

## Contents

<b>4</b>	<b>Fractals</b>	<b>1</b>
4.1	Classical Fractal Constructions . . . . .	1
4.1.1	Generalized Cantor Sets . . . . .	1
4.1.2	The Koch Curve and Koch Snowflake . . . . .	3
4.1.3	The Sierpinski Triangle . . . . .	5
4.1.4	The Sierpinski Carpet and Menger Sponge . . . . .	6
4.2	Topological Dimension and Box-Counting Dimension . . . . .	8
4.2.1	The Topological Dimension of a Set . . . . .	9
4.2.2	The Box-Counting Dimension of a Set . . . . .	12
4.3	Fractals as Invariant Sets of Iterated Function Systems . . . . .	15
4.3.1	Iterated Function Systems and Self-Similar Sets . . . . .	15
4.3.2	The Box-Counting Dimension of a Self-Similar Set . . . . .	16
4.3.3	The “Chaos Game” . . . . .	19

## 4 Fractals

In this chapter, our goal is to discuss fractals, which are (roughly speaking) sets that possess a self-similarity on every scale: in other words, a set that appears to contain copies of itself when magnified. *A priori*, it may seem that fractals have nothing to do with dynamical systems, but as we will discuss, the stable sets of certain dynamical systems are fractals, and many examples of fractals arise as the attracting sets of iterated function systems in higher-dimensional spaces.

We will begin by discussing a few standard fractal constructions, and then turn our attention toward studying fractals from a topological perspective: in particular, we will define the Minkowski dimension of a self-similar set, which gives a reasonably intuitive notion of how much “larger” a fractal is relative to a typical one-dimensional object like a smooth curve. We will then study how a number of common fractals arise as the stable sets of iterated function systems, which are dynamical systems in Euclidean space whose iterations involve random motion.

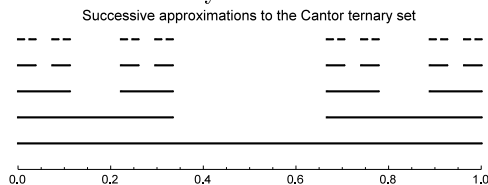
### 4.1 Classical Fractal Constructions

- We will give a brief tour of a number of standard fractal constructions.

#### 4.1.1 Generalized Cantor Sets

- An example of a fractal we have already seen is the Cantor ternary set  $\Gamma = \bigcap_{n=0}^{\infty} C_n$ , where  $C_0 = [0, 1]$  and  $C_{n+1}$  is obtained by deleting the open middle third of each interval in  $C_n$ , for each  $n \geq 0$ .

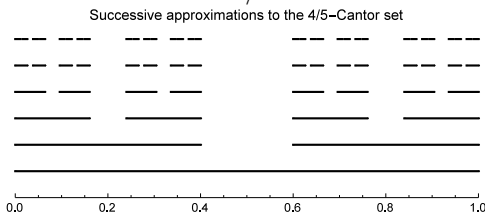
- Here (once again) is a picture of the Cantor ternary set:



- Recall also that we showed that the points in the Cantor set are those elements of  $[0, 1]$  having base-3 decimal expansion containing only 0s and 2s.
- Notice that the Cantor set contains two identical copies of itself, as the part of  $\Gamma$  in the interval  $[0, 1/3]$  and the part of  $\Gamma$  in the interval  $[2/3, 1]$  are (geometrically) similar to  $\Gamma$  itself: they are identical copies of  $\Gamma$  at  $1/3$ -scale.
  - In particular, by the iterative construction of  $\Gamma$ , we also see that it contains four scale- $1/9$  copies of itself: one in  $[0, 1/9]$ , one in  $[2/9, 1/3]$ , one in  $[2/3, 7/9]$ , and one in  $[8/9, 1]$ .
  - From the recursive nature of the definition, we can see in general that the Cantor ternary set contains  $2^n$  smaller copies of itself, each of which can be scaled by a factor of  $3^n$  to obtain the original set. In particular, no matter how closely we zoom in on the Cantor ternary set, it will display the same complicated structure.
  - Contrast this behavior with that of a differentiable curve: the foundational idea of differential calculus is that a differentiable function can be closely approximated by its tangent line, so zooming in on the graph of a differentiable function will yield a graph that approaches the graph of the tangent line.
- Also notice that, although the Cantor ternary set's construction is fairly simple, the set itself is quite complicated, and there is no simple geometric or algebraic condition that describes which points lie in the set.
  - The condition about the base-3 decimal expansion is not really an algebraic or geometric statement: there is no simple equation satisfied by the points lying in the Cantor set but not by the points outside it.
- There are a number of ways to generalize the construction of the Cantor ternary set. One way is to remove a different fixed proportion from the middle of each interval:

- **Definition:** For a real number  $\alpha$  with  $0 < \alpha < 1$ , the open middle- $\alpha$  Cantor set is defined to be  $\Gamma_\alpha = \bigcap_{n=0}^{\infty} C_n^{(\alpha)}$ , where  $C_0^{(\alpha)} = [0, 1]$  and  $C_{n+1}^{(\alpha)}$  is obtained by deleting the open middle  $\alpha$ th of each interval in  $C_n^{(\alpha)}$ , for each  $n \geq 0$ .

- Thus, for example,  $C_1 = [0, (1 - \alpha)/2] \cup [(1 + \alpha)/2, 1]$  and so forth. If  $\alpha = 1/3$ , of course, then this construction gives the Cantor ternary set.
- Here is a picture of the first few iterations for  $\alpha = 1/5$ :



- It is easy enough to see that the set  $C_n$  consists of  $2^n$  intervals each of length  $\left(\frac{1 - \alpha}{2}\right)^n$ , so the total lengths of these intervals is  $(1 - \alpha)^n$ , which goes to zero exponentially fast as  $n \rightarrow \infty$ .
- It is quite clear that the open middle- $\alpha$  Cantor set shares the same general properties as the Cantor ternary set. (In fact, the sets are homeomorphic, although this is not entirely trivial to prove.)

- There are a number of other ways to generalize the Cantor set construction. One possibility is to remove a different proportion from the middle at each stage.

- Explicitly, if  $\alpha_1, \alpha_2, \dots$  is a sequence of real numbers lying in the interval  $(0, 1)$ , then we can define a generalized Cantor set by taking  $\Gamma_{\{\alpha_i\}} = \bigcap_{n=0}^{\infty} C_n$ , where  $C_0 = [0, 1]$  and  $C_{n+1}$  is obtained by removing the open middle  $\alpha_{n+1}$ th of each interval in  $C_n$ .

- In this case, it can be verified that the total length of all the intervals in  $C_n$  is  $\prod_{i=1}^{n-1} (1 - \alpha_i)$ .

- If the values  $\alpha_i$  are chosen appropriately, it is possible to make the infinite product  $\prod_{i=1}^{n-1} (1 - \alpha_i)$  positive.

For example, we could choose  $\alpha_n = \frac{1}{n^2}$ : then one can verify by induction that  $\prod_{i=1}^{n-1} (1 - \alpha_i) = \frac{n+1}{2n}$ , so

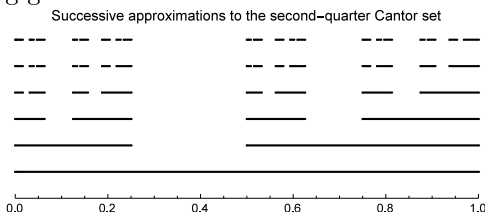
the infinite product is  $\frac{1}{2}$ .

- The generalized Cantor set associated to this sequence has the strange property that it contains no intervals and is a closed set, yet it cannot be covered by a union of intervals of total length less than  $1/2$ .

- Still another generalization would be to remove an open interval from the interior of each interval at each stage, but not necessarily the middle.

- One such construction is as follows: begin with  $C_0 = [0, 1]$ , and then, to construct  $C_{n+1}$  for  $n \geq 0$ , for each interval  $[a, b]$  in  $C_n$ , divide the interval into four equal quarters and remove the second open quarter, thus leaving  $[a, a + (b - a)/4] \cup [(a + b)/2, b]$ .

- Here is a picture of the resulting generalized Cantor set:



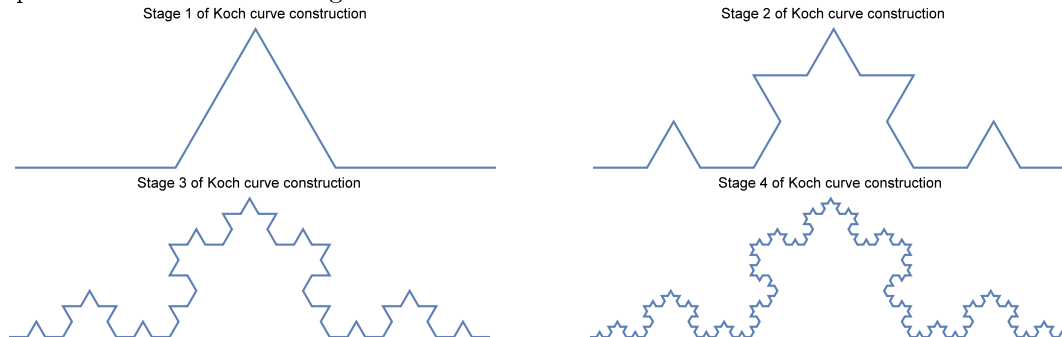
- In this case, the intervals at each stage no longer have equal lengths, although the sum of the lengths of the intervals at the  $n$ th stage of the construction is  $(3/4)^n$ , which still tends to zero.

- This Cantor set contains a  $1/4$ -scale copy of itself on the left and a  $1/2$ -scale copy of itself on the right.

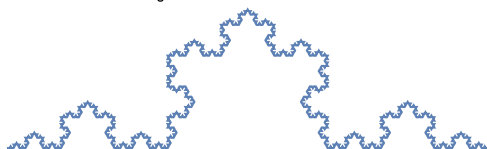
#### 4.1.2 The Koch Curve and Koch Snowflake

- Another classical example of a fractal is the Koch curve: let  $E_0$  be a line segment of length 1. Then, for each  $n \geq 1$ , define the set  $E_n$  to be the set obtained by removing the middle third of each segment in  $E_{n-1}$  and replacing it with the other two sides of the equilateral triangle sharing those endpoints. The Koch curve is the limiting set as  $n \rightarrow \infty$ .

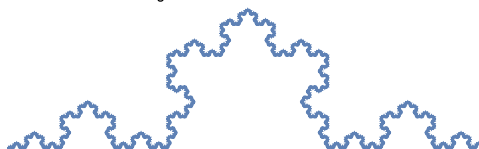
- Here are pictures of the first few stages of the construction:



Stage 5 of Koch curve construction

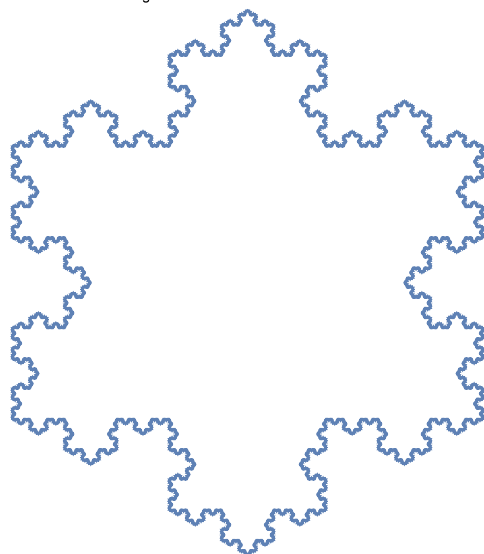


Stage 6 of Koch curve construction

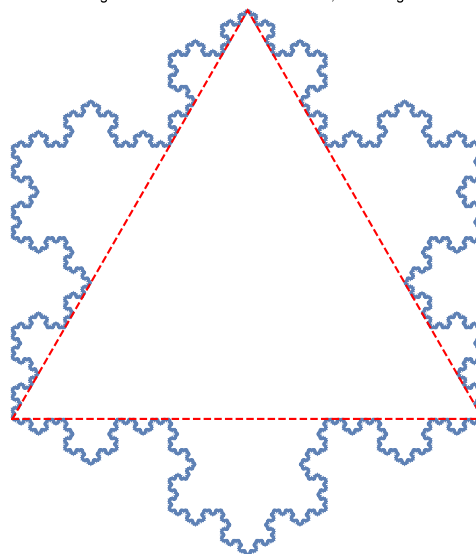


- The  $n$ th stage of the Koch curve is constructed out of  $4^n$  line segments each of length  $3^{-n}$ , so the total length is  $(4/3)^n$ . Thus, as  $n \rightarrow \infty$ , the total length of the segments goes to  $\infty$ .
- Notice that the Koch curve is also self-similar, in that it contains identical smaller copies of itself. For example, each of the four intervals in  $E_1$  sprouts a  $1/3$ -scale copy of the curve.
- It can be proven that the Koch curve is continuous (which justifies calling it a “curve”): namely, that there exists a continuous function  $f : [0, 1] \rightarrow \mathbb{R}^2$  whose image in  $\mathbb{R}^2$  is the Koch curve.
  - The construction of this function is as follows: take  $f_n : [0, 1] \rightarrow \mathbb{R}^2$  to be the continuous, piecewise-smooth function that traces out the  $4^n$  segments in the  $n$ th stage of the Koch curve at constant speed.
  - Then we claim that  $|f_{n+1}(t) - f_n(t)| \leq 3^{-n}$  for every  $t \in [0, 1]$ : first notice that  $f_{n+1}$  and  $f_n$  agree at all of the points in the space  $C_n$  that do not change when we construct  $C_{n+1}$  (i.e., near the endpoints of each of the segments in  $C_n$ ). Therefore,  $f_{n+1}$  and  $f_n$  only differ on the new segments appearing in  $C_{n+1}$ , and the largest distance between any point on either of these new segments and any point on the old segment is  $3^{-n}$ .
  - Now define  $f(t) = \lim_{n \rightarrow \infty} f_n(t)$ . This sequence converges by the inequalities above, and we see that  $|f(t) - f_n(t)| \leq 2 \cdot 3^{-n}$  for any  $t \in [0, 1]$ .
  - This means that the sequence of functions  $\{f_n\}_{n \geq 1}$  converges “uniformly” to its limit  $f$  on the interval  $[0, 1]$ , in the sense that  $\lim_{n \rightarrow \infty} \left[ \max_{0 \leq t \leq 1} |f_n(t) - f(t)| \right] = 0$ . It is then a standard theorem of real analysis that a uniform limit of continuous functions on a closed and bounded interval is continuous. (In fact, it can be shown that  $f$  is actually a homeomorphism.)
  - We also note that the function  $f$  is nondifferentiable everywhere, since the arclength of its graph is infinite on any arbitrarily small interval.
- The Koch snowflake is obtained by constructing three copies of the Koch curve along the edges of an equilateral triangle:

Stage 6 of Koch snowflake construction



Stage 6 of Koch snowflake construction, with triangle



- Interestingly, the Koch snowflake has an infinite perimeter, but encloses a finite area.
  - From our calculations above, the perimeter of the  $n$ th stage is  $3 \cdot (4/3)^n$ , which goes to  $\infty$  as  $n \rightarrow \infty$ .

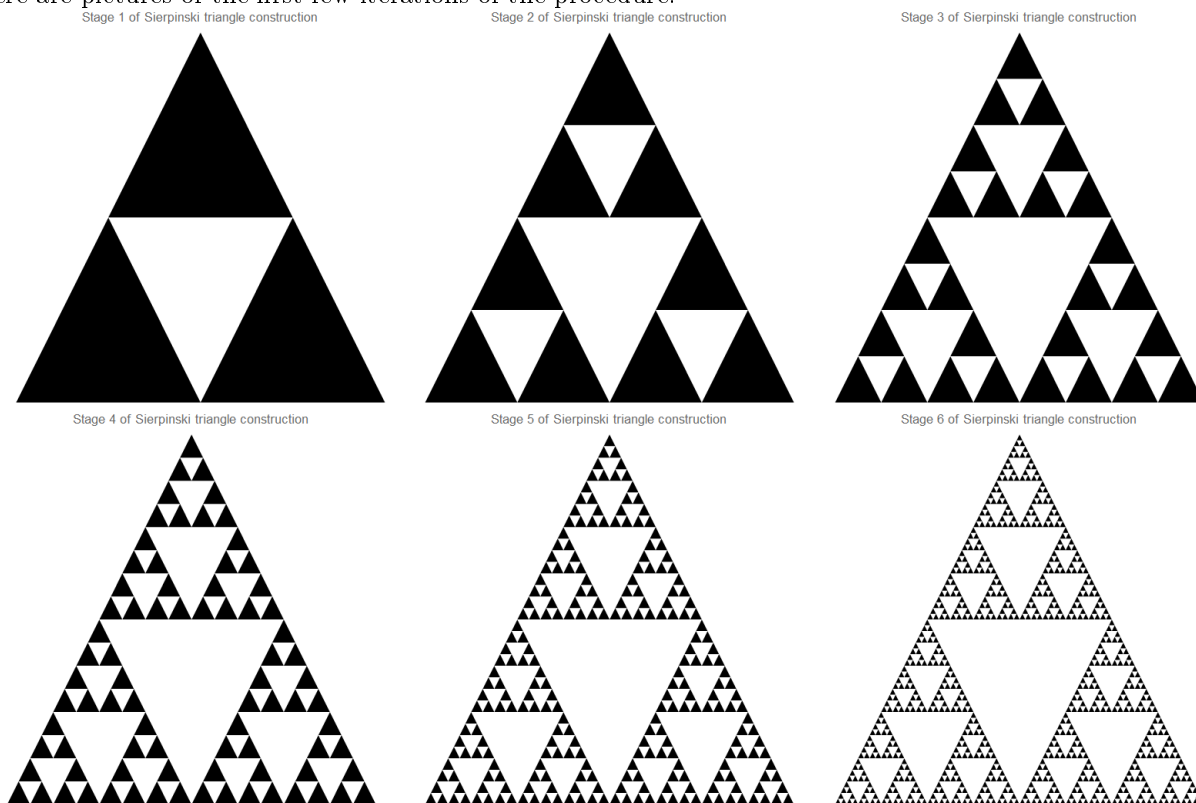
- For the area, in going from the  $n$ th stage to the  $(n + 1)$ st stage, each of the  $3 \cdot 4^n$  segments gets a new equilateral triangle of side length  $\frac{1}{3} \cdot 3^{-n}$  embedded into it, for a total area of  $3 \cdot 4^n \cdot \frac{3^{-2n} \sqrt{3}}{9 \cdot 4} = \frac{\sqrt{3}}{12} (4/9)^n$ .
- Thus, the total area inside the Koch snowflake is  $\frac{\sqrt{3}}{4} + \sum_{n=0}^{\infty} \frac{\sqrt{3}}{12} (4/9)^n = \frac{2\sqrt{3}}{5}$ .

### 4.1.3 The Sierpinski Triangle

- Another well-known fractal is the Sierpinski triangle (also called the Sierpinski gasket): let  $T_0$  be any solid triangle. Then to construct  $T_1$ , we remove the “midpoint triangle” of  $T_0$  (namely, the triangle whose vertices are the midpoints of the sides of  $T_0$ ), forming three smaller triangles similar to  $T_0$ . We then repeat this procedure on each of the three smaller triangles, and continue iterating.

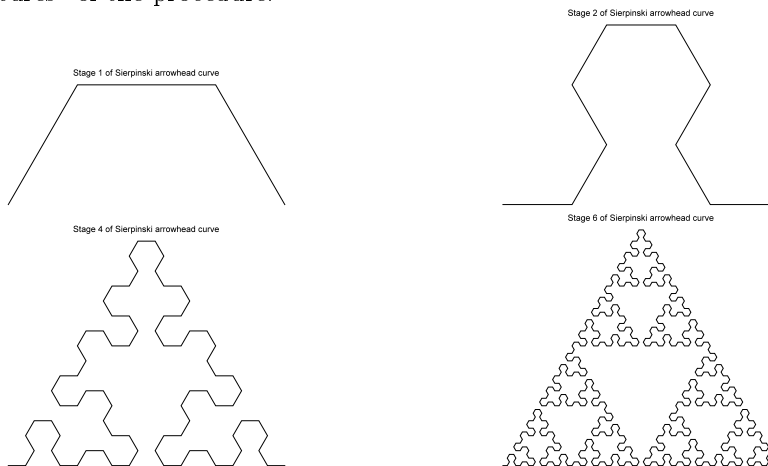
- Explicitly,  $T_{n+1}$  is obtained by removing the midpoint triangle of each triangle in  $T_n$ . The Sierpinski triangle  $T$  is the intersection  $T = \bigcap_{n=0}^{\infty} T_n$ .

- Here are pictures of the first few iterations of the procedure:



- As with the other fractals we have seen, the Sierpinski triangle is self-similar: it contains three  $1/2$ -scale copies of itself. Its area is also zero, since the  $n$ th stage  $T_n$  has area  $(3/4)^n$ .
- The Sierpinski triangle (rather unexpectedly) can also be generated as a curve, called the Sierpinski arrowhead curve, using an iterative replacement procedure similar to the one that generates the Koch curve.
  - The method uses an algorithm often referred to as “turtle graphics” (where a “turtle” moves through the plane drawing line segments to create a piecewise-linear path).
  - To describe the algorithm, the letter  $F$  stands for “move forward one unit”, the symbol  $+$  stands for “rotate left 60 degrees”, the symbol  $-$  stands for “rotate right 60 degrees”, and the letters  $X$  and  $Y$  are placeholders that mean “do nothing”.

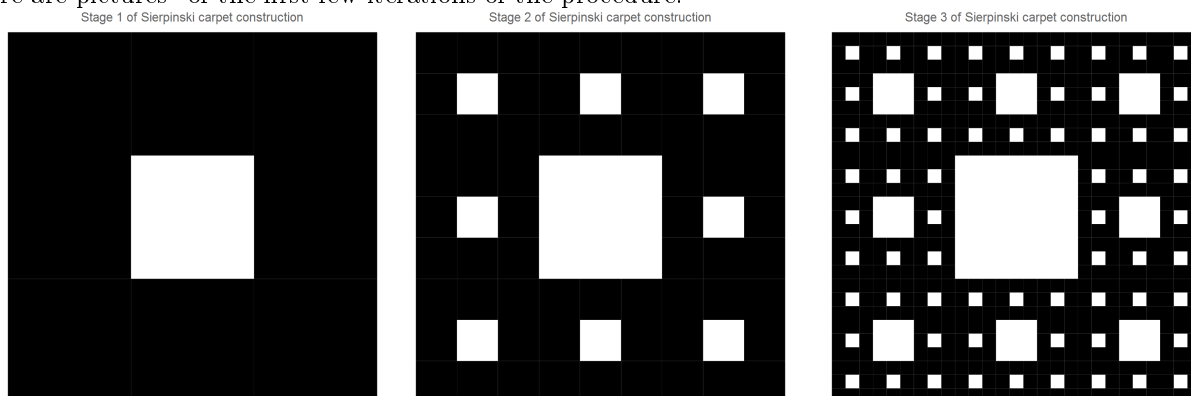
- Thus, the string  $F + X + F ++F$  tells the turtle to move forward one unit, rotate left 60 degrees, do nothing, rotate left 60 degrees, move forward one unit, rotate left 120 degrees, and move forward one unit: thus drawing an equilateral triangle.
- To describe the Sierpinski arrowhead path, we use an iterative replacement: we start with the string  $YF$ , and then generate a new string by replacing all occurrences of  $X$  with  $YF + XF + Y$  and all occurrences of  $Y$  with  $XF - YF - X$ .
- Thus, after one iteration we obtain the string  $XF - YF - XF$ , after two iterations we obtain  $YF + XF + YF - XF - YF - XF - XF - YF - XF$ , and so forth.
- Here are a few pictures<sup>1</sup> of the procedure:



- It can be proven (in a manner similar to the argument we gave for the Koch curve) that, if we rescale each curve to be a function  $f_n : [0, 1] \rightarrow \mathbb{R}^2$  mapping into the appropriate equilateral triangle, then the limit  $\lim_{n \rightarrow \infty} f_n$  exists and is a continuous function  $f : [0, 1] \rightarrow \mathbb{R}^2$  whose image is the Sierpinski triangle.

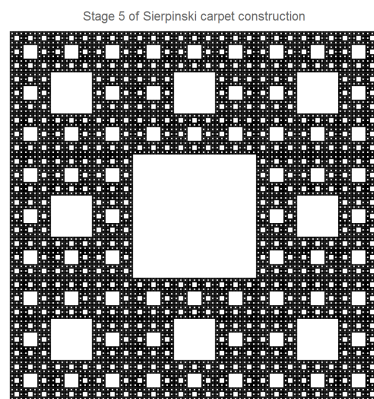
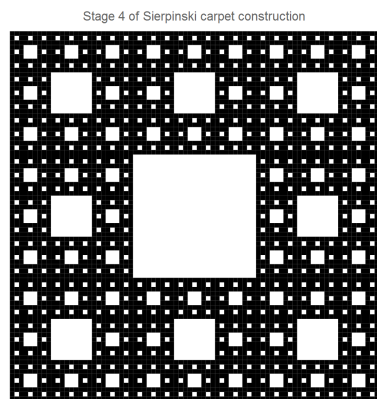
#### 4.1.4 The Sierpinski Carpet and Menger Sponge

- Another fractal whose construction is similar to that of the Cantor set and the Sierpinski triangle is the Sierpinski carpet: let  $S_0$  be a solid square. Then to construct  $S_1$ , we subdivide  $S_0$  into nine congruent squares and remove the center square, thus forming eight smaller squares. We then repeat the construction, iteratively, on each of the smaller squares.
- Here are pictures<sup>2</sup> of the first few iterations of the procedure:



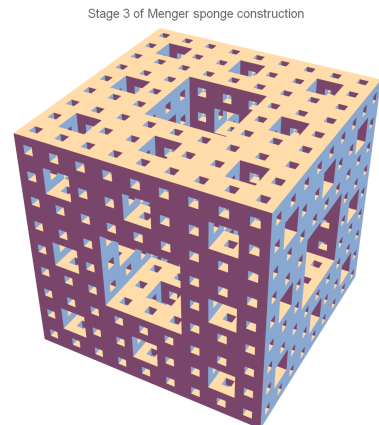
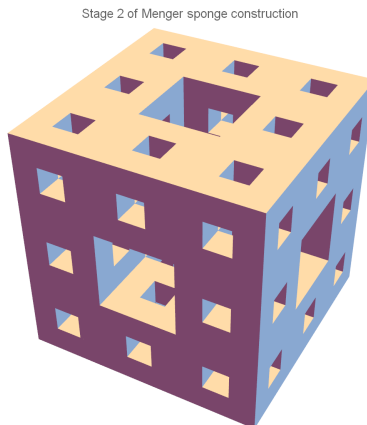
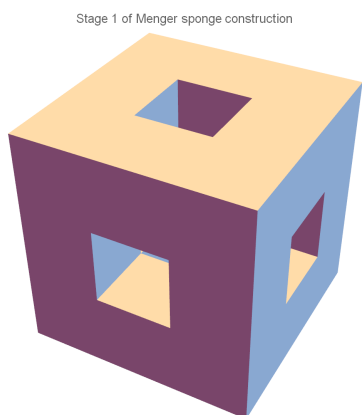
<sup>1</sup>Thanks to Robert M. Dickau for providing the code used to create these graphics.

<sup>2</sup>Thanks to Peter House and the Wolfram Demonstrations Project for providing the code used to create these graphics.



- As with all the other fractals we have discussed, the Sierpinski carpet contains a number of smaller copies of itself: specifically, it contains 8 copies of itself at  $1/3$ -scale.
- The points lying in the Sierpinski carpet have a reasonably nice description, similar to the one given for the Cantor set: they are those points  $(x, y) = (0.x_1x_2x_3\dots, 0.y_1y_2y_3\dots)$  in base 3 such that  $(x_i, y_i) \neq (1, 1)$  for every  $i$ .
  - This follows by observing (by induction) that the points removed at the  $n$ th stage of the construction are those of the form  $(x, y)$  where the  $n$ th base-3 digit of both  $x$  and  $y$  is a 1.
- Despite how it may appear in the pictures, the Sierpinski carpet actually has area zero, and has no interior (i.e., any open disc around any point will always be missing some points).
  - At each stage,  $1/9$  of the remaining area is removed, so the area of the  $n$ th stage  $S_n$  is  $(8/9)^n$ , which tends to zero as  $n \rightarrow \infty$ .
  - Only six stages of the construction are shown, so the total area remaining in the last picture is  $(8/9)^6 \approx 0.493$ , just under half of the total area of the square.
  - The statement about the interior follows from the fact that the carpet has area zero: since the area of any open disc is positive but the area inside the Sierpinski carpet is zero, most of the points in the disc must be missing from the Sierpinski carpet.
  - Alternatively, we can actually find specific points that are missing: any open disc of positive radius will contain some points of the form  $(x, y)$  where both  $x$  and  $y$  have a digit 1 in the same place in their base-3 decimal expansions, and these points will be missing from the set.
- The three-dimensional version of the Sierpinski carpet is called the Menger sponge: let  $C_0$  be a solid cube of unit side length. Then to construct  $C_1$ , we subdivide  $C_0$  into twenty-seven congruent cubes and remove the cube in the middle of each face along with the center cube, thus leaving 20 smaller cubes. We then repeat the construction, iteratively, on each of the smaller cubes.

- Here are pictures<sup>3</sup> of the first few iterations:



<sup>3</sup>Thanks to Jaime Rangel-Mondragon and the Wolfram Demonstrations Project for providing the code used to create these graphics.

- The Menger sponge is quite similar to the Sierpinski carpet.
  - For example, it has zero volume, and thus contains no open sets in its interior, since the volume at the  $n$ th stage of the construction is  $(20/27)^n$ .
  - Its surface area, however, is infinite: it can be verified, either by writing down a recursive formula or by a careful count, that the total surface area after the  $n$ th stage is  $2\left(\frac{20}{9}\right)^n + 4\left(\frac{8}{9}\right)^n$ , which clearly goes to  $\infty$  as  $n \rightarrow \infty$ .
  - Also, the points in the Menger sponge are those points  $(x, y, z) = (0.x_1x_2x_3\dots, 0.y_1y_2y_3\dots, 0.z_1z_2z_3\dots)$  in base 3 such that  $(x_i, y_i), (x_i, z_i), (y_i, z_i) \neq (1, 1)$  for every  $i$ .

## 4.2 Topological Dimension and Box-Counting Dimension

- In the previous section, we gave examples of fractals but did not actually define what a fractal was. We can, at least, identify a few common themes to the examples:
  - Each of the sets possessed some kind of self-similarity.
  - Each of the sets had detailed structure at every scale.
  - Each of the sets had a reasonably simple recursive definition.
  - Each of the sets was too “irregular” to study using geometry. (The planar fractals each had area zero, for example.)
- Ultimately, giving a more precise definition of a fractal in a similar manner to the precise definition of a chaotic system is somewhat difficult, and many proposed definitions do not exactly capture all behaviors that could be viewed as “fractal-like”.
  - To motivate our definition, notice that, at least based on the pictures, it seems reasonable to say that the Sierpinski carpet seemed somehow takes up “more space” than the Sierpinski triangle, which in turn seems to take up “more space” than the Koch curve.
  - Of course, as it stands these statements do not really make any sense: all of these sets have zero area, so in that sense they are all the same size. (Their cardinalities are all equal as well, so as sets they are the same size too.)
  - We would like to be able to draw a finer distinction between these sets than “area zero”, since the sets visually seem to have different sizes: the Koch curve seems very “curve-like”, whereas the Sierpinski carpet seems much more “area-like”.
- Ultimately, the idea we are seeking to describe is that of “dimension”: the Sierpinski carpet has a larger dimension than the Sierpinski triangle, for an appropriate definition of dimension.
- There are, in fact, many different definitions of dimension.
  - The common understanding of the dimension of a space like the real line or the plane is the number of coordinates needed to specify a point lying within it.
  - With this definition, the surface of a sphere is 2-dimensional, because one needs two parameters to specify a location on a sphere (e.g., longitude and latitude). On the other hand, a differentiable curve has dimension 1, because it is only necessary to give one parameter to specify a location on a curve (e.g., the distance along the curve measured relative to a fixed point).
  - Another notion of dimension, from linear algebra, is the dimension of a vector space: this quantity is always a nonnegative integer, and the vector space dimension of  $\mathbb{R}^n$  as an  $\mathbb{R}$ -vector space is  $n$ , as we would expect it to be.
- However, our examples of planar fractals are not vector spaces (certainly not in any obvious sense), nor is it easy to decide whether they require a single parameter or pair of parameters to describe them.
  - The Sierpinski triangle, for example, was initially defined by removing pieces of a two-dimensional solid triangle, so it seems reasonable to declare that it is a two-dimensional object.

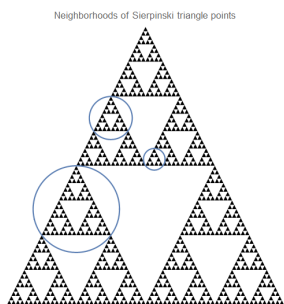


- But we also demonstrated how to construct it using a limit of continuous curves, which are one-dimensional objects.
- It is not easy to reconcile these two facts: one construction suggests that the Sierpinski triangle is two-dimensional, while the other suggests it is one-dimensional.
- Ultimately, the best answer turns out to be that the dimension of the Sierpinski triangle is somewhere between 1 and 2.

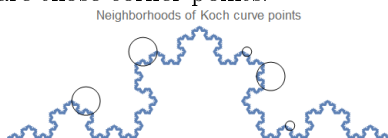
#### 4.2.1 The Topological Dimension of a Set

- In order to understand how fractals differ from more typical sets like curves and regions, we will need to discuss some topology. (Many of the definitions may be familiar, but we will collect them all here for easy reference.)
- Most of our results will be rather technical, and we will not refer in very much depth to the details of these topics in our later discussions.
  - Everything we discuss can be constructed in the setting of a general metric space, but we will work exclusively in  $\mathbb{R}^n$  for concreteness.
  - If  $x$  is a point in  $\mathbb{R}^n$ , the open ball of radius  $r$  centered at  $x$  is the set of points  $B_r(x) = \{y \in X : |x - y| < r\}$  within a distance  $r$  of  $x$ .
  - A subset of  $\mathbb{R}^n$  is called a neighborhood of a point  $x$  if it contains some open ball  $B_r(x)$  for some  $r > 0$ .
  - A subset  $U$  of  $\mathbb{R}^n$  is open if, for every  $x \in U$ , there is some open ball of some positive radius centered at  $x$  contained in the set  $U$ . Equivalently,  $U$  is open if it contains a neighborhood of each of its points.
  - If  $S$  is a subset of  $\mathbb{R}^n$ , the boundary of  $S$  is the set of points  $x \in \mathbb{R}^n$  such that any open ball centered at  $x$  contains some points in  $S$  and some points outside of  $S$ .
  - Example: The boundary of the open ball  $B_r(x)$  is the “sphere” of radius  $r$  centered at  $x$ , namely, the set of points of distance exactly  $r$  away from  $x$ . In particular, in  $\mathbb{R}^1$  the boundary of an interval  $[a, b]$  is the set of endpoints  $\{a, b\}$ .
  - Example: The boundary of the set of rational numbers (inside  $\mathbb{R}$ ) is the entire line  $\mathbb{R}$ , since any open ball will contain some points of  $\mathbb{Q}$  and some points not in  $\mathbb{Q}$ .
- Our first step will be to define the topological dimension of a set of points in a metric space, which is constructed inductively.
- Definition: A nonempty subset  $S$  of  $\mathbb{R}^n$  has topological dimension zero if every point in  $S$  possesses arbitrarily small neighborhoods whose boundaries do not intersect  $S$ .
  - Any finite set of points has topological dimension zero: if the smallest distance between any of the two points is  $d$ , then the boundary of the ball of radius  $d/2$  around any point contains none of the other points.
  - More generally, a countable subset of points (such as the set of rational numbers) has topological dimension zero.
  - The idea is to observe that for any point  $x$ , there are uncountably many disjoint sets that are the boundary of a ball centered at  $x$ , namely, the ‘spheres’ of varying radius  $r > 0$ , but since the set only has countably many points, most of the spheres must contain no points of the set.
- Example: The Cantor ternary set  $\Gamma$  has topological dimension zero.
  - If  $x \in \Gamma$  is any point in the set, there are points  $a < x$  and  $b > x$  not lying in  $\Gamma$  that are arbitrarily close to  $x$ : then the interval  $[a, b]$  is a neighborhood of  $x$  whose boundary does not intersect  $\Gamma$ .
- Non-Example: A line segment does not have topological dimension zero.
  - Any sufficiently small neighborhood of any point on the line segment will intersect the line segment in at least one point.

- Non-Example: Any continuous curve (i.e., the set of points in the image of any continuous map  $f : [0, 1] \rightarrow \mathbb{R}^n$ ) will not have topological dimension zero.
  - The logic is the same as for a line segment: any sufficiently small neighborhood around any point on the curve will have its boundary intersect the curve somewhere.
  - In particular, the Koch curve and the Sierpinski triangle do not have topological dimension 0, since they are examples of continuous curves.
- We can now define the topological dimension of a space inductively.
- Definition: A nonempty subset  $S$  of  $\mathbb{R}^n$  has topological dimension  $k$  if each point in  $S$  has arbitrarily small neighborhoods whose boundaries intersect  $S$  in a set of dimension  $k - 1$ , and  $k$  is the smallest nonnegative integer with this property.
  - Notice, by this definition, the topological dimension of any set of points is always a nonnegative integer.
  - To show a set has topological dimension  $k$ , it is sufficient (i) to show the condition about neighborhood boundaries intersecting the set in sets of dimension  $k - 1$ , and (ii) to show that the set does not satisfy neighborhood boundary condition for having topological dimension  $k - 1$ .
- Example: A line segment has topological dimension 1.
  - It does satisfy the definition for topological dimension  $\leq 1$ , since any ball of sufficiently small radius around a point on a line segment will intersect the line in 1 or 2 points (which is a set of topological dimension 0).
  - Above, we saw that a line segment does not have topological dimension zero.
- Example: The Sierpinski triangle has topological dimension 1.
  - Again as we showed above, the Sierpinski triangle does not have topological dimension 0.
  - So we just need to check that any point has arbitrarily small neighborhoods that intersect the Sierpinski triangle in a finite number of points.
  - For any point  $x$  in the Sierpinski triangle, we can draw a small circle containing that point in its interior passing through the vertices of one of the small triangles in the iterative construction, as shown in the diagram:

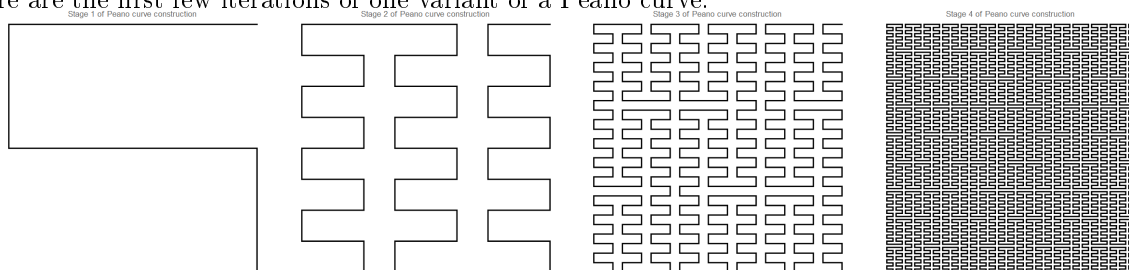


- This circle (which is the boundary of a neighborhood of  $x$ ) will intersect the Sierpinski triangle in exactly 3 points, which is a set of topological dimension 0. Since the set is self-similar, we can find arbitrarily small such circles.
- Example: The Koch curve has topological dimension 1.
    - Using a similar argument as with the Sierpinski triangle, if we draw a small circle around a point on the Koch curve in such a way that it goes through the equilateral triangle formed by three consecutive “corner points” of some stage of the curve’s construction, we can show that the only intersection points of the Koch curve with that circle are those corner points:



- Each circle will intersect the Koch curve in three points (a set of topological dimension 0).
- Furthermore, since the Koch curve does not have topological dimension 0 as we saw before, it must have topological dimension 1.
- **Example:** The set of points in the unit square  $[0, 1] \times [0, 1]$  has topological dimension 2.
  - Any rectangle drawn around any point in the unit square will intersect the square in a union of line segments, which has topological dimension 1.
  - Furthermore, any other neighborhood of any point in the unit square will intersect the square in (at least) a curve, which has topological dimension at least 1.
- One of the most useful properties of the topological dimension is that it is invariant under homeomorphism: if  $f : U \rightarrow V$  is a homeomorphism, then the topological dimension of  $U$  is equal to the topological dimension of  $V$ .
  - Thus, any set that is homeomorphic to the interval  $[0, 1]$  has topological dimension 1. This provides another way to see that the Koch curve has topological dimension 1, because the map we defined from  $[0, 1]$  to the curve is actually a homeomorphism (though this is not so easy to show directly).
- It may seem, based on our discussion and the above fact, that the image of a continuous curve  $f : [0, 1] \rightarrow \mathbb{R}^n$  should have topological dimension 1. However, this is not true! There are in fact continuous functions  $f : [0, 1] \rightarrow \mathbb{R}^n$  whose image is a solid region in space.

- Such curves are called “space-filling curves”, and, like the other kinds of curves we have discussed, are defined using a limiting procedure.
- A particularly famous space-filling curve is called the Peano curve. Its construction is made recursively using the same kind of “turtle graphics” algorithm we gave for drawing the Sierpinski arrowhead curve. Here are the first few iterations of one variant of a Peano curve:



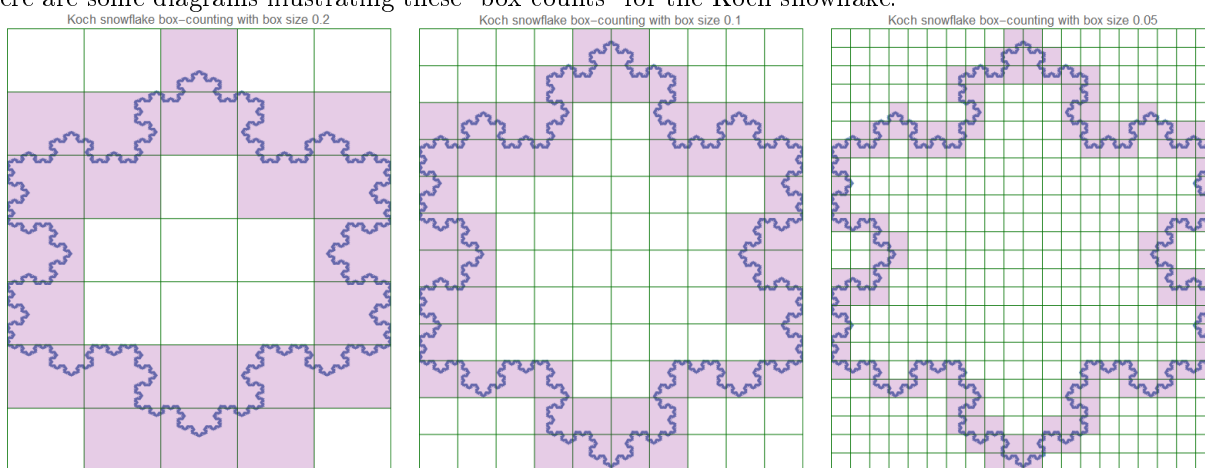
- As with the Koch curve and the Sierpinski arrowhead curve, the Peano functions  $f_n : [0, 1] \rightarrow \mathbb{R}^2$  representing the curves drawn above have a limit  $f$  which is a continuous map from  $[0, 1]$  to the unit square  $[0, 1] \times [0, 1]$ : but this limit function is actually surjective!
- In particular, this “space filling curve” is a continuous curve that actually passes through every point in the square!
- The Peano function  $f : [0, 1] \rightarrow \mathbb{R}^2$  is then an example of a continuous function whose image has topological dimension 2.
- The Peano function is not injective, however: the curve goes through many points many times.
  - \* In fact, there is a theorem that says any continuous bijection from a compact metric space to another compact metric space is necessarily a homeomorphism. But the interval  $[0, 1]$  and the square  $[0, 1] \times [0, 1]$  are not homeomorphic.
  - \* Thus, rather bizarrely, the Peano function is actually forced not to be injective.
- There are higher-dimensional versions of the Peano curve that fill up a box in  $n$ -dimensional space.

### 4.2.2 The Box-Counting Dimension of a Set

- The topological dimension does not allow us to see the difference between the interval  $[0, 1]$  and the Sierpinski triangle, since both sets have topological dimension 1.
- We will now give a different definition of dimension that can take non-integer values.
- To motivate the definition, consider how many “half-scale copies” of various objects we would need to glue together in order to create the original object.
  - For a line segment, we need to glue together two copies.
  - For a square (or a plane region) we need to glue together four copies.
  - For a cube (or a solid region in space) we need to glue together eight copies.
  - In general, for an  $n$ -dimensional cube, we need to glue together  $2^n$  copies.
  - If we had some other kind of object whose size doubled when we glued together 3 half-size copies, then (arguing entirely by analogy) it would be at least somewhat reasonable to say that the dimension is  $\log_2 3$ , since we decided that an  $n$ -dimensional object requires  $2^n$  copies.
  - But we actually have an example of an object with this property: the Sierpinski triangle is constructed out of three half-scale copies of itself. So we would like to say that the Sierpinski triangle “has dimension  $\log_2 3$ ”, under some definition of dimension.
  - In a similar way we could try using a different scaling factor in place of  $1/2$ : for example, an  $n$ -dimensional set should also have the property that we can recover it by gluing together  $3^n$  copies each of which is  $1/3$  the size of the original.
  - Thus, if we have a set (like the Cantor ternary set) that we can create by gluing together two  $1/3$ -scale copies, we would like to say that the dimension of the set should be  $\log_3 2$ .
- In order to deal with more general kinds of sets, we require a more general definition.
  - From the above discussion, the dimension of a set is capturing something about the growth rate of the number of small pieces that are required to obtain a larger version of the set.
  - Specifically, if we can obtain the original set by gluing together  $N$  copies each of which is an  $\epsilon$ -scale version of the set (where  $\epsilon > 0$  is small), then the dimension  $d$  will satisfy the relation  $N = (1/\epsilon)^d$ .
  - Solving for the dimension  $d$  yields  $d = \frac{\log(N)}{\log(1/\epsilon)}$ . (The base of the logarithm does not matter since the ratio will be the same for any base.)
  - If we then want to try to compute this value  $d$  for a particular set, one way we could try to do this is to replace the count of the number of small copies  $N$  with something easier to determine.
  - This is the idea behind the box-counting dimension: we choose a small value of  $\epsilon$  and then try to approximate the value of the ratio  $\frac{\log(N)}{\log(1/\epsilon)}$  by counting how many small rectangular boxes of size  $\epsilon$  are needed to cover the set, rather than counting how many scaled copies of the original set of size  $\epsilon$  are needed. (It is much easier to count rectangular boxes than copies of a complicated set!)
  - Our hope is that sufficiently small rectangular boxes should be a good enough approximation to the set that if we take the limit of the ratio as  $\epsilon \rightarrow 0$ , we should recover the dimension of the set that agrees with what we were trying to describe above.
- Here is the formal definition:
- **Definition:** Suppose  $S$  is a bounded set in  $\mathbb{R}^n$ . For  $\epsilon > 0$ , subdivide  $\mathbb{R}^n$  into boxes having all sides of length  $\epsilon$  using (hyper)planes parallel to the coordinate (hyper)planes, and then define  $N(\epsilon)$  to be the number of such boxes that contain at least one point of  $S$ . We define the box-counting dimension of  $S$  to be the limit  $\dim_B(S) = \lim_{\epsilon \rightarrow 0} \frac{\log(N(\epsilon))}{\log(1/\epsilon)}$ , assuming the limit exists.

- Note: Here, and whenever we use the box-counting dimension, the limit is always to be interpreted as the limit with  $\epsilon \rightarrow 0^+$ , since none of the quantities are defined when  $\epsilon < 0$ . (We will not bother writing  $\epsilon \rightarrow 0^+$  explicitly.)
- The box-counting dimension has many different names: it is variously called the Minkowski dimension, the entropy dimension, the Kolmogorov capacity or Kolmogorov entropy, the capacity dimension, the metric dimension, and the fractal dimension. (We explicitly will avoid the term “fractal dimension”, because that term is often interpreted inconsistently by different authors.)
- In general, a reasonably simple set will have box-counting dimension equal to its topological dimension. A fractal, on the other hand, will usually have a box-counting dimension strictly larger than its topological dimension.
- There is another well-known dimension that is closely related to the box-counting dimension, called the Hausdorff dimension. Its definition is significantly more technical but it has a very similar flavor, and for most reasonable sets it is equal to the box-counting dimension. (Roughly speaking, the Hausdorff dimension is a generalization that allows boxes of varying sizes, and also allows them to have different shapes than just boxes.)

- Here are some diagrams illustrating these “box counts” for the Koch snowflake:



- In the first diagram, we count a total of  $N(0.2) = 21$  boxes of side length  $\epsilon = 0.2$ , so the ratio  $\frac{\log(N(\epsilon))}{\log(1/\epsilon)} = \frac{\log(21)}{\log(5)} \approx 1.89$ .
- In the second diagram, we count a total of  $N(0.1) = 54$  boxes of side length  $\epsilon = 0.1$ , so the ratio  $\frac{\log(N(\epsilon))}{\log(1/\epsilon)} = \frac{\log(54)}{\log(10)} \approx 1.732$ .
- In the third diagram, we count a total of  $N(0.05) = 129$  boxes of side length  $\epsilon = 0.05$ , so the ratio  $\frac{\log(N(\epsilon))}{\log(1/\epsilon)} = \frac{\log(129)}{\log(20)} \approx 1.622$ .
- Even though we have drawn very accurate pictures and used quite small boxes, the results do not seem to be converging very quickly.
- They certainly seem to be decreasing, but it is not easy to tell what value they are converging to: certainly it is not at all clear whether they are converging to our expected dimension  $\log_3 4 \approx 1.261$ .
- But we can at least see that the values are a lot larger than 1, which is what we would expect the dimension of a smooth curve to be.

- Example: Show that the box-counting dimension of a line segment is 1.

- If we embed the line segment along the  $x$ -axis starting at the origin, it is easy to see that the line segment intersects exactly  $N(\epsilon) = \left\lceil \frac{L}{\epsilon} \right\rceil$  (i.e.,  $L/\epsilon$  rounded up to the nearest integer) boxes of size  $\epsilon$ . We can therefore take  $N(\epsilon) = 1 + \frac{L}{\epsilon}$ .

- Then the box-counting dimension is  $\lim_{\epsilon \rightarrow 0} \frac{\log(1 + L/\epsilon)}{\log(1/\epsilon)} = \lim_{\epsilon \rightarrow 0} \frac{\log(L + \epsilon)}{\log(1/\epsilon)} + 1 = 1$  since the limit  $\lim_{\epsilon \rightarrow 0} \frac{\log(L + \epsilon)}{\log(1/\epsilon)}$  is zero.
- Example: Show that the box-counting dimension of the unit square is 2.
  - If we again embed the square in the plane as the region  $[0, 1] \times [0, 1]$  then the square will intersect precisely  $N(\epsilon) = \left\lceil \frac{1}{\epsilon} \right\rceil^2$  boxes of size  $\epsilon$ . We can therefore take  $N(\epsilon) = (1 + \frac{1}{\epsilon})^2$ .
  - Then the box-counting dimension is  $\lim_{\epsilon \rightarrow 0} \frac{\log((1 + 1/\epsilon)^2)}{\log(1/\epsilon)} = 2 \lim_{\epsilon \rightarrow 0} \frac{\log(1 + 1/\epsilon)}{\log(1/\epsilon)} = 2$ .
- There are a number of ways to simplify the computations required to find the box-counting dimension.
  - First, an equivalent way to define the box-counting definition is instead to define  $N(\epsilon)$  to be the smallest number of boxes having all sides of length  $\epsilon$  (whose positions can be arranged arbitrarily) whose union contains the set  $S$ . The only difference between this definition and the original one we gave is whether we are allowed to use  $\epsilon$ -boxes with arbitrary orientation and center, or whether we must use the ones that lie in the grid system.
    - \* Let us explain why these two seemingly different definitions will yield the same dimension. Let  $N_{\text{arbitrary}}$  be the minimal possible number of boxes where we allow arbitrary positions, and  $N_{\text{grid}}$  be the count of boxes where they lie in the grid system.
    - \* Clearly,  $N_{\text{arbitrary}} \leq N_{\text{grid}}$ , since any arrangement of boxes in the grid system gives an arrangement using the arbitrary system.
    - \* Also,  $N_{\text{grid}} \leq N_{\text{arbitrary}} \cdot (1 + 2\sqrt{n})^n$ , because any box having all sides of length  $\epsilon$  with arbitrary position and orientation will be covered by the  $(1 + 2\sqrt{n})^n$  boxes in the grid system that lie within a distance at most  $\sqrt{n}\epsilon$  away in each coordinate direction from the starting box.
    - \* Thus, we have  $N_{\text{arbitrary}} \leq N_{\text{grid}} \leq N_{\text{arbitrary}} (1 + 2\sqrt{n})^n$ .
    - \* By the squeeze theorem, we can conclude that the two limits  $\lim_{\epsilon \rightarrow 0} \frac{\log(N_{\text{grid}}(\epsilon))}{\log(1/\epsilon)}$  and  $\lim_{\epsilon \rightarrow 0} \frac{\log(N_{\text{arbitrary}}(\epsilon))}{\log(1/\epsilon)}$  are equal, since the constant factor  $(1 + 2\sqrt{n})^n$  does not depend on  $\epsilon$ .
  - Another equivalent way to define the box-counting dimension is to use  $\epsilon$ -balls (that is, balls of radius  $\epsilon$ ) in places of boxes of side length  $\epsilon$ .
    - \* By a similar argument as for the “grid boxes versus arbitrary boxes” definitions above, this definition is also equivalent to both of those definitions: any ball can intersect a bounded number of boxes, and any box can be covered by a bounded number of balls.
  - Finally, in practice, it is unnecessary to work with boxes of an arbitrary size  $\epsilon$ .
    - \* We can in fact get away with using boxes of particular fixed sizes  $\epsilon_1, \epsilon_2, \dots$  approaching zero, since the limit of the sequence  $\lim_{n \rightarrow \infty} \frac{\log(N(\epsilon_n))}{\log(1/\epsilon_n)}$  will be the same as the value of the continuous limit  $\lim_{\epsilon \rightarrow 0} \frac{\log(N(\epsilon))}{\log(1/\epsilon)}$ .
    - \* If we choose the box sizes as  $\epsilon_n = m^{-n}$  for some positive real number  $m$ , then by using the squeeze theorem in a manner similar to the above, one can show that the limit  $\lim_{n \rightarrow \infty} \frac{\log(N(\epsilon_n))}{\log(1/\epsilon_n)}$  exists if and only if the continuous limit  $\lim_{\epsilon \rightarrow 0} \frac{\log(N(\epsilon))}{\log(1/\epsilon)}$  exists.
    - \* In other words, it is completely sufficient to make the counts for the box sizes  $\epsilon_n = m^{-n}$  only.
- Example: Show that the box-counting dimension of the Cantor ternary set is  $\log_3 2$ .
  - We will use 1-dimensional boxes (i.e., intervals) of size  $\epsilon_n = 3^{-n}$ .
  - From the construction of the Cantor ternary set, we can cover the Cantor set using  $2^n$  intervals of size  $3^{-n}$ , since in fact the  $n$ th stage of the construction consists precisely of  $2^n$  intervals each of length  $3^{-n}$ . Furthermore, it is not possible to use fewer intervals, since no interval can intersect the interior of more than one of the segments at a time, so  $N(3^{-n}) = 2^n$ .

- Thus the limit is  $\lim_{n \rightarrow \infty} \frac{\log(2^n)}{\log(1/3^{-n})} = \lim_{n \rightarrow \infty} \frac{n \log(2)}{n \log(3)} = \frac{\log 2}{\log 3} = \log_3 2$ .
- So, as claimed, the box-counting dimension of the Cantor ternary set is  $\log_3 2$ .
- Example: Find the box-counting dimension of the Sierpinski carpet.
  - We will use 2-dimensional boxes of size  $\epsilon_n = 3^{-n}$ .
  - From the construction of the Sierpinski carpet, we can cover it using  $8^n$  boxes of size  $3^{-n}$  (since in fact the  $n$ th stage of the construction consists of exactly  $8^n$  solid boxes each of side length  $3^{-n}$ ). Furthermore, it is not possible to use fewer boxes of this size, so  $N(3^{-n}) = 8^n$ .
  - Thus the limit is  $\lim_{n \rightarrow \infty} \frac{\log(8^n)}{\log(1/3^{-n})} = \lim_{n \rightarrow \infty} \frac{n \log(8)}{n \log(3)} = \frac{\log 8}{\log 3} = \log_3 8$ .
  - So the box-counting dimension of the Sierpinski carpet is  $\log_3 8$ .

### 4.3 Fractals as Invariant Sets of Iterated Function Systems

- Using the definition of the box-counting dimension directly is rather difficult for complicated sets. Since we are primarily interested in studying fractals, which demonstrate self-similarities, we would like to find a way to compute the box-counting dimension for self-similar sets that does not require as much effort as using the definition directly.
- We will approach this problem almost in reverse, namely: given a collection of similarities that we want a fractal to possess, we will prove that there is a unique set having those similarities, and then compute the box-counting dimension of that set.
- Using our descriptions of various fractals as invariant sets for these collections of similarities, we will also give another method for “randomly” constructing and drawing fractals that is often called the “chaos game”.

#### 4.3.1 Iterated Function Systems and Self-Similar Sets

- Definition: If  $D$  is a subset of  $\mathbb{R}^n$ , we say a mapping  $T : D \rightarrow D$  is a contraction if there is a positive constant  $c < 1$  such that  $|T(x) - T(y)| \leq c|x - y|$  for all  $x, y \in D$ . If equality holds everywhere (i.e., if  $|T(x) - T(y)| = c|x - y|$  for all  $x, y \in D$ ) then we say  $T$  is a similarity.
  - A contraction is essentially a map that, when applied, moves points closer to one another.
  - Example: If  $f$  is any function that has an attracting fixed point, then (as we saw in our analysis of attracting fixed points) sufficiently near the attracting fixed point,  $f$  behaves as a contraction.
  - A similarity is a contraction that also preserves relative distances.
  - It can be proven that every similarity can be written as the composition of a rotation, a scaling, and a translation. In particular, every similarity is a linear function, in the sense that it can be written in the form  $T(x) = Ax + B$  for some  $n \times n$  matrices  $A$  and  $B$ .
  - Conversely, it is easy to see that any composition of rotations and translations (which do not change distances) with scalings (which multiply distances by a positive constant) is a similarity.
  - In particular, the similarities from  $\mathbb{R}^1$  to  $\mathbb{R}^1$  are the functions of the form  $f(x) = ax + b$  where  $0 < |a| < 1$ .
- All of the fractals we have constructed are self-similar sets. We will now describe a general recipe for constructing fractals using similarities.
- Definition: An iterated function system consists of a finite number of contractions  $T_1, \dots, T_k$  defined on some region  $D$  in  $\mathbb{R}^n$ .
  - Example: On the interval  $[0, 1]$ , the two maps  $T_1(x) = \frac{1}{3}x$  and  $T_2(x) = \frac{1}{3}x + \frac{2}{3}$  form an iterated function system, since both maps are similarities.

- **Theorem** (Invariant Set for an Iterated Function System): Suppose  $T_1, \dots, T_k$  form an iterated function system on the bounded region  $D$ . Then there exists a unique closed invariant set  $\Lambda = \Lambda(T_1, \dots, T_k)$  such that 
$$\Lambda = \bigcup_{i=1}^k T_i(\Lambda).$$
  - The set  $\Lambda$  is called an invariant set because it does not change when we apply any of the maps  $T_i$  to it. It is also often called an attractor for the iterated function system, for reasons we will discuss later.
  - **Proof** (outline): We can define the invariant set explicitly using a procedure quite similar to the method we used in symbolic dynamics.
  - Fix any  $y \in D$ . Define the set  $\Sigma$  to be the set of all points that can be written as a limit of a sequence of the form  $\lim_{n \rightarrow \infty} T_{x_0} \circ T_{x_1} \circ \dots \circ T_{x_n}(y)$ , where each  $x_i \in \{1, 2, \dots, k\}$  and  $y$  is an arbitrary element of  $D$ .
  - The limit of every such sequence must exist: if  $C$  is any closed set containing  $y$ , then the sequence  $T_{x_0}(C), T_{x_0}(T_{x_1}(C)), \dots$  is a nested sequence of closed and bounded sets, so by a generalization of Cantor's nested interval theorem, the intersection is nonempty, and must in fact consist of a single point because the maps  $T_{x_i}$  are contractions.
  - We can represent limits of such sequences using an itinerary: namely, by an infinite string of symbols  $(d_0 d_1 d_2 \dots)$  where each  $d_i \in \{1, 2, \dots, k\}$ .
  - If  $z$  has itinerary  $(d_0 d_1 d_2 \dots)$ , then  $T_i(z)$  has itinerary  $(i d_0 d_1 d_2 \dots)$ , so we immediately see that the set of points in  $\Lambda$  is invariant under each map  $T_i$ .
  - Furthermore,  $\Lambda = \{(0 d_1 d_2 \dots)\} \cup \{(1 d_1 d_2 \dots)\} \cup \dots \cup \{(k d_1 d_2 \dots)\} = T_1(\Lambda) \cup \dots \cup T_k(\Lambda)$ .
  - Thus,  $\Lambda$  is invariant under the maps  $T_i$  and is also an invariant set. It is also straightforward to show that it is closed.
  - The uniqueness of  $\Lambda$  is a bit more involved, but it can be obtained from a theorem called the contraction mapping theorem (which says that a contraction on a sufficiently nice metric space like  $\mathbb{R}^n$  necessarily has a unique fixed point).
- In particular, if the maps  $T_1, \dots, T_k$  are similarities, this theorem says that there exists a unique closed set  $\Lambda$  that is equal to the union of each of the (smaller, similar) pieces  $T_i(\Lambda)$ .
  - This is precisely the kind of construction we have been using to define our examples of fractals. The theorem allows us to run the argument the other way around: for any collection of similarities, there exists an invariant set (which will usually be a fractal) associated to those similarities.
  - As an example, the invariant set for the iterated function system  $\{T_1, T_2\}$  on  $[0, 1]$ , where  $T_1(x) = \frac{1}{3}x$ ,  $T_2(x) = \frac{1}{3}x + \frac{2}{3}$ , is the Cantor ternary set. The theorem guarantees that the Cantor ternary set is the *only* set that can be pieced together from the smaller copies of itself generated by these two similarities.
  - As another example, the Sierpinski carpet is the invariant set for the iterated function system  $\{T_1, T_2, \dots, T_8\}$  on the square  $[0, 1] \times [0, 1]$ , where each function  $T_i$  has the form  $T_i(x, y) = \frac{1}{3}(x, y) + (a_i, b_i)$ , and  $(a_i, b_i)$  is the  $i$ th point in the list  $(0, 0), (0, \frac{1}{3}), (0, \frac{2}{3}), (\frac{1}{3}, 0), (\frac{1}{3}, \frac{2}{3}), (\frac{2}{3}, 0), (\frac{2}{3}, \frac{1}{3}), (\frac{2}{3}, \frac{2}{3})$ .

### 4.3.2 The Box-Counting Dimension of a Self-Similar Set

- We would like to describe when the invariant set of an iterated function system of similarities is a fractal, and how to compute its box-counting dimension.
- We require a technical definition:
- **Definition:** An iterated function system satisfies the open set condition if there exists an open set  $U \subset D$  such that the sets  $T_1(U), T_2(U), \dots, T_k(U)$  are all contained in  $U$  and are disjoint.
  - What this condition essentially says is: the images of the maps  $T_1, \dots, T_k$  do not overlap "too much".



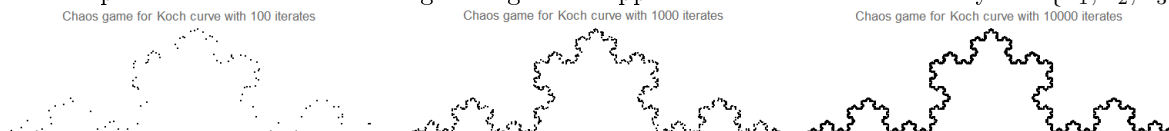
- **Theorem (Moran):** If  $S_1, \dots, S_k$  is an iterated function system of similarities with similarity constants  $c_1, \dots, c_k$  that satisfies the open set condition, then the box-counting dimension of the invariant set is the unique positive real number  $d$  satisfying  $1 = c_1^d + c_2^d + \dots + c_k^d$ .
  - As an immediate corollary, if all the similarity ratios are equal to a given constant  $c$  (where  $0 < c < 1$ ), then the dimension  $d$  is  $d = \frac{\log(k)}{\log(1/c)}$ , provided that the similarities do not overlap.
  - We will not give the proof of this theorem. It is not conceptually difficult, but the arguments are rather computationally messy even for the case of two similarities.
  - Essentially the idea is to construct a list of sets of the form  $T_{x_1}T_{x_2}\dots T_{x_n}(U)$  for appropriate sequences  $x_1, \dots, x_n$  that (i) contain the invariant set, (ii) can each be covered by a single box of size  $\epsilon$ , and (iii) cannot be covered by a box of size less than  $\epsilon/k$  for some fixed constant  $k$ .
  - Then one counts the number of such boxes based on the number and types of sequences  $x_1, \dots, x_n$ , and converts the counting problem into a problem of minimizing a particular expression, whose minimum value can be used to show that the dimension  $d$  satisfies the indicated relation.
- Using this theorem we can compute the box-counting dimension of each of the fractals we have constructed: the idea is to write down the similarity ratios  $c_1, \dots, c_k$  of each of the maps used to construct the fractal, and then solve the “Moran equation”  $1 = c_1^d + c_2^d + \dots + c_k^d$ .
- **Example:** Find the box-counting dimension of the open middle- $\alpha$  Cantor set, for  $0 < \alpha < 1$ .
  - The first stage of the construction replaces the interval  $[0, 1]$  with the union  $[0, (1 - \alpha)/2] \cup [(1 + \alpha)/2, 1]$ .
  - We can model this construction using an iterated function system  $\{T_1, T_2\}$  on the interval  $[0, 1]$ , where  $T_1(x)$  maps  $[0, 1]$  linearly onto  $[0, (1 - \alpha)/2]$  and  $T_2(x)$  maps  $[0, 1]$  linearly onto  $[(1 + \alpha)/2, 1]$ .
  - It is easy to see that we can take  $T_1(x) = \frac{1 - \alpha}{2}x$  and  $T_2(x) = \frac{1 - \alpha}{2}x + \frac{1 + \alpha}{2}$ . The open set condition is satisfied because the ranges of these functions do not overlap at all.
  - Thus, the similarity ratios for the two maps are both  $\frac{1 - \alpha}{2}$ .
  - The dimension therefore satisfies  $1 = \left(\frac{1 - \alpha}{2}\right)^d + \left(\frac{1 - \alpha}{2}\right)^d$ , from which we obtain  $\left(\frac{2}{1 - \alpha}\right)^d = 2$ .
  - Taking logarithms yields  $d = \boxed{\log_{2/(1-\alpha)} 2}$ .
  - Note that the box-counting dimension is always between 0 and 1. As  $\alpha \rightarrow 0$ , the dimension approaches 1, and as  $\alpha \rightarrow 1$ , the dimension approaches 0.
- **Example:** Find the box-counting dimension of the “second-quarter” Cantor set.
  - The first stage of the construction replaces the interval  $[0, 1]$  with the union  $[0, 1/4] \cup [1/2, 1]$ .
  - We can model this construction using an iterated function system  $\{T_1, T_2\}$  on the interval  $[0, 1]$ , where  $T_1(x) = \frac{1}{4}x$  and  $T_2(x) = \frac{1}{2}x + \frac{1}{2}$ . The open set condition is satisfied because the ranges of these functions do not overlap at all.
  - The similarity ratios are therefore  $\frac{1}{2}$  and  $\frac{1}{4}$ , so the Moran equation gives  $1 = \left(\frac{1}{2}\right)^d + \left(\frac{1}{4}\right)^d$ .
  - If we set  $2^d = y$ , then the equation is equivalent to  $1 = y^{-1} + y^{-2}$ , or  $y^2 - y - 1 = 0$ .
  - The positive solution is  $y = \frac{1 + \sqrt{5}}{2}$ , so the dimension is  $d = \boxed{\log_2 \left(\frac{1 + \sqrt{5}}{2}\right) \approx 0.694}$ .
  - Notice that the dimension of this set is slightly lower than the dimension of the middle-1/4 Cantor set, which (from the previous example) is equal to  $\log_{8/3} 2 \approx 0.707$ .
  - It is interesting, and perhaps counterintuitive, that merely shifting the location of the removed interval actually changes the dimension of the resulting fractal.

- Example: Find the box-counting dimension of the Koch curve.
  - The first stage of the construction replaces the line segment having endpoints  $(0, 0)$  and  $(1, 0)$  with four smaller line segments joining successively the points  $(0, 0)$ ,  $(\frac{1}{3}, 0)$ ,  $(\frac{1}{2}, \frac{\sqrt{3}}{6})$ ,  $(\frac{2}{3}, 0)$ , and  $(1, 0)$ .
  - We can model this construction using an iterated function system  $\{T_1, T_2, T_3, T_4\}$  on the the square  $[0, 1] \times [0, 1]$ , where  $T_1(x, y) = (\frac{x}{3}, \frac{y}{3})$ ,  $T_2(x, y) = (\frac{2+x-y\sqrt{3}}{6}, \frac{x\sqrt{3}+y}{6})$ ,  $T_3(x, y) = (\frac{3+x+y\sqrt{3}}{6}, \frac{\sqrt{3}-x\sqrt{3}+y}{6})$ , and  $T_4(x, y) = (\frac{x}{3} + \frac{2}{3}, \frac{y}{3})$ .
  - It is a fairly straightforward computation to see that each of these maps is a similarity with similarity ratio  $\frac{1}{3}$  that sends the line segment having endpoints  $(0, 0)$  and  $(1, 0)$  to the appropriate one of the four segments making up the next iterate of the curve.
  - The similarity ratios are all equal to  $\frac{1}{3}$  so the Moran equation gives  $1 = 4 \left(\frac{1}{3}\right)^d$ . Rearranging gives  $3^d = 4$ , so  $d = \frac{\log 4}{\log 3} = \boxed{\log_3 4 \approx 1.262}$ .
- Example: Find the box-counting dimension of the Sierpinski triangle.
  - The first stage of the construction replaces the full equilateral triangle with vertices at  $(0, 0)$ ,  $(1/2, \sqrt{3}/2)$ ,  $(1, 0)$  with three half-scale equilateral triangles.
  - We can model this construction using an iterated function system  $\{T_1, T_2, T_3\}$  on the region inside the triangle, where  $T_1(x, y) = \frac{1}{2}(x, y)$ ,  $T_2(x, y) = \frac{1}{2}(x, y) + (\frac{1}{2}, 0)$ , and  $T_3(x, y) = \frac{1}{2}(x, y) + (\frac{1}{4}, \frac{\sqrt{3}}{4})$ .
  - Note that as  $(x, y)$  ranges over the entire triangle,  $T_1$  ranges over the lower left quarter-triangle,  $T_2$  ranges over the lower-right quarter-triangle, and  $T_3$  ranges over the upper quarter-triangle.
  - The open set condition is satisfied (with  $U$  equal to the interior of the full triangle) since  $T_1(U)$ ,  $T_2(U)$ , and  $T_3(U)$  do not intersect.
  - The similarity ratios are therefore  $\frac{1}{2}$ ,  $\frac{1}{2}$ , and  $\frac{1}{2}$ , so the Moran equation gives  $1 = 3 \left(\frac{1}{2}\right)^d$ . Rearranging yields  $2^d = 3$ , so  $d = \frac{\log 3}{\log 2} = \boxed{\log_2 3 \approx 1.585}$ .
- Example: Find the box-counting dimension of the Sierpinski carpet
  - The first stage of the construction replaces the full square with 8 smaller squares at  $1/3$ -scale.
  - We can model this construction using an iterated function system  $\{T_1, T_2, \dots, T_8\}$  on the region inside the square, where each function  $T_i$  has the form  $T_i(x, y) = \frac{1}{3}(x, y) + (a_i, b_i)$ , where  $(a_i, b_i)$  runs through the 8 points  $(0, 0)$ ,  $(0, \frac{1}{3})$ ,  $(0, \frac{2}{3})$ ,  $(\frac{1}{3}, 0)$ ,  $(\frac{1}{3}, \frac{2}{3})$ ,  $(\frac{2}{3}, 0)$ ,  $(\frac{2}{3}, \frac{1}{3})$ ,  $(\frac{2}{3}, \frac{2}{3})$ .
  - The open set condition is satisfied (with  $U$  equal to the interior of the unit square) since the interiors of the eight smaller squares do not intersect.
  - The similarity ratios are all equal to  $\frac{1}{3}$  so the Moran equation gives  $1 = 8 \left(\frac{1}{3}\right)^d$ . Rearranging yields  $3^d = 8$ , so  $d = \frac{\log 8}{\log 3} = \boxed{\log_3 8 \approx 1.893}$ .
- Example: Find the box-counting dimension of the Menger sponge.
  - The first stage of the construction replaces the full cube with 20 smaller cubes at  $1/3$ -scale.

- We can model this construction using an iterated function system  $\{T_1, T_2, \dots, T_{20}\}$  on the region inside the cube, where each function  $T_i$  has the form  $T_i(x, y, z) = \frac{1}{3}(x, y, z) + \mathbf{v}_i$ , where  $\mathbf{v}_i$  is the corner of the corresponding subcube that is closest to the origin. This definition ensures that if  $C$  is the full cube, then  $T_i(C)$  is the  $i$ th of the smaller cubes for  $1 \leq i \leq 20$ .
- The open set condition is satisfied (with  $U$  equal to the interior of the cube) since the interiors of the small cubes do not intersect.
- The similarity ratios are all equal to  $\frac{1}{3}$  so the Moran equation gives  $1 = 20 \left(\frac{1}{3}\right)^d$ . Rearranging yields  $3^d = 20$ , so  $d = \frac{\log 20}{\log 3} = \boxed{\log_3 20 \approx 2.727}$ .

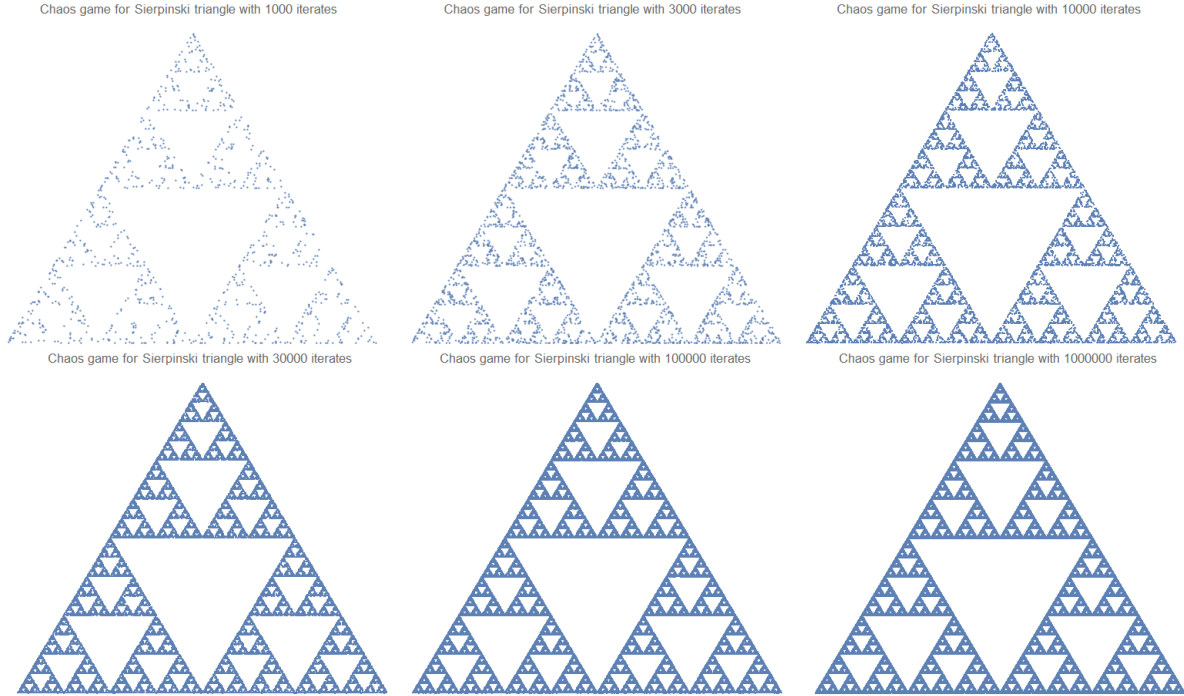
### 4.3.3 The “Chaos Game”

- Let us revisit the ideas from the theorem we proved about the existence of an invariant set for an iterated function system.
  - The main idea of the proof was that the invariant set  $\Lambda$  was the set of all points that can be written as a limit of a sequence of the form  $\lim_{n \rightarrow \infty} T_{x_0} \circ T_{x_1} \circ \dots \circ T_{x_n}(y)$ , where each  $x_i \in \{1, 2, \dots, k\}$  and  $y$  is an arbitrary element of  $D$ .
  - In particular, this suggests a computational method for finding this invariant set, namely: choose a starting point  $y$ , and then compute points of the form  $T_{x_0} \circ T_{x_1} \circ \dots \circ T_{x_n}(y)$  for a “random” sequence  $x_0, x_1, x_2, \dots$  of elements in  $\{1, 2, \dots, k\}$ .
- **Algorithm** (“Chaos Game”): To draw the invariant set  $\Lambda$  for an iterated function system  $\{T_1, \dots, T_k\}$  on a region  $D$ , choose an arbitrary  $y \in D$ , and then plot the points  $\{y, y_1, y_2, y_3, y_4, \dots\}$  where  $y_m$  is obtained by applying a randomly-chosen map from  $\{T_1, \dots, T_k\}$  to  $y_{m-1}$ .
  - It can be proven that, for almost every starting point  $y$  in  $D$  and sequence of maps  $T_{x_i}$ , the list of iterates  $\{y_1, y_2, y_3, y_4, \dots\}$  will be dense in  $\Lambda$ .
  - In particular, if we compute a large number of iterates and discard the first few, plotting the remainder should give a good approximation to the actual invariant set.
  - This also explains why the invariant set is sometimes called an attractor: the orbit of any point  $y$  under application of a random map  $T_i$  will (almost always) converge to the invariant set, in the sense that distance from the  $n$ th point in the orbit to the nearest point of the invariant set tends to zero as  $n \rightarrow \infty$ .
  - This algorithm is quite computationally efficient, since all that is needed is to apply a randomly-chosen function and keep track of the resulting list of points.
- **Example:** Use an iterated function system to plot the Koch curve.
  - The Koch curve is invariant under each of the maps  $\{T_1, T_2, T_3, T_4\}$  on the the square  $[0, 1] \times [0, 1]$ , where  $T_1(x, y) = \left(\frac{x}{3}, \frac{y}{3}\right)$ ,  $T_2(x, y) = \left(\frac{2+x-y\sqrt{3}}{6}, \frac{x\sqrt{3}+y}{6}\right)$ ,  $T_3(x, y) = \left(\frac{3+x+y\sqrt{3}}{6}, \frac{\sqrt{3}-x\sqrt{3}+y}{6}\right)$ , and  $T_4(x, y) = \left(\frac{x}{3} + \frac{2}{3}, \frac{y}{3}\right)$ .
  - Furthermore, applying each of these maps to the Koch curve produces four 1/3-scale copies of the curve whose union is the original curve.
  - Here are plots of iterates of the chaos game algorithm applied to the iterated function system  $\{T_1, T_2, T_3, T_4\}$ :



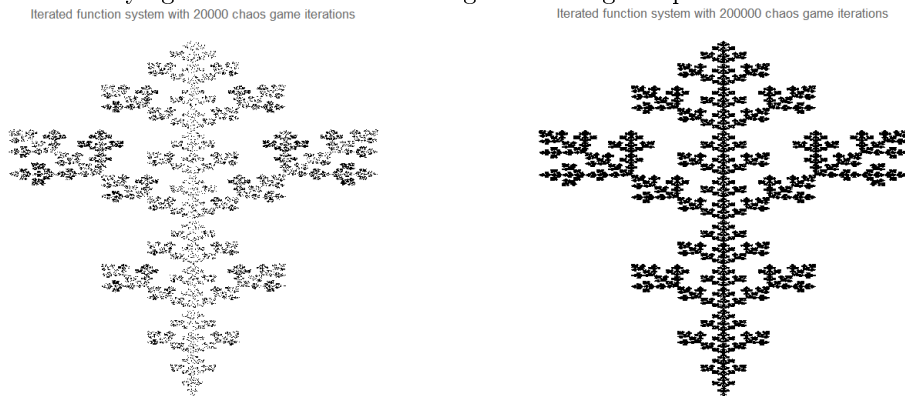
- **Example:** Use an iterated function system to plot the Sierpinski triangle.

- The Sierpinski triangle is invariant under each of the maps  $\{T_1, T_2, T_3\}$ , where  $T_1(x, y) = \frac{1}{2}(x, y)$ ,  $T_2(x, y) = \frac{1}{2}(x, y) + (\frac{1}{2}, 0)$ , and  $T_3(x, y) = \frac{1}{2}(x, y) + (\frac{1}{4}, \frac{\sqrt{3}}{4})$ .
- Furthermore, applying each of these maps to the Sierpinski triangle produces three 1/2-scale copies whose union is the original triangle.
- Here are plots of iterates of the chaos game algorithm applied to the iterated function system  $\{T_1, T_2, T_3\}$ :



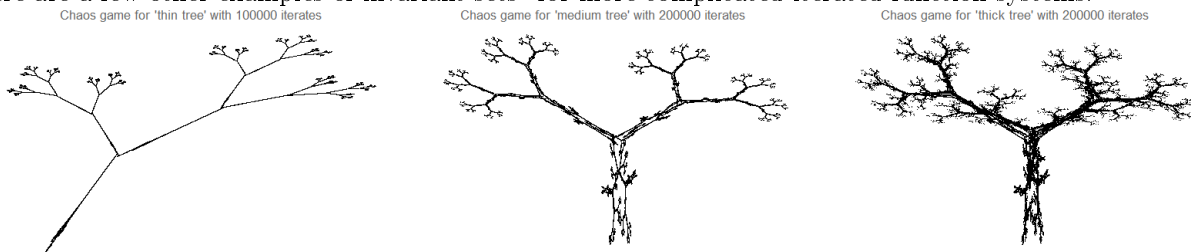
- We can also generate new fractals using iterated function systems and draw them using the chaos game procedure.
  - If we choose all the maps to be similarities that obey the open set condition, then we can even compute the box-counting dimension of the resulting sets.
- Example: Plot the fractal that is the invariant set for the iterated function system  $\{T_1, T_2, T_3, T_4\}$  with  $T_1(x, y) = \left(\frac{x+y}{4}, \frac{1-x+y}{4}\right)$ ,  $T_2(x, y) = \left(\frac{x-y}{4}, \frac{1+x+y}{4}\right)$ ,  $T_3(x, y) = \left(\frac{x}{2}, \frac{y}{2}\right)$ ,  $T_4(x, y) = \left(\frac{x}{2}, \frac{2+y}{2}\right)$ , and then find its box-counting dimension.

- Here are plots with varying numbers of iterates using the chaos game procedure:

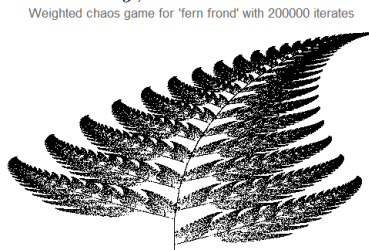


- It is straightforward to check that the maps  $T_1$  and  $T_2$  are similarities with scaling factor  $\frac{\sqrt{2}}{4}$  and the maps  $T_3$  and  $T_4$  are similarities with scaling factor  $\frac{1}{2}$ . It is also straightforward to verify that the open set condition holds.
- Thus, the Moran equation yields  $1 = 2 \left(\frac{\sqrt{2}}{4}\right)^d + 2 \left(\frac{1}{2}\right)^d$ .
- It is not easy to give an exact solution to this equation: even if we set  $2^{d/2} = y$  the equation becomes  $1 = 2y^{-3} + 2y^{-2}$ , or  $y^3 - 2y - 2 = 0$ . The exact solution is rather unpleasant:  $y = \frac{1}{3} \left( \sqrt[3]{27 + 3\sqrt{57}} + \sqrt[3]{27 - 3\sqrt{57}} \right)$
- It is much easier to compute the approximate solution, however: Newton's method gives  $y \approx 1.7693$ , from which we obtain  $d \approx \boxed{1.646}$ .

- Here are a few other examples of invariant sets<sup>4</sup> for more complicated iterated function systems:



- Each picture is for an iterated function system whose functions are of the form  $T(x) = (ax + by + e, cx + dy + f)$ .
- The first picture is for the iterated function system  $\{T_1, T_2, T_3\}$  with  $T_1(x, y) = (-0.387x + 0.43y + 0.522, 0.43x + 0.387y + 0.256)$ ,  $T_2(x, y) = (-0.322x - 0.009y + 0.5059, -0.091x + 0.441y + 0.4219)$ , and  $T_3(x, y) = (0.015x - 0.113y + 0.4, 0.02x - 0.468y + 0.4)$ .
- The second picture is for the iterated function system  $\{T_1, T_2, T_3, T_4\}$  with  $T_1(x, y) = (0.462x + 0.414y + 0.2511, -0.252x + 0.361y + 0.5692)$ ,  $T_2(x, y) = (0.195x - 0.488y + 0.4431, 0.344x + 0.443y + 0.2452)$ ,  $T_3(x, y) = (-0.058x - 0.07y + 0.5976, 0.453x - 0.111y + 0.0969)$ , and  $T_4(x, y) = (-0.035x + 0.07y + 0.4884, -0.469x - 0.022y + 0.5069)$ .
- The third picture is for the second iterated function system, along with a fifth map  $T_5(x, y) = (0.8562 - 0.637x, 0.501y + 0.2513)$ .
- We also mention that it is possible to make more complicated shapes using a “weighted chaos game”, where instead of choosing one of the maps  $T_i$  with equal probability, we choose them randomly according to some weighted (possibly unequal) probabilities.
  - Here is one such fractal<sup>5</sup> arising from the iterated function system  $\{T_1, T_2, T_3, T_4\}$  with weights  $\{0.01, 0.85, 0.07, 0.07\}$ , having  $T_1(x, y) = (0, 0.16y)$ ,  $T_2(x, y) = (0.85x + 0.04y, 1.6 - 0.04x + 0.85y)$ ,  $T_3(x, y) = (0.2x - 0.26y, 1.6 + 0.23x + 0.22y)$ , and  $T_4(x, y) = (-0.15x + 0.28y, 0.44 + 0.26x + 0.24y)$ :



Well, you're at the end of my handout. Hope it was helpful.

Copyright notice: This material is copyright Evan Dummit, 2015-2023. You may not reproduce or distribute this material without my express permission.

<sup>4</sup>Thanks to Bernard Vuilleumier and the Wolfram Demonstrations Project for providing the code used to create these graphics.

<sup>5</sup>This example originally from M. Barnsley's text *Fractals Everywhere*.