prove it.

If we return to our iterative procedure from Section 2, we can try to un-dualize this to show that the answer to Theorem 8.5 is "yes." But let's not return to that iterative procedure from before. There's no reason to use that exact construction. We have more flexibility. In fact, we can modify Theorem 8.5 slightly, which we do tomorrow (see also Exercise 9.4).

9 Day 4 exercises

Exercise 9.1. ( ) Very important: Make sure you understand the primal-dual correspondence for the sliding triangles construction. We will need it for tomorrow.

Exercise 9.2. ( ) Using Theorem 8.1, show that there exist Besicovitch sets in $\mathbb{R}^3$ with zero volume. (Does the argument generalize to higher dimensions?)

Exercise 9.3. ( ) When we un-dualized the sliding triangle construction, we ended up with a bunch of horizontal line segments. What would a non-horizontal line segment mean?

Exercise 9.4. ( ) (This is a fun problem. We'll go over one way to answer this tomorrow.) Can you find a way to construct a set $E \subset \mathbb{R}^2$ with the following properties?

1. E is made up of a finite union of closed line segments.

2. The orthogonal projection of E onto the horizontal axis (i.e., projection in the vertical direction) contains the interval $[-1, 1]$.

3. For every $t \in [0, 1]$, the orthogonal projection of E in direction $(1, -t)$ is small. (I'm not being very precise by "small" here. Try to get it as small as you can.)

Hint: See https://en.wikipedia.org/wiki/Window_blind#Types for some inspiration. (A projection is like a shadow that the sun casts on your set. You want something with large shadows when the sun is directly on top, but small shadows when the sun is in other positions.)

10 Day 5

10.1 A slightly modified approach

Let's say we want to rotate the needle from slope -1 to slope 1. We are allowed to teleport it finitely many times. We also want to keep it in the strip $\{0 \le x \le 1\}$. And we want it to cover small area.

A non-vertical line segment in $\mathbb{R}^2_{\text{primal}}$ corresponds to a rotation of a line with a point kept fixed. In the examples described above, the fixed point in the rotation was always on the y-axis $x = 0$. That is why we had horizontal line segments in $\mathbb{R}^2_{\text{primal}}$. But there's no reason to also rotate around points on the y-axis. We can allow line segments in other directions.

We will show the following

Theorem 10.1. *Let $\varepsilon > 0$. We can find a set $E \subset \mathbb{R}$ with the following properties.*

1. *E is made up of a finite union of closed line segments.*

2. *The orthogonal projection of E onto the horizontal axis (i.e., projection in vertical direction) contains the interval $[-1, 1]$.*

3. *For every $t \in [0, 1]$, the orthogonal projection of E in direction $(1, -t)$ has length $\le \varepsilon$.*

Moving along a line segment is like rotating the needle. Vertical teleportation (by changing the b-coordinate) is like translating the needle without rotating it. The second condition in Theorem 10.1 is to make sure that we can indeed start at $a = -1$ and make it to $a = 1$ by moving along line segments and teleporting vertically.

One way to show that the answer to Theorem 10.1 is "yes" is to use the *iterated Venetian blinds construction*. See Figure 13. Here is a description in words (which may be hard to understand... just look at the picture). Start with E_0 the horizontal line segment from $(-1, 0)$ to $(1, 0)$. Next, choose an angle θ_1 and a number $N_1 \in \mathbb{N}$, and replace E_0 with N_1

segments of equal length such that each segment has one endpoint in E_0, and makes angle θ_1 with E_0. Call this new set E_1. Now choose θ_2 and N_2. For each segment S in E_1, replace it with N_2 segments, each with an endpoint in S, and makes angle θ_2 with S.

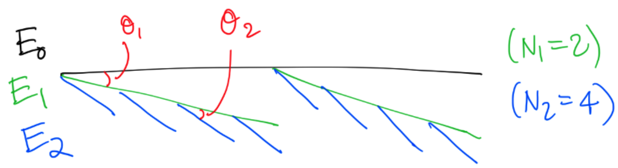


Figure 13: Iterated Venetian blinds

Note that by construction, the first two conditions of Theorem 10.1 are automatically satisfied. If you choose the θ_i very small and the N_i very large, then after finitely many steps, you can get something which also satisfies the third condition. Let's explain this in more detail.

Sketch of proof of Theorem 10.1. Fix $\varepsilon > 0$. In what follows, when we say a set has "small projection" in some direction, we mean that the length of the orthogonal projection is strictly less than ε.

Start with E_0 the horizontal line segment from $(-1, 0)$ to $(1, 0)$. Let $\theta_1 > 0$ be the angle such that the orthogonal projection of E_0 in direction $\theta_1 > 0$ has length $\varepsilon/2$. (Here, direction θ means direction $(\cos\theta, -\sin\theta)$.

Then the projection of E_0 is small for all directions in $[0, \theta_1]$. In fact, we can take a small open neighborhood $U_0 \supset E_0$ such that the projection of U_0 is small for all directions in $[0, \theta_1]$. Pick N_1 so large that the resulting Venetian blind $E_1 \subset U_0$. (Recall that E_1 is determined by E_0, θ_1, and N_1.) Since $E_1 \subset U_0$, we know that the projection of E_1 is small for all directions in $[0, \theta_1]$

In fact, $P_{\theta_1}(E_1)$ is made up of N_1 points, so it has length 0. As before, let $\theta_2 > 0$ be the angle such that the orthogonal projection of E_2 in direction $\theta_1 + \theta_2$ has length $\varepsilon/2$. Let $U_1 \supset E_1$ be an open neighborhood of E_1 such that the projection of U_1 is small for all directions in $[0, \theta_1]$. Then pick N_2 so that $E_2 \subset U_1$.

Repeat for $E_3, E_4, \ldots$. This doesn't go on forever, because of the following.

1. You're done when $\theta_1 + \cdots + \theta_n = \pi/4$ (since this means you reached slope -1).

2. For all i, we have $\theta_i \geq \frac{1}{100}\varepsilon$. This is because of how θ_i was chosen. (The number $\frac{1}{100}$ is not important. What's important is that it is some absolute constant. I just did a quick calculation and $\frac{1}{4\sqrt{2}}$ works too. You might even be able to do better.)

Once the process terminates, we're done! $\square$

This iterated Venetian blind construction gives us another way to construct Kakeya sets!

10.2 Construction of a Besicovitch set

We can do the iterated Venetian blind construction without stopping. And we can go around and around and around, i.e. $\sum_{i=1}^{\infty} \theta_i = \infty$. If you check the details carefully, you can get the following:

Theorem 10.2. *Let $\varepsilon > 0$. We can find a set $E \subset \mathbb{R}^2$ with the following properties.*

1. *The orthogonal projection of E onto the horizontal axis (i.e., projection in direction $(0, 1)$) contains the interval $[-1, 1]$.*

2. *For every $t \in \mathbb{R}$, the orthogonal projection of E in direction $(1, t)$ has zero length.*

In other words, E projected onto the a-axis contains the interval $[-1, 1]$. But when projected in any other direction, it has zero length! From this, it follows that every vertical cross section of $L(E)$ (the dual of E) has length zero, and therefore $L(E)$ has area zero! So we just showed that $L(E)$ has area zero, and for every $a \in [-1, 1]$, there is a line in $L(E)$ of slope a. This implies the following (by making a rotated copy):

Corollary 10.3. *There exists a set in $\mathbb{R}^2$ of area zero that contains a full line in every direction.*

Fun fact 10.4. There's something more general than Theorem 10.2 called the digital sundial theorem. The proof is by iterated Venetian blinds as well. See https://en.wikipedia.org/wiki/Digital_sundial#Fractal_sundial

10.3 Construction of another Besicovitch set

Let's use some very powerful theorems to prove some powerful results.

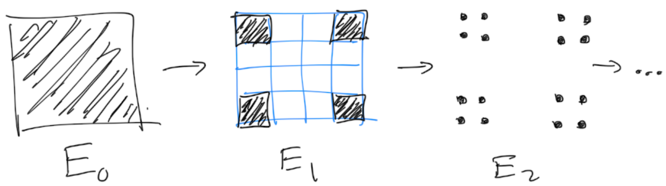
Consider the four-corner Cantor set, the fractal generated by the process shown in Figure 14a. There are several basic facts about this set.

1. It is self-similar, with 4 copies of itself rescaled by 1/4. So it has box dimension and Hausdorff dimension equal to 1.

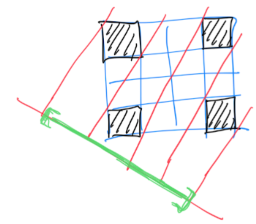
2. If you project the four-corner Cantor set in the direction $(1, 2)$, you get a full interval. See Figure 14b.

The following is also true about the four-corner Cantor set, but it is highly nontrivial.

Theorem 10.5. *The projection of the four-corner Cantor set in almost every direction has length zero.*



(a) The four corner Cantor set is $\bigcap_n E_n$



(b) Projection in direction $(1, 2)$

Figure 14

(To define "almost every" requires some measure theory. It means that the set where the statement fails is a set of measure zero. Don't worry about this)

Let's take Theorem 10.5 for granted now. This, together with Figure 14b, gives us the following: If you rotate the four-corner Cantor set by angle $\arctan(1/2)$, then you get a set A which we projected vertically is an interval, but in almost every direction, the projection has length zero. Then using our prior discussion, $L(A)$ has area zero. By making finitely many rotated copies of $L(A)$, we get another set of area zero which has a line in every direction. Nice!

(The next few paragraphs don't make much sense.)

To prove Theorem 10.5, we could do it directly, using some self-similar properties along with some "density" arguments. Or we could use the following very powerful theorem.

Theorem 10.6 (Besicovitch projection theorem)**.** *Let E be a compact purely unrectifiable set of positive and finite 1-dimensional Hausdorff measure. Then the projection of E in almost every direction has length zero.*

The four-corner Cantor set satisfies all the conditions of Theorem 10.6. Purely unrectifiable means that any smooth curve intersects E in a set of length zero.

Uhh, like I said, that probably didn't make sense. Take a look instead at Kenneth Falconer's *The geometry of fractal sets*, especially Chapters 6 and 7.

10.4 Proof of Kakeya conjecture in 2-dimensions

Theorem 10.7. *There exists a constant $C > 0$ such that if $E \subset \mathbb{R}^2$ is any set with a unit line segment in every direction, then $N_\delta(E) \gtrsim (1/\delta)^2(\log(1/\delta))^{-1}$.*

The proof is very combinatorial. The basic idea is two line segments cannot intersect "too much." If you want an efficient cover of a Besicovitch set with boxes, most of the boxes will not be able to cover many line segments at once.

More precisely, if two line segments L_1 and L_2 make an angle θ with each other and $\theta \geq \delta$, then the $\{x \in L_1 : \text{distance from } x \text{ to } L_2 \text{ is } \leq \delta\}$ has length $\lesssim \delta/\theta$. So if we try to cover $L_1 \cup L_2$ with balls of size δ, then most of the balls will only intersect one of the two segments.

Corollary 10.8. *Every Besicovitch set in the plane has box dimension* 2.

10.5 Circular arcs

We showed that we can move a needle (unit line segment) from any position in the plane to any other position in arbitrarily small area. What happens if we have a curved needle? Kornélia Héra and Miklós Laczkovich showed that if you have a short circular arc, you can do the same. (Their proof works for an arc that is 1/5 of the circle, but not for 1/4 of the circle.) They did this by adapting the construction used by Cunningham in his paper "The Kakeya Problem for Simply Connected and for Star-Shaped Sets." (Cunningham does not use Perron trees.)

See https://arxiv.org/abs/1802.00290

Marianna Csörnyei (my PhD advisor) and I showed that we could do it for circular arcs smaller than 1/2 of a circle. We actually proved something for rectifiable curves in general. The result for rectifiable curves is a little technical to state here.

We looked at what Cunningham's construction was like in the dual, and discovered that it was made up of Venetian blind-like zigzags. Then we realized that this path could also be used to move a circular arc and other curves. Point-line duality led us to our construction even though our construction does not use point-line duality in the end.

See https://arxiv.org/abs/1609.01649

For arcs longer than 1/2 of a circle, we don't if it's possible.

10.6 Connections to Fourier analysis

In the 1970s, Charles Fefferman connected the Kakeya needle problem to a question in Fourier analysis. He did this in a 7-page paper called "The multiplier problem for the ball."

Here is the question, in intuitive terms. Suppose you have a function $f : \mathbb{R}^2 \to C$, and you remove the high frequencies of f in a very particular way. As you add these frequencies back in, will you get back f in the limit? Most mathematicians believed the answer to be "for some notions of convergence, yes." But Fefferman showed that the answer was no! He was able to construct a counterexample using Perron tree-like sets.

Here is the above discussion more precisely. Start with a function $f : \mathbb{R}^2 \to \mathbb{C}$. Define the Fourier transform $\widehat{f}(\xi) = \int_{\mathbb{R}^2} f(x) e^{-2\pi i x\cdot\xi}\, dx$ and the ball multiplier operator $T_R f(x) = \int_{|\xi|\le R} \widehat{f}(\xi) e^{2\pi i x\cdot\xi}\, d\xi$. Is is true that $\lim_{R\to\infty} T_R f = f$? If the limit is viewed in $L^2(\mathbb{R}^2)$, the answer is yes (easily, by Plancherel), but Fefferman showed that for $p \neq 2$, convergence does not necessarily hold in $L^p(\mathbb{R}^2)$.

11 Day 5 exercises

Exercise 11.1. () Check the details of the proof of the iterated Venetian blind construction used in Theorem 10.1.

Exercise 11.2. () Prove Theorem 10.7. It's really a combinatorial argument.

Exercise 11.3. () Show that Corollary 10.8 follows from Theorem 10.7.