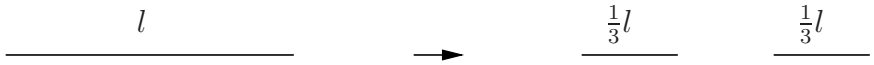
Lecture 27

Measuring the sizes of "fractal" sets in $\mathbb{R}^k$

Let's recall the construction of the ternary Cantor set $C \subset [0,1]$ discussed in Lecture 24. Recall that C was produced by the repeated application of a "middle-thirds dissection" procedure, shown on Pages 3.62 and 3.63: From the set $C_0 = [0,1]$, we remove the open set $(\frac{1}{3}, \frac{2}{3})$ to produce the set $C_1 = [0, \frac{1}{3}] \cup [\frac{2}{3}, 1]$. From C_1 we remove the open sets $(\frac{1}{9}, \frac{2}{9})$ and $(\frac{7}{9}, \frac{8}{9})$ to produce $C_2 = [0, \frac{1}{9}] \cup [\frac{2}{9}, \frac{1}{3}] \cup [\frac{2}{3}, \frac{7}{9}] \cup [\frac{8}{9}, 1]$. We may consider this process as the repeated application of a "generator" G that operates on sets of points. Think of the generator as a kind of computer which is equipped with a pattern recognition device. It searches an input object for line segments. When it finds a line segment, regardless of its length, it removes the middle thirds of it. In summary the generator G may be represented by the procedure



The middle thirds dissection procedure associated with the Cantor set may then be considered as a dynamical system involving the generator G and its action on sets iteratively. We use the input "seed" set, $C_0 = [0,1]$ and then define the iteration sequence,

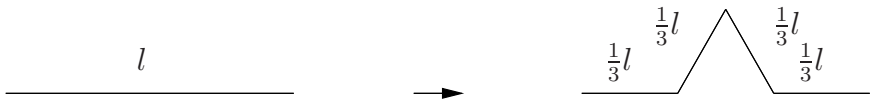
$$C_{n+1} = G(C_n)\,, \quad n \geq 0\,. \tag{1}$$

Referring to the diagrams in Lecture 24, you should be able to see the following,

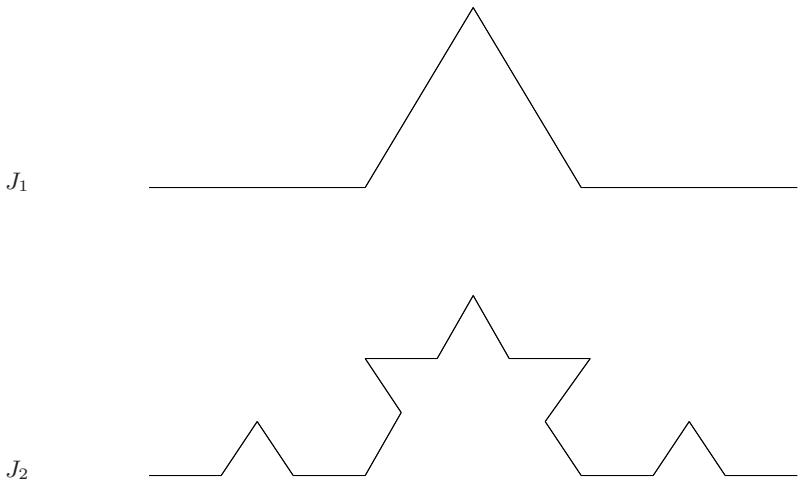
$$\begin{aligned} C_1 &= G(C_0) = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right] \\ C_2 &= G(C_1) = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right] \\ &\vdots \end{aligned} \tag{2}$$

For an $n \geq 1$, the set C_n is composed of 2^n intervals of length $\frac{1}{3^n}$. In the limit, the sets C_n converge to the Cantor set C. We shall return to this set in the near future for the purpose of assigning a fractal dimension D to it.

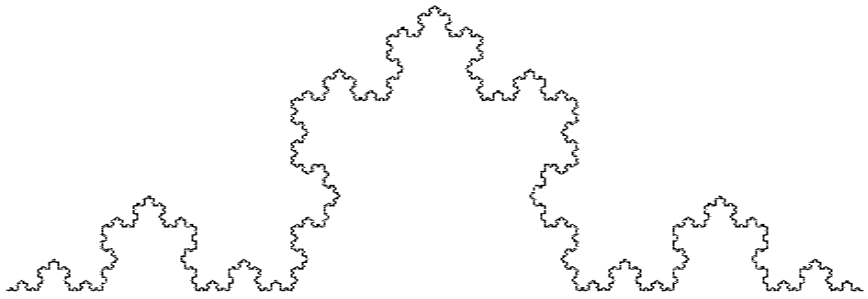
Now consider the following generator that produces a limiting curve in $\mathbb{R}^2$:



This generator removes the middle third of a line segment of length l and replaces it with two segments of length $\frac{l}{3}$. If we begin with the seed $J_0 = [0,1]$, the next two sets in the iteration sequence $J_{n+1} = G(J_n)$ are sketched below.



We state, without proof, that the sets J_n converge to a very complicated curve in $\mathbb{R}^2$, the so-called "von Koch curve," which we shall denote as C, shown below.



The von Koch curve C

The curve C is a very irregular curve – an example of a **fractal curve**. It is a **self-similar set** in the sense that C is a union of four contracted copies of itself, C_i, $1 \leq i \leq 4$. Each copy C_i can be obtained by contracting C, say toward its left-most point (here considered as the origin of our coordinate system), and then translating and/or rotating the contracted copy appropriately. It is not surprising that the positions of the four copies C_i are determined by the four components of the piecewise linear curve $J_1 = G(J_0)$ shown earlier.

The copying property of the curve C can be continued: Since each subset C_i may be expressed as a union of four contracted copies C_{ij} of itself, the original curve C may be expressed as a union of sixteen contracted copies of itself.

Let us now try to measure the length of the von Koch curve, using the method of ϵ-measuring

rods discussed in the previous section. We start at the left endpoint of C and place these rods end to end until we reach the right endpoint of C. A most natural choice for the lengths of these rods, with increasing refinement, i.e. $\epsilon \to 0$, is the sequence $\epsilon_n = (\frac{1}{3})^n$, with $n \to \infty$.

If we use a measuring rod of length $\epsilon_0 = 1$, then clearly only one such rod is needed to get us from the left endpoint of C to the right endpoint. Clearly, this length estimate, $L_0 = 1$, is a very gross one since it ignores the quite significant middle "inlet" of the curve, not to mention all the tinier "nooks and crannies" that are present in the curve at all sizes. If we use a measuring rod of length $\epsilon_1 = \frac{1}{3}$, then four rods are necessary to get us from the left endpoint of C to the right endpoint, yielding a length estimate of $L_1 = \frac{4}{3}$. This rod at least detects the middle "inlet." Nevertheless, it cannot detect any smaller inlets. It should be clear that $4^2 = 16$ measuring rods of length $\epsilon_2 = (\frac{1}{3})^2$ are necessary to cover C, yielding a length estimate of $L_2 = \frac{16}{9} = (\frac{4}{3})^2$.

The reader should notice that the "covering" of C with ϵ_n-measuring rods yields length estimates L_n that are equivalent to the lengths of the curves J_n obtained by the repeated action of the generator G above. The results are summarized below:

Length of measuring rod: $\epsilon_n = \frac{1}{3^n}$
Number of measuring rods needed to cover C: $N(\epsilon_n) = 4^n$
Estimate of length: $L_n = N(\epsilon_n)\epsilon_n = (\frac{4}{3})^n$.

Therefore, we conclude that the length of C is given by

$$L^1(C) = \lim_{n\to\infty} L_n = \infty, \tag{3}$$

i.e. **the von Koch curve C has infinite length**. In retrospect, we are not surprised since the generator essentially increases the length of each curve J_k by a factor of $\frac{4}{3}$ to produce the curve J_{k+1}. However, the fact that C has infinite length is quite intriguing. From a look at the curve C in terms of how it was constructed, i.e., as a sequence of curves J_n as shown in the figure on the previous page, one can deduce that C is **non-rectifiable**, i.e., it has no tangent vectors. Therefore, it cannot be parametrized in the usual way, i.e., $(x(t), y(t))$, $a \le t \le b$, where $x(t)$ and $y(t)$ are piecewise smooth. One cannot obtain a finite length via standard methods of first- or second-year Calculus.

In fact, the results summarized above already indicate that something "non-Euclidean" is going on here. Note the relationship between consecutive numbers of ϵ_n-rods necessary to cover C:

$$N(\epsilon_{n+1}) = 4N(\epsilon_n). \tag{4}$$

For any n, we may set $\epsilon = \epsilon_n$ and $\epsilon' = r\epsilon$, where $r = \frac{1}{3}$, to obtain the relation

$$\boxed{N\left(\frac{\epsilon}{3}\right) = 4N(\epsilon)\,.} \tag{5}$$

From this equation, we see that **the von Koch curve cannot be a one-dimensional set**. If it were, then the factor "4" would have to be replace by "3". To see this, recall the the following scaling equation for one-dimensional sets that was presented in the previous lecture (Lecture 26, Eq. (69)),

$$N(\epsilon') = N(r\epsilon) = N(\epsilon)r^{-1}\,. \tag{6}$$

In this case $r = \frac{1}{3}$ so that $r^{-1} = 3$. Once again, we suspect that something "non-Euclidean" is going on here.

One may well wonder whether these deviations are due to the fact that the von Koch "curve" C may have some area, due to its complicated jaggedness. If C were a two-dimensional object, then the number of ϵ-tiles, $N(\epsilon)$, would have to scale according to the following equation for two-dimensional objects, presented in the previous lecture (Lecture 26, Eq. (73)),

$$N(\epsilon') = N(r\epsilon) = N(\epsilon)r^{-2}\,. \tag{7}$$

From this equation, since $r = \frac{1}{3}$, the the factor "4" would have to be replaced by "9."

We therefore know that C **cannot be one-dimensional nor can it be two-dimensional**. But let's put this aside for a moment and attempt to compute the the area of C using the method of ϵ-tiles of area ϵ^2 outlined earlier. The reader may worry that the covering of C with square tiles of size $\epsilon_n \times \epsilon_n$ will yield gross overestimates of the area, since the tiles overlap significantly with each other. (One can choose to cover C, not with squares but with **circular discs** of diameter ϵ_n. The resulting estimate will differ only by a factor of $\frac{\pi}{4}$. Why?) Our estimate will be, up to a constant factor,

$$L^2(C) \cong N(\epsilon_n)\epsilon_n^2. \tag{8}$$

A little reflection will reveal that the number of ϵ_n-sided square tiles or ϵ_n-diameter circular discs required to cover C is identical to the number $N(\epsilon_n)$ of ϵ_n-rods needed to "cover" C when we were approximating its area. Therefore, we have the following results:

Area of disc (up to factor K): $\epsilon_n^2 = (\frac{1}{3^n})^2$
Number of discs needed to cover C: $N(\epsilon_n) = 4^n$
Estimate of area: $A_n = N(\epsilon_n)\epsilon_n^2 = (\frac{4}{9})^n$.

We conclude that the area of C is given by

$$L^2(C) = \lim_{n\to\infty} L_n = 0. \tag{9}$$

Even an overestimate of the area yields a result of zero. In summary, we have found that

the von Koch curve C has infinite length yet zero area!

Doesn't that look familiar to something that was discussed at the end of the previous lecture?

An explanation of these two results, in terms of our measuring methods, is rather simple. In both cases, $N(\epsilon_n) = 4^n$ measuring rods or discs are required to cover C at each resolution. In our attempts to determine the length and area of C, we encountered the following limits, respectively,

$$\left(\frac{4}{3}\right)^n \to \infty \qquad \text{and} \qquad \left(\frac{4}{3^2}\right)^n \to 0, \qquad \text{as} \qquad n \to \infty. \tag{10}$$

In the first case, the factor "3" is not large enough to match the factor "4" to produce a noninfinite limit. In the second case, the factor "3" is raised to too high a power, i.e. 3^2, to produce a nonzero limit. One may suspect that if we were allowed to raise the factor "3" to a suitable power, say $1 < D < 2$, we could "steer" the above limit away from both zero and infinity. In fact, if we choose D so that

$$3^D = 4\,, \tag{11}$$

then the ratio in Eq. (10) becomes

$$\frac{4}{3^D} = 1 \tag{12}$$

so that

$$\lim_{n\to\infty} \frac{4}{3^D} = \lim_{n\to\infty} 1^n = 1\,. \tag{13}$$

We may easily solve for D in Eq. (11) by taking logarithms of both sides. Any logarithm will do, but we'll take natural logarithms,

$$\ln(3^D) = \ln 4 \quad \Longrightarrow \quad D\,\ln 3 = \ln 4\,, \tag{14}$$

which implies that

$$D = \frac{\ln 4}{\ln 3} \approx 1.26\,. \tag{15}$$

Since the dimension D lies between 1 and 2, we may consider the von Koch curve as being a set which is "thicker" than a curve but "thinner" than a region in the plane.

This looks like a rather questionable procedure but it is, in fact, the basis for assigning fractional dimensions as we now discuss.

What we found above with the von Koch curve is generally the situation encountered with fractal sets S that are "embedded" or constructed in the underlying space $\mathbb{R}^k$, k integer - in this case $k = 2$, i.e. the plane. It is typically found that for some integer $k \geq 1$,

$$L^{k-1}(S) = \infty, \qquad L^k(S) = 0. \tag{16}$$

In other words, the set S is thicker than a $(k-1)$-dimensional set, but thinner than a k-dimensional set. Its "dimension" lies somewhere between $k-1$ and k.

This motivated mathematicians to consider the following "q-dimensional measure" of a set S, where $q \in \mathbb{R}$ is allowed to be any real number:

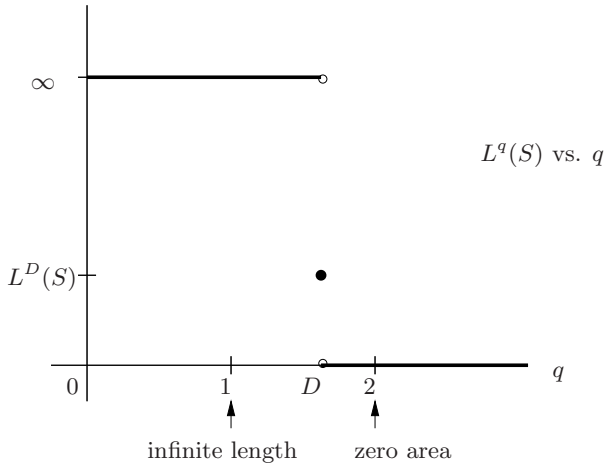
$$L^q(S) = \lim_{\epsilon \to 0} N(\epsilon)\, \epsilon^q \,. \tag{17}$$

Here, $N(\epsilon)$ is the number of lines/tiles/cubes/whatever needed to cover the set S.

We now state, without proof, that for "self-similar" fractal sets $S \in \mathbb{R}^k$ such as the Cantor set and the von Koch curve the following holds: There exists a value D, $k-1 \le D \le k$, such that

$$L^q(S) = \infty, \quad q < D, \qquad L^q(S) = 0 \quad q > D, \qquad q \in \mathbb{R}. \tag{18}$$

We shall call D the "fractal dimension" (or "capacity dimension") of the fractal set S. This generic behaviour of the q-dimensional measure function L^q is sketched below.



To summarize: **The fractal set S is "thicker" than any set of dimension $q < D$ and "thinner" than any set of dimension $q > D$. It "scales" like an object with dimension D.**

Note that Eq. (18) does not say anything about the D-dimensional measure $L^D(S)$ of the set S. This is because a nonzero, finite value for $L^D(S)$ is not always guaranteed. For many "nice" fractals, however, such as the ones that we shall be examining, a nonzero and finite value can be assigned. And, as you'll see below, we'll often assume that a finite value does exist. But even so, we are generally not as interested in the measure $L^D(S)$ of the set as we are in its dimension D which, from the discussion above, is a kind of "boundary point" between "thickness" and "thinness" of the set S.

And one additional note: When the set S is "nonfractal," e.g., a line segment, smooth curve, planar area, etc., then the value D coincides with its topological dimension, i.e., the "fractal dimension" D becomes the topological dimension of the set. Technically, we should have omitted the word "fractal" in our original definition of D earlier, but it was desired to emphasize the use of D to characterize

fractal sets.

Now recall the example given at the end of the previous section, in which we considered a subset S of the plane that had finite area. The structure of S - a simple planar, "Euclidean," region - was such that the dimension $D = 2$ was appropriate to describe its "size." We could work with integer values to zero in on its dimension, determining that its length ($q = 1$) was infinite and its volume ($q = 3$) was zero. However, integer values of dimension are not enough to analyze fractal sets such as the von Koch curve - we must go to non-integer values to zero in on an appropriate "dimension" of the set.

For most self-similar fractal sets S, their appropriate "size" is then given by the D-dimensional volume or measure

$$L^D(S) = \lim_{\epsilon \to 0} N(\epsilon)\epsilon^D \,, \tag{19}$$

assuming that the limit exists. Let's now see the consequence of this result. For ϵ sufficiently close to zero, the D-dimensional volume or measure of a set S is well approximated as follows,

$$L^D(S) = N(\epsilon)\epsilon^D = N(r\epsilon)(r\epsilon)^D \,. \tag{20}$$

If we divide both sides by $(r\epsilon)^D$, we obtain the following important result,

$$\boxed{N(r\epsilon) = N(\epsilon)r^{-D} \,,} \tag{21}$$

which is a scaling relationship between the number of ϵ-tiles and the number of $r\epsilon$-tiles needed to cover set S. Note that this is a non-integer generalization of the following scaling relationship observed for Euclidean sets, presented as Eq. (77) in the previous lecture (Lecture 26),

$$\boxed{N(r\epsilon) = N(\epsilon)r^{-k} \,,} \tag{22}$$

for k an integer.

Return to von Koch curve: In this case $r = \frac{1}{3}$. If $N(\epsilon)$ is the number of ϵ-tiles needed to cover the von Koch curve C, then the number of $N(r\epsilon) = N(\frac{1}{3}\epsilon)$ tiles needed to cover C is $4N(\epsilon)$. Eq. (21) then becomes

$$N\left(\frac{1}{3}\epsilon\right) = 4N(\epsilon) = N(\epsilon)\left(\frac{1}{3}\right)^{-D} \,. \tag{23}$$

Division of both sides of the second equality by $N(\epsilon)$ and a little rearrangement yields

$$3^D = 4 \quad \Longrightarrow \quad D = \frac{\ln 4}{\ln 3} \,, \tag{24}$$

which is in agreement with the value of D in Eq. (15).

Return to ternary Cantor set C on [0,1]: In Section 3.7, we found that its length was zero. Recall that this followed from a look at the lengths of the sets J_n that converge to C in the limit $n \to \infty$: $L^1(J_n) = \frac{2}{3}^n$. In fact these are precisely the length estimates of the Cantor set that would be obtained if we used ϵ measuring rods of length $\epsilon_n = \frac{1}{3^n}$. In summary:

Length of measuring rod: $\epsilon_n = \frac{1}{3^n}$
Number of measuring rods needed to cover C: $N(\epsilon_n) = 2^n$
Estimate of length: $L_n = N(\epsilon_n)\epsilon_n = (\frac{2}{3})^n$.

It follows that the length of C is given by

$$L^1(C) = \lim_{n\to\infty} L_n = 0. \tag{25}$$

It now remains to determine the fractal dimension of the Cantor set C. Once again, we begin with the fact that the number of square tiles of length $\epsilon_n = \frac{1}{3^n}$ needed to cover C is the same as the number of rods we used above: $N(\epsilon_n) = 2^n$. We may now proceed in either of two ways (each of which, of course, is related to the other):

1. From $N(\epsilon_n) = 2^n$, it follows that

$$\begin{aligned} N(\epsilon_{n+1}) &= N(\frac{1}{3}\epsilon_n) \\ &= 2N(\epsilon_n) \end{aligned} \tag{26}$$

Insert these results into Eq. (21):

$$N\left(\frac{1}{3}\epsilon\right) = 2N(\epsilon) = N(\epsilon)\left(\frac{1}{3}\right)^{-D}. \tag{27}$$

Division of both sides of the second equality by $N(\epsilon)$ and a little rearrangement yields

$$3^D = 2 \quad \Longrightarrow \quad D = \frac{\ln 2}{\ln 3} \approx 0.63\,. \tag{28}$$

2. Since the q-dimensional measure of C, $q \in \mathbb{R}$, will be given by

$$\begin{aligned} L^q(C) &= \lim_{n\to\infty} N(\epsilon_n)(\epsilon_n)^q \\ &= \lim_{n\to\infty} \left(\frac{2}{3^q}\right)^n, \end{aligned} \tag{29}$$

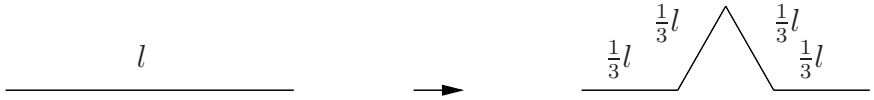
it follows that the critical value for q occurs when $3^q = 2$ or $q = \frac{\ln 2}{\ln 3}$. When $q < \frac{\ln 2}{\ln 3}$, $3^q < 2$ so that $\frac{2}{3^q} > 1$ which implies that $L_q(C) = \infty$. When $q > \frac{\ln 2}{\ln 3}$, $3^q > 2$ so that $\frac{2}{3^q} < 1$ which implies that $L_q(C) = 0$. Therefore, $D = \frac{\ln 2}{\ln 3}$, in agreement with the first method.

The fact that the fractal dimension D of the Cantor set lies between 0 and 1 suggests that it is "thicker" than a set of points but "thinner" than a curve/line.

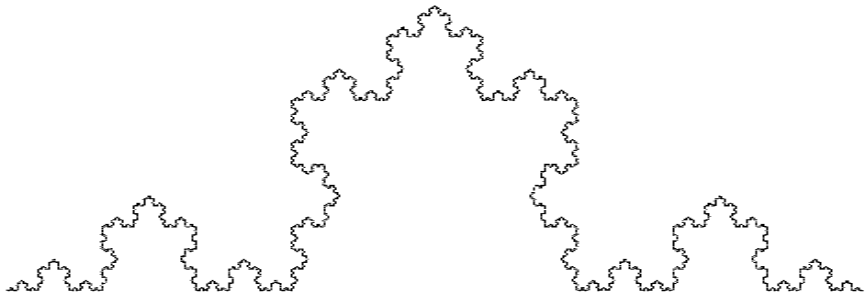
A summary of the above results

A summary is in order, in an effort to get an idea of the "big picture."

- When the following generator, is applied repeatedly to the unit interval $C_0 = [0, 1]$ the resulting



  sequence of sets C_n defined by the iteration sequence $C_{n+1} = G(C_n)$, $n = 0, 1, 2, \cdots$, converges to the following set,



  known as the **von Koch curve**. The length of the von Koch curve was found to be infinite, and its area was found to be zero. Using a scaling argument, we found its fractal dimension to be

$$D = \frac{\ln 4}{\ln 3} \approx 1.26\,. \tag{30}$$

  Since $D > 1$, this set may be viewed as "thicker" than a curve but "thinner" than a region in the plane.

  Finally, the von Koch curve may be viewed as **a union of four copies of itself**, each copy of which is obtained by shrinking the von Koch curve by a factor $r = \frac{1}{3}$ in each direction.

- When the following generator,

  is applied repeatedly to the unit interval $J_0 = [0, 1]$ the resulting sequence of sets J_n defined by the iteration sequence $J_{n+1} = G(J_n)$, $n = 0, 1, 2, \cdots$, converges to the following set,

known as the **ternary Cantor set**. The length of the Cantor set was found to be zero. Using a scaling argument, we found its fractal dimension to be

$$D = \frac{\ln 2}{\ln 3} \approx 0.63\,. \tag{31}$$

Since $D < 1$, this set may be viewed as "thinner" than a curve, which has topological dimension one. Furthermore, since $D > 0$, this set may be viewed as "thicker" than a set of points which have topological dimension zero.

Finally, the von Koch curve may be viewed as **a union of two copies of itself**, each copy of which is obtained by shrinking the von Koch curve by a factor $r = \frac{1}{3}$ in each direction.

Let us now consider the following generator, which may be viewed as a kind of "middle road" between the two generators above – a generator which neither adds nor removes sets from a set:

Yes, this may be viewed as a generator: It simply returns the "middle-thirds" part of an interval that the Cantor set generator removed. In effect, it represents the identity operator. It simply maps the unit interval $J_0 = [0, 1]$ to itself. In the limit, the sequence of sets $J_{n+1} = G(J_n)$ converges trivially to the interval $[0, 1]$. Of course, the dimension of this set is $D = 1$. But let's derive this result using the scaling associated with this generator. (We actually did this in the previous lecture.)

We view the unit interval $[0, 1]$ as **a union of three copies of itself**, each copy of which is obtained by shrinking the interval by a factor of $r = \frac{1}{3}$. Let $N(\epsilon)$ be the number of ϵ-sticks needed to cover the unit interval $[0, 1]$. If we now use measuring sticks of length $\frac{1}{3}\epsilon$, we need $3N(\epsilon)$ sticks to cover the interval, i.e.,

$$N\left(\frac{1}{3}\epsilon\right) = 3N(\epsilon)\,. \tag{32}$$

We now use the scaling result derived earlier,

$$N(r\epsilon) = N(\epsilon) r^{-D}\,, \tag{33}$$

with $r = \frac{1}{3}$ in Eq. (32):

$$N\left(\frac{1}{3}\epsilon\right) = 3N(\epsilon) = N(\epsilon)\left(\frac{1}{3}\right)^{D}\,. \tag{34}$$

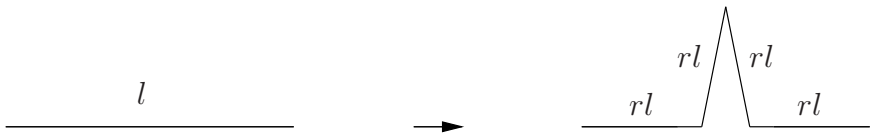
Dividing both sides of the final equation by $N(\epsilon)$ and rearranging yields

$$3^D = 3 \quad \Longrightarrow \quad D = 1\,. \tag{35}$$

Lecture 28

Measuring the sizes of fractal sets (cont'd)

We now consider a generalization of the von Koch curve construction as defined by the following generator:



This generator, to be denoted as G_r in order to explicitly state its dependence upon the parameter r, replaces a line segment of length l with four shorter line segments of length rl as shown above. The feasible range of the parameter is $\frac{1}{4} \leq r \leq \frac{1}{2}$. When $r = \frac{1}{4}$, the line segment is simply replaced by an identical line segment, so we expect the limiting curve C_r to be the original seed line segment. The parameter value $r = \frac{1}{3}$ corresponds to the standard von Koch curve discussed earlier.

We now consider the limiting set of points obtained from the iteration of G_r. Without loss of generality, we assume that our initial "seed" set is $I_0 = [0, 1]$ from which the sequence $I_{n+1} = G_r(I_n)$ is then generated. In the limit $n \to \infty$, the sequence of piecewise linear curves I_n converges to a curve C_r. Some limiting sets for various r values are shown on the next page. As r increases, the curves C_r demonstrate a greater degree of jaggedness or "fractality." In each case, C_r is self-similar and may be expressed as a union of four contracted copies of itself:

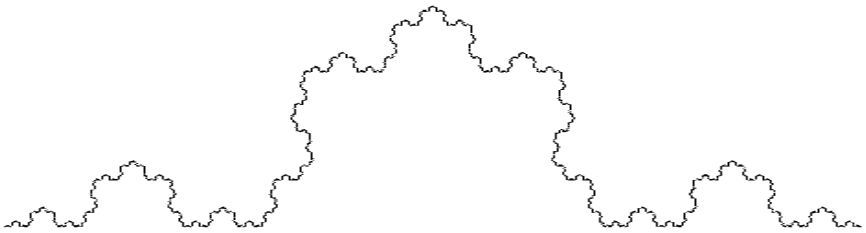
$$C_r = \bigcup_{k=1} C_{r,k}. \tag{36}$$

Each copy $C_{r,k}$ is obtained by shrinking C_r toward the origin (0,0) by a factor of r and then rotating and translating it appropriately. Perhaps the most striking limiting "curve" C_r occurs in the case $r = \frac{1}{2}$. The limiting set is, in fact, the triangular region shown on the next page. Our curve C_r is, in fact, a "space-filling" curve. The reader is encouraged to work out the first few iterates I_n to see how these sets approach the limiting area. (See the book, *The Fractal Geometry of Nature* by Benoit Mandelbrot, for more examples of generators that yield space-filling curves.)

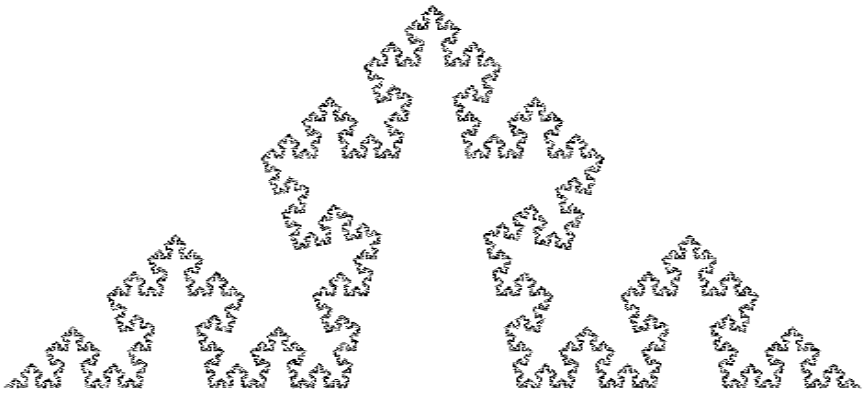
The next step is to determine the fractal dimension D of the limiting curve C_r. The self-similar nature of C_r described above will be very useful here. For an $\epsilon > 0$, suppose that we can tile the curve C_r with $N(\epsilon)$ circular discs of diameter ϵ. Also assume that ϵ is sufficiently small so that we may make the approximation

$$L^D(C_r) \approx N(\epsilon)\epsilon^D \tag{37}$$

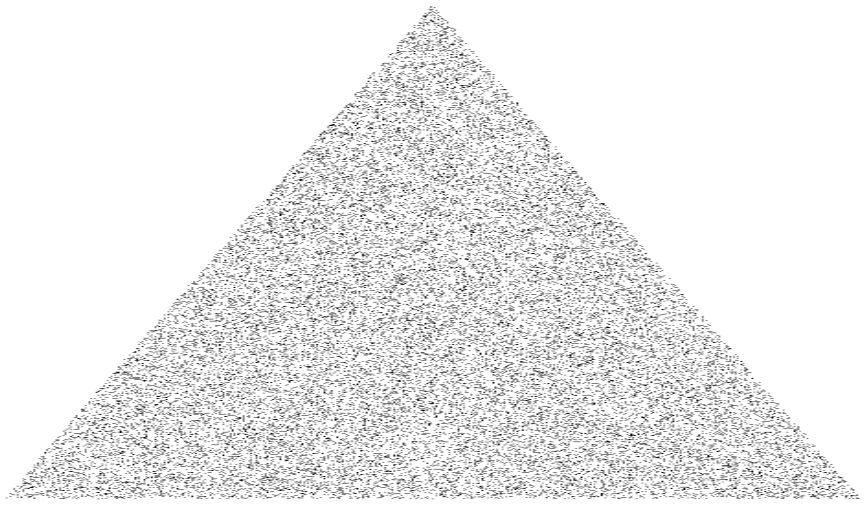
(This approximation may be made as closely as desired by choosing ϵ sufficiently close to 0.) We sketch this covering below.



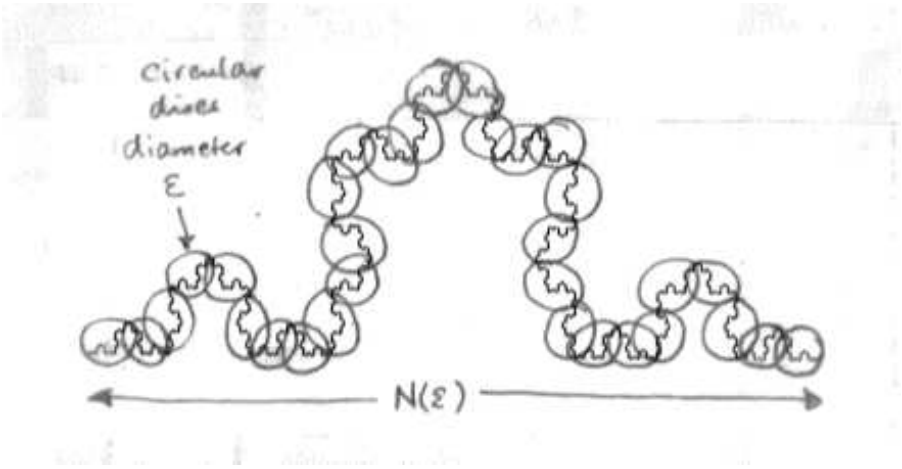
Generalized von Koch curve: $r = 0.3$



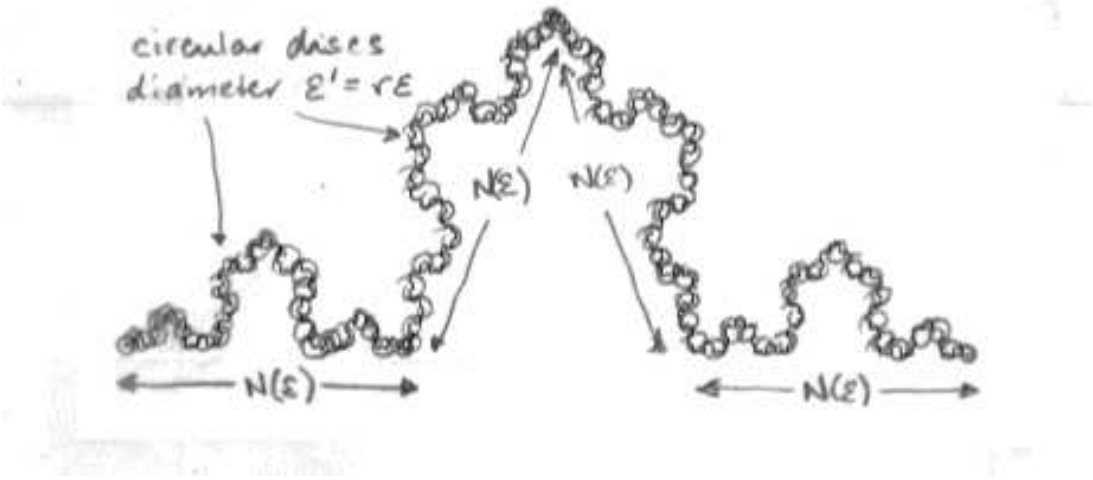
Generalized von Koch curve: $r = 0.4$



Generalized von Koch curve: $r = 0.5$



Now construct four shrunken copies of the above figure, i.e. the curve C_r as well as the $N(\epsilon)$ discs, with contraction factor r, and arrange these copies so that the the curve C_r is reconstructed. The resulting figure, shown below, shows curve C_r being covered with $4N(\epsilon)$ circular discs of diameter $r\epsilon$.



The D-dimensional volume of C_r may therefore be approximated as

$$L^D(C_r) \approx 4N(\epsilon)(r\epsilon)^D \tag{38}$$

We now equate the two D-dimensional volume estimates,

$$N(\epsilon)\epsilon^D = 4N(\epsilon)(r\epsilon)^D \,. \tag{39}$$

Dividing both sides by $N(\epsilon)\epsilon^D$ yields the relation

$$4r^D = 1 \,, \tag{40}$$

which may be rewritten as

$$r^{-D} = 4 \quad \Longrightarrow \quad \left(\frac{1}{r}\right)^D = 4 \,. \tag{41}$$

Solving for D yields

$$\boxed{D_r = \frac{\ln 4}{\ln \frac{1}{r}}, \qquad \frac{1}{4} \leq r \leq \frac{1}{2}.} \tag{42}$$

This is the fractal dimension of the generalized von Koch curve C_r.

We could also have derived the above result in a quick way by using our earlier scaling equation, which we repeat here:

$$N(r\epsilon) = N(\epsilon) r^{-D}\,. \tag{43}$$

The number of $N(r\epsilon)$ balls needed to cover the set is 4 times the number of $N(\epsilon)$ balls, i.e.,

$$N(r\epsilon) = 4N(\epsilon)\,. \tag{44}$$

Substituting this into (43) yields,

$$4N(\epsilon) = N(\epsilon) r^{-D}\,, \tag{45}$$

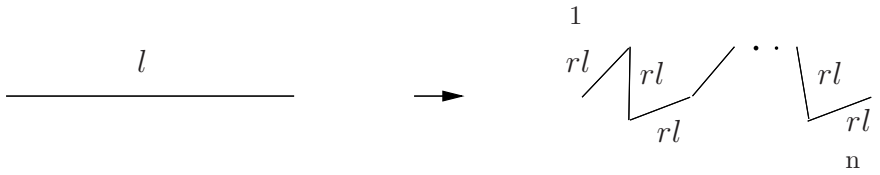
which yields (40), from which we obtain the result in (42).

Special cases of D_r in Eq. (42):

1. $r = \frac{1}{4}$: $D = 1$; the generator replaces a line with itself.
2. $r = \frac{1}{2}$: $D = 2$; the limiting set is the triangular region shown in the figure.

Further generalizations of the von Koch curve generator

The factor "4" appearing in the expression for the fractal dimension D_r is due to the fact that C_r is composed of four contracted copies of itself. Now consider an even more generalized generator, shown below, in which a line segment is replaced by n copies of itself that are contracted by a factor r.



From an argument that is almost identical to the one used above, the "4" in Eq. (38) is replaced by "n" so that the fractal dimension of the attractor curve is given by

$$D(n,r) = \frac{\ln n}{\ln \frac{1}{r}}\,. \tag{46}$$

Note the relationship between n, r and D may also be left in the following form:

$$nr^D = 1\,. \tag{47}$$

Let's step back and derive this result in the way that we did for the four-copy case. Suppose that we can cover the fractal curve with $N(\epsilon)$ circular boxes or disks of radius ϵ. We can imagine ϵ to be sufficiently, in fact, infinitesimally, small so that the D-dimensional measure of the fractal curve C is well approximated by (up to a multiplicative constant),

$$L^D(C) \cong N(\epsilon)\epsilon^D\,. \tag{48}$$

(Here, we are ignoring the $\pi/4$ factor that gives the area of a circular disk of radius ϵ. Even if we used it, it will eventually disappear. We could also have used rectangles with sides of length ϵ.) In fact, it may make things a little less confusing if we simply write that

$$L^D(C) = \lim_{\epsilon\to 0} N(\epsilon)\epsilon^D\,. \tag{49}$$

(Although that being said, the limit itself may be infinite!)

Now recall that each component of the fractal curve C is a contracted (and rotated, translated) copy of C. We now shrink C, along with the $N(\epsilon)$ ϵ-disks that cover it, by a factor of r. The result if a shrunken copy of C that is covered with $N(\epsilon)$ disks of radii $r' = r\epsilon$. We replace each segment of curve C (as well as the ϵ-disks that cover it) with the shrunked copy of C with $r\epsilon$-disks. The net result is that we have reconstructed curve C, but it is now covered with $nN(\epsilon)$ disks of radius $r' = r\epsilon$. The approximation to the D-dimensional measure of the curve is

$$L^D(C) \cong N(r\epsilon)(r\epsilon)^D = nN(\epsilon)(r\epsilon)^D\,. \tag{50}$$

But we can once again take limits,

$$L^D(C) = \lim_{\epsilon\to 0} N(r\epsilon)(r\epsilon)^D = \lim_{\epsilon\to 0} nN(\epsilon)(r\epsilon)^D \,. \tag{51}$$

We now combine the two limiting results for $L^D(C)$,

$$\lim_{\epsilon\to 0} N(\epsilon)\epsilon^D = \lim_{\epsilon\to 0} nN(\epsilon)(r\epsilon)^D \,. \tag{52}$$

Notice that we can divide out $N(\epsilon)\epsilon^D$ from both sides to obtain

$$nr^D = 1 \,, \tag{53}$$

from which it follows that

$$\boxed{D(n,r) = \frac{\ln n}{\ln \frac{1}{r}} \,.} \tag{54}$$

And if that was too confusing, just keep in mind the basic scaling result

$$N(r\epsilon) = N(\epsilon)r^{-D} \,, \tag{55}$$

that we discussed in the previous lecture. (It's actually an approximation - the equality holds in the limit $\epsilon \to 0^+$.) The number of measuring rods/balls/whatever with length/side length/whatever $\epsilon' = r\epsilon$ needed to cover the set is

$$N(r\epsilon) = nN(\epsilon) \,. \tag{56}$$

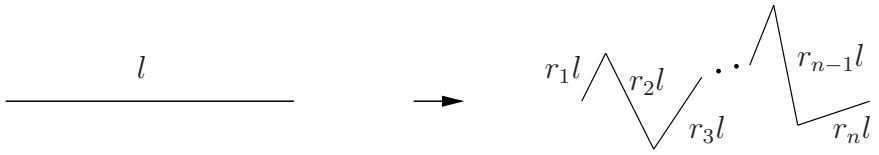
From (55) and (56), we have

$$nN(\epsilon) = N(\epsilon)r^{-D} \quad \Longrightarrow \quad n = \left(\frac{1}{r}\right)^D \quad \Longrightarrow \quad D = \frac{\ln n}{\ln \frac{1}{r}} \,, \tag{57}$$

in agreement with the result obtained earlier.

And even further generalizations of the von Koch curve

One may generalize the above generator G_r by employing scaling factors r_i that are not necessarily equal to each other, as sketched below.



This generalization makes it possible to generate a much wider class of fractal curves. Some examples are shown in the figure on the next page. On the left of each of the first rows is shown the generator G. On the right is the resulting fractal curve. The second and third sets show that when two lines of the generator are superimposed, tree-like patterns result. As we travel from the leftmost point of the generator (0,0) to the rightmost point (1,0), the fractal curve always lies to our left. As such, the two superimposed lines have different orientations, and a copy of the curve lies on each side.

In the final entry, the generator and fractal curve are shown on the same plot. The generator is slightly more complicated, in the sense that the path from (0,0) to (1,0) is not straight. The fractal curve is much more "cloudlike." In fact, a slight perturbation of this result, shown in the next figure, produces a curve which intersects back on itself infinitely often, giving rise to a more "curly" fractal set.

A fractal set C obtained from a generalized generator will once again be self-similar, i.e., it is a union of contracted copies C_i of itself. The major difference is that each copy C_i will be obtained by contracting C by the factor r_i. The fractal dimension D of these generalized curves is a more complicated function of the scaling functions r_i as compared to the *homogeneous* case $r_i = r$ studied earlier. Here we peform a quick derivation of the result.

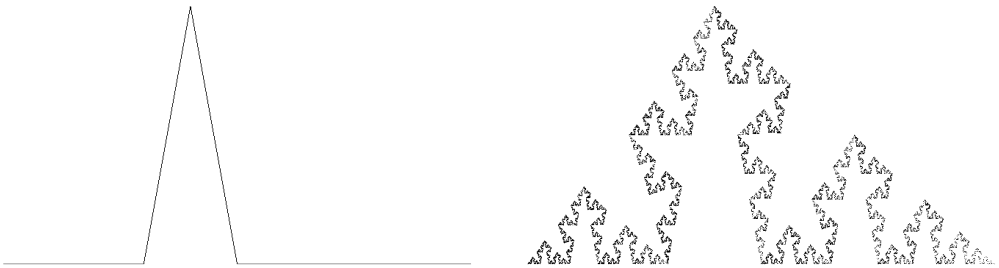
Once again, let us assume that ϵ is sufficiently small so that we can cover the entire curve C with $N(\epsilon)$ circular disks of radius ϵ. The D-dimensional measure of the curve is well approximated by

$$L^D(C) \cong N(\epsilon)\epsilon^D \,. \tag{58}$$

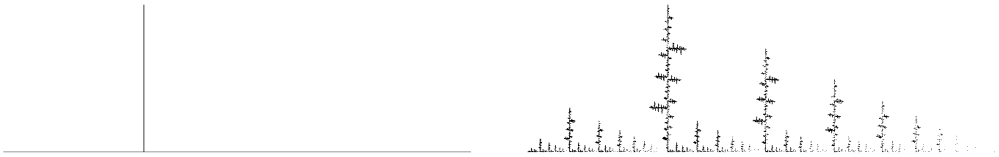
Now make n copies of this curve, along with the ϵ-radius disks on it. Each copy, along with the ϵ-disks will be contracted by the factor r_i. Now replace each piece of the fractal curve C with the appropriate r_i-contracted copy of C with the disks. The curve C will remain the same. The first piece of C will be covered with with $N(\epsilon)$ disks of radius $r_1\epsilon$, the second piece with $N(\epsilon)$ disks of radius $r_2\epsilon$, etc., until we arrive at the nth piece covered with $N(\epsilon)$ disks of radius $r_n\epsilon$. As a result, we now have another estimate of the D-dimensional measure,

$$L^D(C) = N(\epsilon)(r_1\epsilon)^D + N(\epsilon)(r_2\epsilon)^D + \cdots + N(\epsilon)(r_n\epsilon)^D \,. \tag{59}$$

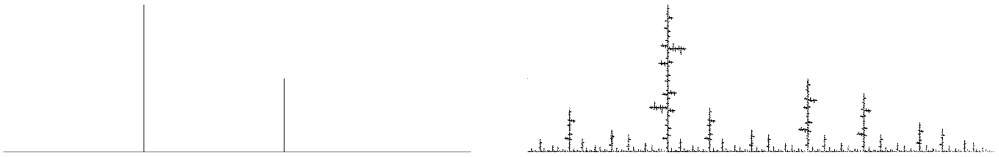
Some examples of generalized von Koch curves



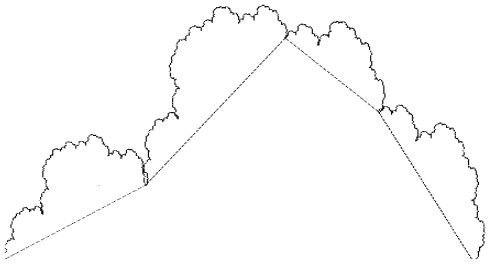
Generator G with four copies – the result is a modified von Koch curve



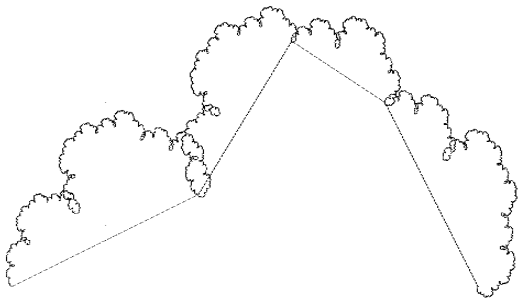
Generator G with four copies – the result resembles a collection of trees



Generator G with seven copies – another tree-like scene



Rather than being dendritic, the fractal curve is cloudlike. Here we show the generator and the curve on the same plot.



Generator G producing a more "curly" fractal set.

If we now combine these two estimates for $L^D(C)$, and understand that they are equal in the limit $\epsilon \to 0^+$, we obtain, after cancelling the $N(\epsilon)$ and ϵ^D factors from both sides, the following equation for D,

$$r_1^D + r_2^D + \cdots + r_n^D = 1. \tag{60}$$

Note that in the special case $r_i = r$, i.e., all scaling factors are equal, the above equation becomes

$$nr^D = 1\,. \tag{61}$$

But we have seen this equation before – it is Eq. (53) which, of course, was derived for the case in which all scaling factors are equal. In this case,

$$D = \frac{\ln D}{\ln \frac{1}{r}}\,. \tag{62}$$

Returning to Eq. (60) above, we claim that there is a unique solution to this equation and leave the proof of this claim as an exercise for the reader. The basic steps are as follows:

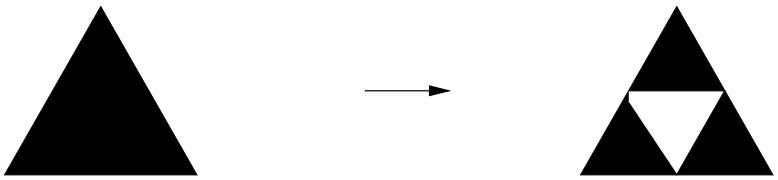
- For a fixed set of scaling factors, $0 < r_i < 1$ for $1 \le i \le n$, $n \ge 2$, define the following function of D:
$$f(D) = r_1^D + r_2^D + \cdots r_n^D\,. \tag{63}$$
- Note that $f(0) = n$.
- Also note that $f(D) \to 0^+$ as $D \to \infty$ since each of the r_i lie between 0 and 1.
- Show that $f'(D) < 0$ for $D > 0$, which implies that $f(D)$ is strictly decreasing for $D > 1$.
- From the above information, it should follow that the graph of $f(D)$ vs. D should intersect the line $y = 1$ only once. At this value of D, Eq. (60) is satisfied.

Lecture 29

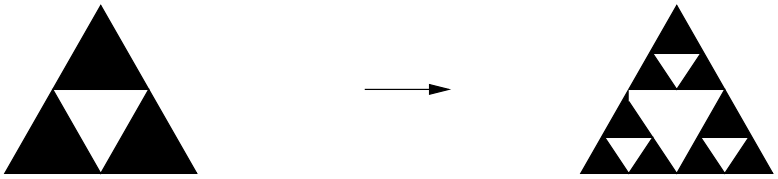
Measuring the sizes of fractal sets in $\mathbb{R}^n$ (cont'd)

The "Sierpinski triangle" or "Sierpinski gasket"

We now conclude this section with a classic fractal set in the plane, the so-called "Sierpinski triangle". We begin with an equilateral triangle and define a generator G that removes the middle quarter of triangular regions (open sets) as shown below:



The action of the generator on the triangular region may also be interpreted geometrically as follows: G produces three contracted copies of the region by shrinking in the x and y directions by a factor of $r = \frac{1}{2}$ and then placing the three copies so that they touch at the midpoints of the triangular boundary. The repeated action of G produces the following sets:



In the limit, the sequence of sets converges to the following fractal set S, the Sierpinski triangle, shown in the figure below.

The set S is easily seen to be a union of three contracted copies of itself, with contraction factor $r = \frac{1}{2}$. A measurement procedure involving circular discs of diameter $\epsilon_n = \frac{1}{2^n}$ shows that S has zero two-dimensional area and infinite one-dimensional length. Let us now recall the basic scaling result for self-similar fractals, relating the number of disks needed to cover a set with fractal dimension N,

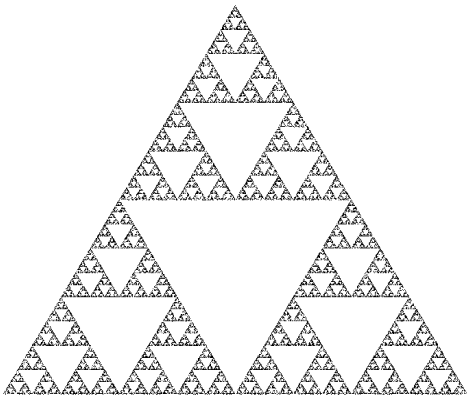
$$N(r\epsilon) = N(\epsilon) r^{-D} \,. \tag{64}$$

In the case of the Sierpinski triangle,

$$r = \frac{1}{2} \tag{65}$$

and

$$N\left(\frac{1}{2}\epsilon\right) = 3N(\epsilon) \,, \tag{66}$$



Sierpinski triangle

This comes from the fact that if you use $N(\epsilon)$ disks of radius ϵ to cover the entire triangle, and then make three shrunken copies (by factor $r = \frac{1}{2}$), you can cover each of the three subsets of the triangle with $N(\epsilon)$ disks of radius $\frac{1}{2}\epsilon$. As a result, you need $3N(\epsilon)$ disks of radius $\frac{1}{2}\epsilon$ to cover the triangle.

From Eqs. (64)-(66), we have

$$3N(\epsilon) = N(\epsilon)\left(\frac{1}{2}\right)^{-D}, \tag{67}$$

or

$$3 = 2^D, \tag{68}$$

which implies that

$$D = \frac{\ln 3}{\ln 2} \cong 1.58496 \tag{69}$$

Note that $1 < D < 2$ – the Sierpinski triangle, as expected, is "thicker" than a line but "thinner" than a two-dimensional region in the plane.

Another look at fractal dimensions in terms of scaling:

Let's examine the above result one more time to see why the fractal dimension D lies between the values of 1 and 2. The Sierpinski triangle is a union of **three** contracted copies of itself. Each smaller copy is obtained by contracting the triangle by a factor of $r = \frac{1}{2}$ in the x- and y-directions. Since the Sierpinski triangle is a union of **three** contracted copies of itself, this means that the "size" of each copy is **one-third** the "size" of the entire Sierpinski triangle. (By "size", we can mean "mass" or "measure", or whatever.)

Now constrast this fact with the way that "normal" Euclidean objects scale, using the scaling factor $r = \frac{1}{2}$ that was used for the Sierpinski triangle:

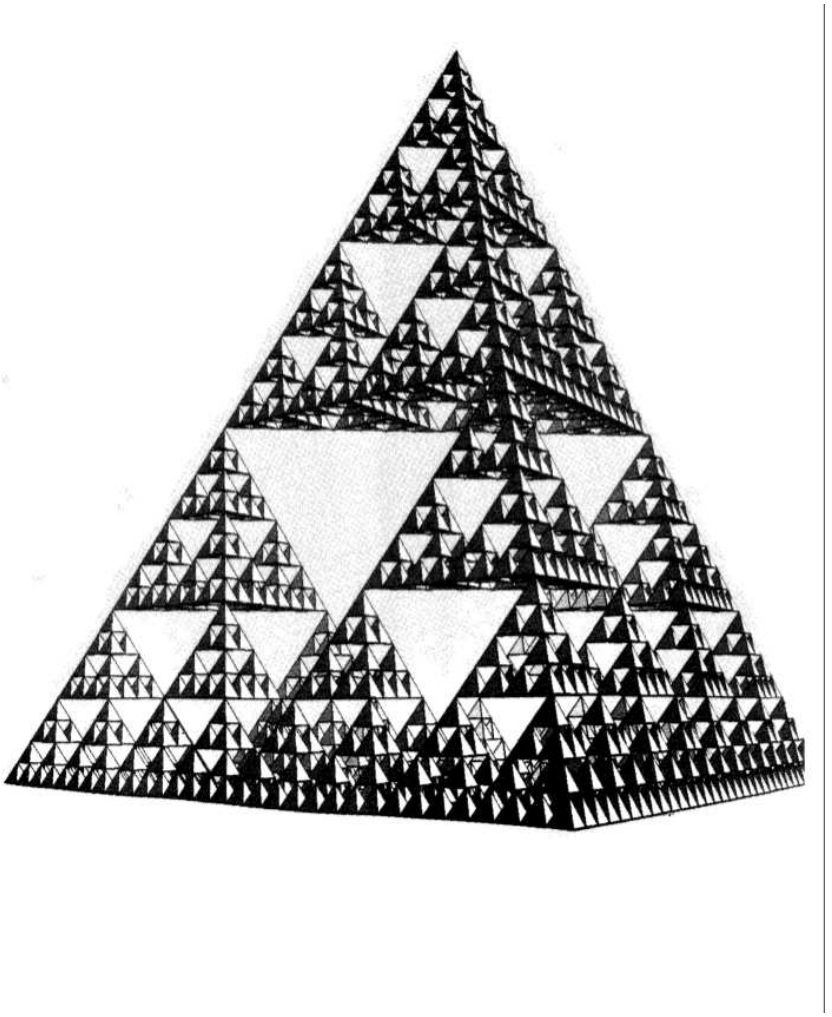
1. Consider a line of length 1 in the plane: say, the interval $[0, 1]$ on the x-axis. Let us now make **two** contracted copies of this interval, namesly $[0, \frac{1}{2}]$ and $[\frac{1}{2}, 1]$. Both of then are contractions of the interval $[0, 1]$ by a factor of $r = \frac{1}{2}$. And the length of each of these copies is **one-half** the length of the interval $[0, 1]$. This is simple – it is the way that one-dimensional sets scale.

2. Now consider a square with sides of length 1, say the interval $[0, 1]^2$. If we shrink the sides of the square by factor $\frac{1}{2}$, we obtain a square that is **one-quarter** the area of the square. As such, we'll need **four** of these copies to reconstruct the original square. This is simple as well – it is the way that two-dimensional sets scale.

3. But when we shrink the Sierpinski triangle in the x- and y-directions by the factor $r = \frac{1}{2}$, we produce a smaller triangle which is **one-third** of the size of the original triangle: Not one-half, as was the case for lines, nor one-quarter, which was the case of regions, but one-third, which lies between one-half and one-quarter. Because the Sierpinski triangle has "holes" in it, it scales neither as a one-dimensional nor as a two-dimensional object. It scales as a D-dimensional object, where $D = \ln 3 / \ln 2$.

Other fractals obtained by dissection

There are many generalizations of the Sierpinski triangle in $\mathbb{R}^2$ – a few of these will be presented in Problem Set No. 6. One can also extend the method to higher dimensions, i.e., a "Sierpinski gasket" in $\mathbb{R}^3$. A sketch of this fractal, taken from B. Mandelbrot's book, *The Fractal Geometry of Nature*, is presented in the figure below.



Three-dimensional Sierpinski gasket, from **The Fractal Geometry of Nature**, by B. Mandelbrot.

To construct this fractal, we begin with a tetrahedron (3D solid with four equilateral triangles as faces). For each of the four triangular faces, connect the midpoints of its three edges with lines, as is done in the construction of the Sierpinski triangle. Then remove all interior points so that four smaller tetrahedra, each of which is a contracted copy of the original tetrahedron, remain. Then repeat this procedure for the four tetrahedra to produce 16 smaller tetrahedra, and so on.

The above 3D Sierpinski gasket is self-similar: It may be expressed as a union of 4 contracted copies of itself. In this case, the contraction factor is $r = \frac{1}{2}$. As such, the fractal dimension D of the

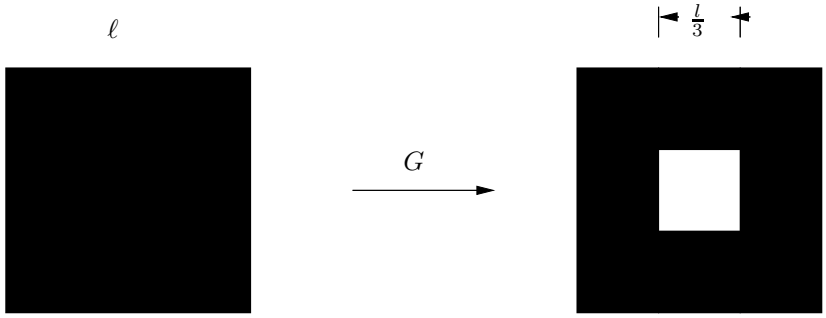
3D Sierpinski gasket is

$$D = \frac{\ln n}{\ln \frac{1}{r}} = \frac{\ln 4}{\ln 2} = 2\,. \tag{70}$$

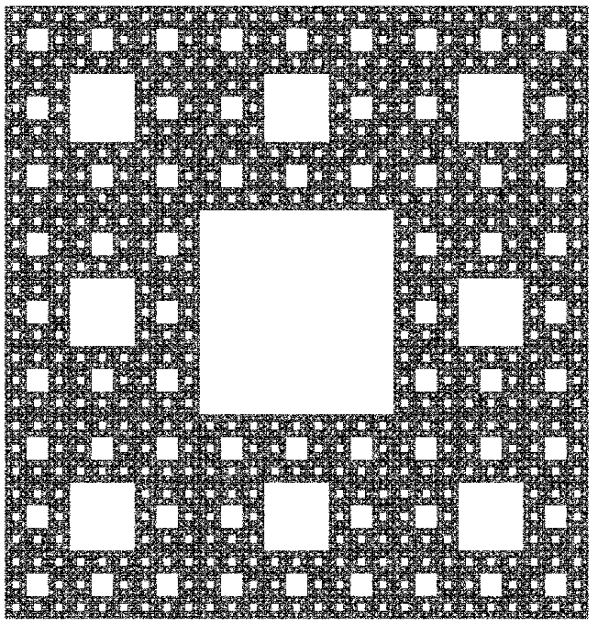
This result may be rather surprising. When we think of a two-dimensional object in $\mathbb{R}^3$, we think of a planar region or, in general, a surface. "Nice" (i.e., smooth) surfaces can be parametrized in terms of two-parameters, a consequence of their two-dimensionality. Clearly, however, the 3D Sierpinski gasket is **not** a surface. It is a "three-dimensional object," but it scales as a two-dimensional one.

Sierpinski carpet and Menger sponge

Consider the following generator in $\mathbb{R}^2$, which operates on regions in the plane. It removes the "middle-ninth" of a square region:



Repeated application of this procedure yields, in the limit, the so-called two-dimensional "Sierpinski carpet," an approximation of which is shown in the figure below.

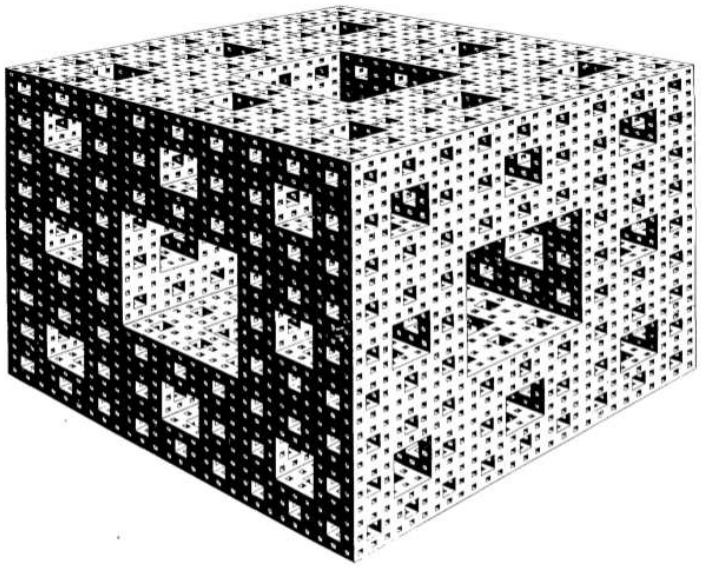


The "Sierpinski carpet"

The Sierpinski carpet is self-similar: It may be expressed as a union of 8 contracted copies of itself. In this case, the contraction factor is $r = \frac{1}{3}$. As such, the fractal dimension D of the Sierpinski carpet is

$$D = \frac{\ln n}{\ln \frac{1}{r}} = \frac{\ln 8}{\ln 3} \cong 1.89279\,. \tag{71}$$

An illustration of a three-dimensional version of the Sierpinski carpet, known as the "Menger sponge," taken from Mandelbrot's book, *Fractal Geometry of Nature* is shown below.



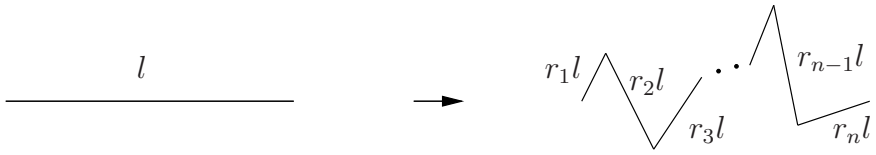
Three-dimensional Menger sponge, from *The Fractal Geometry of Nature* by B. Mandelbrot.

The Menger sponge is a union of $n = 20$ copies of itself, each of which is obtained by shrinking the three sides of the sponge by a factor of $r = \frac{1}{3}$. (Why twenty? We start by taking a cube and extracting six middle-ninth cubes from each side, along with the "core" cube, for a total of seven cubes being taken out of the large cube. This leaves 27-7=20 cubes after the first iteration.)
The dimension D of the sponge is therefore

$$D = \frac{\ln n}{\ln(1/r)} = \frac{\ln 20}{\ln 3} \approx 2.727\,. \tag{72}$$

Generalized Cantor-like sets

In the previous two lectures, we examined the construction of generalized von Koch-like curves. The last family of generators had the form,

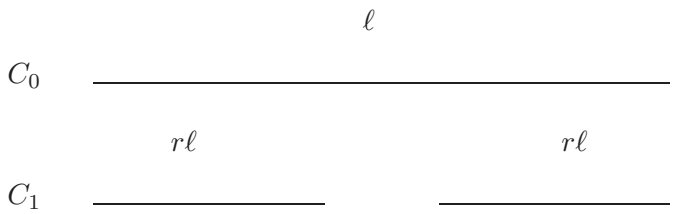


It was never mentioned expicitly, but this family of generators – which includes the von Koch curve as a particular case – have a special form which guarantees that the limiting set C is a **connected set**, i.e., there are no "gaps" in the set: Given any two points $x, y \in C$, one can move from x to y along points in C without ever leaving C. (Admittedly, this is a rather imprecise definition, but it will do.)

The "special form" mentioned above is that **the generator itself is connected** – each "piece" of the generator touches another one, and no "pieces" are left unconnected.

If the pieces are not connected to each other, then the resulting limit set will be **disconnected**. We saw this in the case of the Cantor set, which was constructed with following generator,

The above generator can be generalized as follows,



where $0 < r < \frac{1}{2}$. When $r = \frac{1}{3}$, the generator is the same as that of the ternary Cantor set. In the case $r = \frac{1}{2}$, no part of the original interval C_0 is removed, so $C_1 = C_0$. The limiting set will simply be the original interval C_0.

For simplicity, let $C_0 = [0, 1]$ with length $\ell = 1$. In the case $0 < r < \frac{1}{2}$ it is easy to see that repeated applications of the generator to the interval $[0, 1]$, with length $\ell = 1$, will produce sets C_n which are composed of 2^n subintervals of length r^n. Therefore, the total length L_n of the set C_n is

$$L_n = 2^n r^n \,. \tag{73}$$

Intuitively, we expect $L_n \to 0$ as $n \to \infty$, but can one deduce this from the above formula? The answer is "Yes" if we rewrite it as follows,

$$L_n = \left(\frac{r}{\frac{1}{2}}\right)^n \to 0 \quad \text{as} \quad n \to \infty\,, \tag{74}$$

where we have used the fact that $r < \frac{1}{2}$.

The limiting set C produced by this generator will be a **Cantor-like** set that has the same qualitative properties of the ternary Cantor set, i.e., it has **zero length** and it is **totally disconnected.** (Recall that by "totally disconnected" we mean that it is not possible to move from one point $x \in C$ to another point $y \in C$ without leaving the set C.) The distances between points will be slightly different than for the ternary Cantor set because the factor $\frac{1}{3}$ has been replaced by the factor r. (For example, the length of the middle gap will be $d = 1 - 2r$.) **The set C will be a union of two contracted copies of itself – in this case, the contraction factor will be r.**

In order to find the fractal dimension D of this generalized Cantor set, we can use the same type of scaling argument that was used for the generalized von Koch curves. If we need $N(\epsilon)$ circular tiles of radius ϵ to cover the set C, then the number of circular tiles of radius $r\epsilon$ will be simply

$$N(r\epsilon) = 2N(\epsilon)\,, \tag{75}$$

since there are two copies of the set. Once again, we can use the scaling equation,

$$N(r\epsilon) = N(\epsilon)r^{-D}\,, \tag{76}$$

to determine the fractal dimension D. Equating the RHS of the above two equations,

$$2N(\epsilon) = N(\epsilon)r^{-D}\,, \tag{77}$$

which, after division by $N(\epsilon)$ yields

$$r^{-D} = 2 \quad \Longrightarrow \quad \left(\frac{1}{r}\right)^D = 2 \quad \Longrightarrow \quad \boxed{D(r) = \frac{\ln 2}{\ln(\frac{1}{r})}\,.} \tag{78}$$

Let's examine a few properties of the dimension $D(r)$, considered as a function of the scaling parameter $0 \le r \le \frac{1}{2}$.

1. First, note that as $r \to 0^+$, $\ln(\frac{1}{r}) \to \infty$ which implies that $D \to 0$. In the limit that $r = 0$, the generator removes all points of the line segment except its endpoints. As such, we have reached the limit of the iteration procedure – the result is a set composed of two points, the endpoints of the line segment. The dimension of this set is 0.

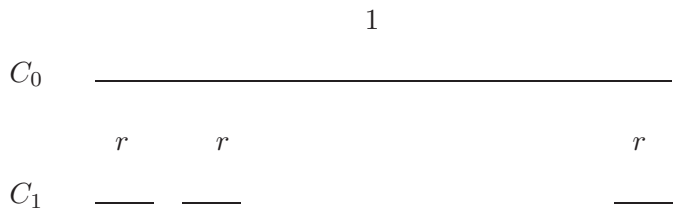
2. Differentiation of $D(r)$ with respect to r yields,

$$D'(r) = \frac{\ln 2}{\left[\ln(\frac{1}{r})\right]^2} \cdot \frac{1}{r^2} > 0\,. \tag{79}$$

This shows that $D(r)$ increases with r. Intuitively, the Cantor-like set C is getting "thicker" as r increases.

3. As $r \to \frac{1}{2}^-$, $D \to 1$. This is consistent with our earlier comment that in the case $r = \frac{1}{2}$, the limiting set is the interval $[0, 1]$ which is a simple Euclidean curve of dimension 1.

The next generalization is to allow the generator to make n identical copies of the interval, each copy being a contracted version with contraction factor r, i.e,



Note that we have omitted the length ℓ from the generator. From now on, it is understood that a segment of whatever length is replaced by n copies of r times its length.

In this case, we must have that $r < \frac{1}{n}$ for "dissection" to occur, i.e., for "holes" to be produced. Notice that we did not specify the spacings between the n "pieces" of the generator. For simplicity, we can assume that they are equally spaced from each other, which implies that the length of each of the $n-1$ spacings is $\frac{1-nr}{n-1}$ (Exercise).

Repeated application of this generator to the unit interval $[0, 1]$ will produce, in the limit, a Cantor-like set C. This set will have zero length and will be "totally disconnected." It will be self-similar in the sense that it can be expressed as a union of n contracted copies of itself.

We now employ the usual scaling argument to find the fractal dimension D of the Cantor-like set C. Once again, if we need $N(\epsilon)$ circular tiles of radius ϵ to cover the set C, we'll need

$$N(r\epsilon) = nN(\epsilon) \tag{80}$$

circular tiles of radius $r\epsilon$ to cover C. Once again, we use the scaling equation,

$$N(r\epsilon) = N(\epsilon)r^{-D}\,, \tag{81}$$

to determine the fractal dimension D. Equating the RHS of the above two equations,

$$nN(\epsilon) = N(\epsilon)r^{-D}\,, \tag{82}$$

which, after division by $N(\epsilon)$ yields

$$r^{-D} = n \quad \Longrightarrow \quad \left(\frac{1}{r}\right)^{D} = n \quad \Longrightarrow \quad \boxed{D(r) = \frac{\ln n}{\ln(\frac{1}{r})}\,.} \tag{83}$$

Note that this result is identical – at least in form – to the result for the generalized von Koch curve in Eq. (54). In this case, however, the scaling parameter r must obey the inequality $r < \frac{1}{n}$ in order to produce a Cantor-like set which, in turn, will have dimension $D < 1$. Note that $D(r)$ in the above equation is but a slight modification of its counterpart in Eq. (78) for the case $n = 2$. We can once again employ a little Calculus to obtain the following results,

1. As $r \to 0^{+}$, $D \to 0$. In the limit that $r =$, the generator removes all but n points from the line segment: The two endpoints along with $n-2$ points which are equally spaced between the endpoints. (Exercise). The dimension of this finite set of points is 0.

2. Differentiation of $D(r)$ with respect to r yields,

$$D'(r) = \frac{\ln n}{\left[\ln(\frac{1}{r})\right]^{2}} \cdot \frac{1}{r^{2}} > 0\,. \tag{84}$$

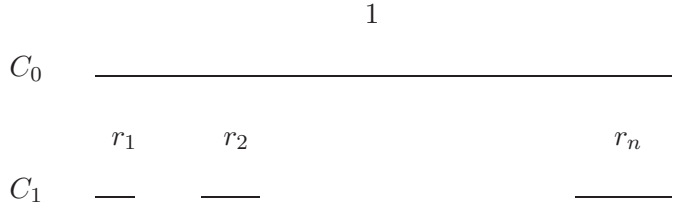
   This once again shows that $D(r)$ increases with r. Intuitively, the Cantor-like set C is getting "thicker" as r increases.

3. As $r \to \frac{1}{n}^{-}$, $D \to 1$. In the case $r = \frac{1}{n}$, the limiting set is the interval $[0, 1]$ which is a simple Euclidean curve of dimension 1.

   Note: Recall that the case of the **generalized von Koch curves**, the parameter r had to be sufficiently large, ie., $r > \frac{1}{n}$, in order to produce a curve with no "gaps." For such curves, $D > 1$.

Another note: From Eq. (83), the fractal dimension $D(r)$ does not depend on the **spacings** between the n "pieces" of the generator. The equation for D does not contain any information regarding these spacings - whether they be equal or nonequal. As such, we could further generalize the dissection process by allowing unequal spacings of the shrunken segments of length r.

Another generalization is the possiblity of having different scaling factors for the different pieces, i.e.,



The condition for dissection would be that

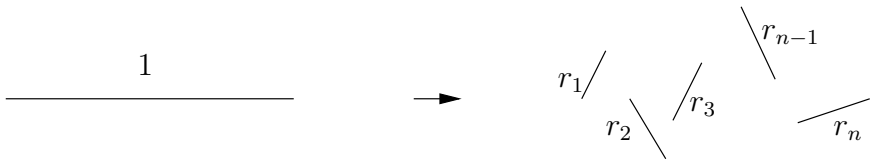
$$\sum_{k=1}^{n} r_k < 1\,. \tag{85}$$

We can treat this case in the same way that we did for the generalized von Koch curve with variable scaling factors and leave the derivation as an exercise. The final result is that D will satisfy the equation

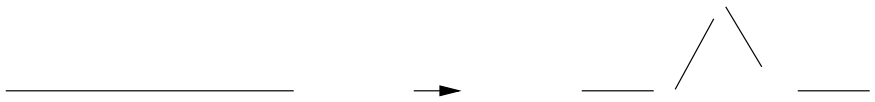
$$r_1^D + r_2^D + \cdots + r_n^D = 1\,, \tag{86}$$

which is identical to Eq. (60) for the generalized von Koch curve.

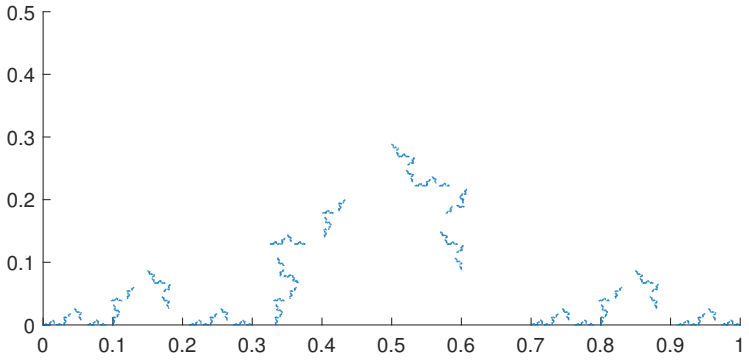
And one final generalization: There is no absolute need to keep the copies of the segment being dissected on the same line as the segment. In other words, we could produce shrunken copies that lie either off the original segment and point in other directions as sketched schematically below.



The resulting limit set will be a Cantor-like set in $\mathbb{R}^2$ which does not lie on a straight line. For example, consider the following generator which is obtained from the original von Koch curve generator by shrinking each of its four components,



The resulting limit set is shown below. The points look like they lie on a von Koch-like curve. In theory, the set is totally disconnected only some of the "gaps" will be visible because of the thickness of the points used to plot the set.



A short note on fractal-like sets occurring in nature

Benoit Mandelbrot's classic work, *The Fractal Geometry of Nature* centered, as its name suggests, on the idea that fractal patterns occur in nature – not just here and there, but everywhere, e.g., clouds, rock, soil, river/tributary patterns, etc..

Diffusion-limited aggregates: Fractal-like sets occurring in nature

An excerpt from the article entitled, "Diffusion-Limited Aggregation: A Model for Pattern Formation," by Thomas C. Halsey from the *Physics Today* online archive at
`http://web.archive.org/web/20070405094836/http://aip.org/pt/vol-53/iss-11/p36.html`

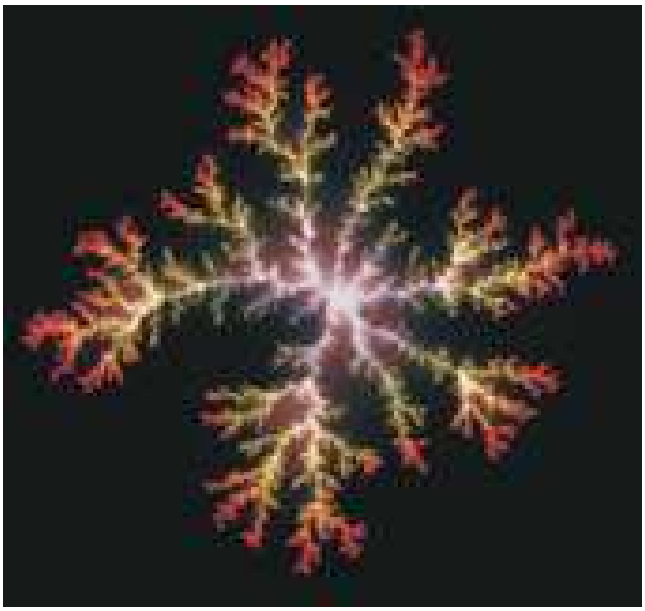
The basic concept

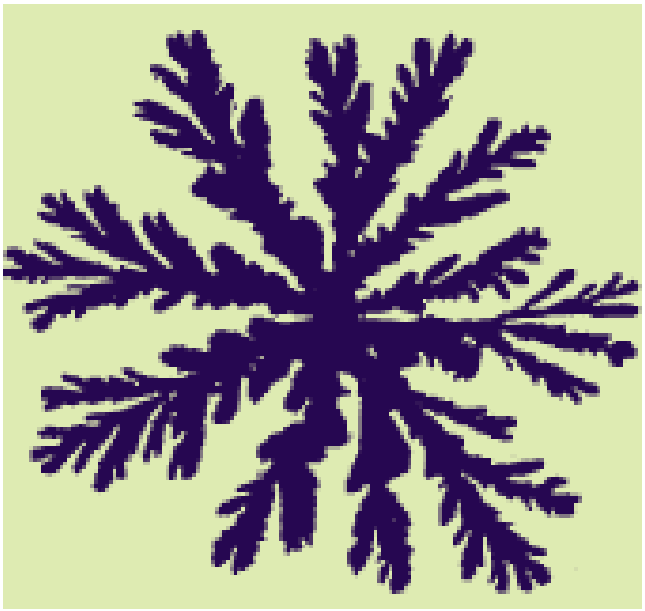
To understand the basics, consider colloidal particles undergoing Brownian motion in some fluid, and let them adhere irreversibly on contact with one another. Suppose further that the density of the colloidal particles is quite low, so one might imagine that the aggregation process occurs one particle at a time. We are then led to the following model.

Fix a seed particle at the origin of some coordinate system. Now introduce another particle at a large distance from the seed, and let it perform a random walk. Ultimately, that second particle will either escape to infinity or contact the seed, to which it will stick irreversibly. Now introduce a third particle into the system and allow it to walk randomly until it either sticks to the two-particle cluster or escapes to infinity. Clearly, this process can be repeated to an extent limited only by the modeler's patience and ingenuity (the required computational resources grow rapidly with n, the number of particles).

The clusters generated by this process are both highly branched and fractal. The cluster's fractal structure arises because the faster growing parts of the cluster shield the other parts, which therefore become less accessible to incoming particles. An arriving random walker is far more likely to attach to one of the tips of the cluster shown in the figure below (left) than to penetrate deeply into one of the cluster's "fjords" without first contacting any surface site. Thus the tips tend to screen the fjords, a process that evidently operates on all length scales. The figure at the right shows the "equipotential lines" of walker probability density near the cluster, confirming the unlikelihood of random walkers penetrating the fjords.
An experimental situation is shown in the figure below.

The fractal dimension D of both the theoretical as well as the experimental cluster in three-dimensions is roughly $D = 2.5$.





Estimating the (global) fractal dimension D of sets/objects

Given a set $S \in \mathbb{R}^k$, let $N(\epsilon)$ be the minimum number of k-dimensional balls of diameter ϵ in $\mathbb{R}^k$ needed to "cover" S, i.e., S is contained in a union of these balls. If S has fractal dimension D, then

$$N(\epsilon) \cong A\epsilon^{-D} = A\left(\frac{1}{\epsilon}\right)^{D}. \tag{87}$$

For mathematical objects, the approximation improves as $\epsilon \to 0^+$. Now take logarithms of both sides of the above equation,

$$\log N(\epsilon) \cong D \log\left(\frac{1}{\epsilon}\right) + \log A\,. \tag{88}$$

Rearranging,

$$D \cong \frac{\log N(\epsilon)}{\log \frac{1}{\epsilon}} - \frac{\log A}{\frac{1}{\log \epsilon}} \,. \tag{89}$$

The mathematical definition of D is

$$D = \lim_{\epsilon \to 0^+} \frac{\log N(\epsilon)}{\log \frac{1}{\epsilon}} \,. \tag{90}$$

For natural objects, one will not be able to let $\epsilon \to 0^+$. As such, we determine $N(\epsilon)$ over a range of reasonable ϵ-values.

Basic idea:

1. Work with a finite set of epsilon values, i.e., $\epsilon_1 > \epsilon_2 > \cdots \epsilon_n$, e.g.,

   $$\epsilon_k = r^k \,, \tag{91}$$

   where $0 < r < 1$ is a convenient scaling factor, possibly relevant to the problem.

2. Let $N(\epsilon_k)$ denote the number of ϵ_k-balls needed to cover S.

3. Plot $\log N(\epsilon_k)$ vs. $\log \dfrac{1}{\epsilon_k}$ for $1 \leq k \leq n$.

4. A least-squares fit to this data will produce a straight line with slope m that will be an approximation to D.

Iterated function systems (IFS) and the construction of fractal sets

We now discuss a very convenient and powerful method of constructing and analyzing self-similar fractal sets – the so-called method of **iterated function systems**, abbreviated as **IFS**. The IFS method will involve the use of a set **contractive mappings** on the line $\mathbb{R}$, or in the plane $\mathbb{R}^2$ or in 3D space $\mathbb{R}^3$. as opposed to the generators discussed in the previous lectures. In this way, one doesn't have to keep track of line segments and then operating on them. A remarkable aspect of the IFS method is that one can generate very good pictures of fractals by simply applying the maps in an IFS in a random manner, as was discussed in Problem Set No. 5.

We'll illustrate the idea with a simple example – the one, in fact, that was used in Problem Set No. 5. From that example, we'll move on to a more general discussion, and then look a little more closely at the mathematics behind IFS.

Consider the following two maps, f_1 and f_2, which map the interval $I = [0,1]$ into itself: For any $x \in [0,1]$, $f_1(x) \in [0,1]$ and $f_2(x) \in [0,1]$:

1. $f_1(x) = \frac{1}{3}x$ with fixed point $\bar{x}_1 = 0$. The iteration dynamics associated with $f_1(x)$ is quite straightforward: For any $x_0 \in [0,1]$, $x_n = f_1^n(x_0) = \frac{x_0}{3^n} \to 0$ as $n \to \infty$.

2. $f_2(x) = \frac{1}{3}x + \frac{2}{3}$ with fixed point $\bar{x}_2 = 1$. The iteration dynamics associated with $f_2(x)$ is also quite straightforward: The distance between $x_{n+1} = f_2(x_n)$ and the fixed point $\bar{x}_2 = 1$ is one-third the distance between the point x_n and $\bar{x}_2$. As such, $x_n = f_2^n(x_0) \to 1$ as $n \to \infty$.

We now examine the (repeated) action of the maps f_1 and f_2 **on the interval** $I = [0,1]$. Instead of looking at where each map f_i sends a point $x \in [0,1]$, we consider where each map f_i will send **sets of points in [0,1]**. It is instructive to examine the action of each of the above maps on the interval [0,1]. To do this, let us define the following associated **interval-** or **set-valued mappings**:

Definition: For $i \in \{1,2\}$ and any subset $S \subseteq [0,1]$, we denote $\hat{f}_i(S) = \{f_i(x) | x \in S\}$.

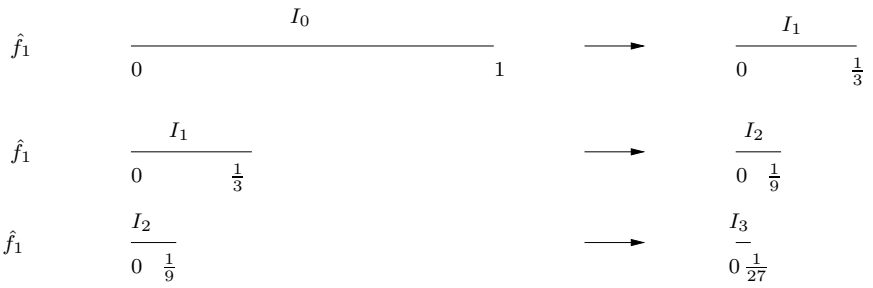
To illustrate:

1. $\hat{f}_1([0,1]) = [0, \frac{1}{3}]$. In other words, $\hat{f}_1$ "shrinks" the interval [0,1] to $[0, \frac{1}{3}]$.

2. $\hat{f}_2([0,1]) = [\frac{2}{3}, 1]$. In other words, $\hat{f}_1$ "shrinks" the interval $[0,1]$ to $[\frac{2}{3}, 1]$.

It might help to use some pictures. Let's consider map f_1 for the moment. If we let $I_0 = [0,1]$, then the interval $I_1 = \hat{f}_1(I_0) = [0, \frac{1}{3}]$ is sketched below. We then apply $\hat{f}_1$ to the set I_1 to produce the set $I_2 = \hat{f}_1(I_1) = [0, \frac{1}{9}]$.
In general, n applications of the map $\hat{f}_1$ to the interval I_0 produces the set

$$I_n = \hat{f}_1(I_0) = \left[0, \frac{1}{3^n}\right] . \tag{92}$$

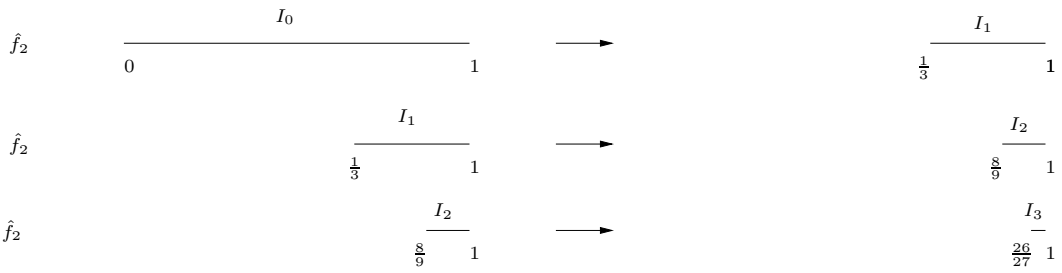


Clearly, the intervals I_n are shrinking in length. Moreover, they appear to be approaching a limit, i.e.,

$$\lim_{n\to\infty} I_n = \lim_{n\to\infty}\left[0, \frac{1}{3^n}\right] = \{0\}\,. \tag{93}$$

They are approaching the single point $\{0\}$. Note that this point is the fixed point of f_1. We may view this single point as a set or an interval consisting of one point.

Let's now consider the map f_2. If we let $I_0 = [0, 1]$, then the interval $I_1 = \hat{f}_2(I_0) = [\frac{2}{3}, 1]$ is sketched below. We then apply $\hat{f}_2$ to the set I_1 to produce the set $I_2 = \hat{f}_1(I_1) = [\frac{8}{9}, 1]$. This procedure is shown graphically below.



In general, n applications of the map $\hat{f}_2$ to the interval I_0 produces the set

$$I_n = \hat{f}_1(I_0) = \left[1 - \frac{1}{3^n}, 1\right] = \left[\frac{3^n - 1}{3^n}, 1\right]\,. \tag{94}$$

Once again, the intervals I_n are shrinking in length. Moreover, they appear to be approaching a limit, i.e.,

$$\lim_{n\to\infty} I_n = \lim_{n\to\infty}\left[1 - \frac{1}{3^n}, 1\right] = \{1\}\,. \tag{95}$$

They are approaching the single point $\{1\}$, which is the fixed point of f_2. We may once again view this single point as a set or an interval consisting of one point.

In summary:

1. Repeated application of the set-valued map $\hat{f}_1$ to the interval $I = [0, 1]$ produces a set of intervals I_n which are "shrinking" toward the fixed point $\bar{x}_1$ of f_1.

2. Repeated application of the set-valued map $\hat{f}_2$ to the interval $I = [0,1]$ produces a set of intervals I_n which are "shrinking" toward the fixed point $\bar{x}_2$ of f_2.

Now, instead of considering the action of each of the set-valued maps $\hat{f}_1$ and $\hat{f}_2$ separately, let's combine their actions as if they were components of a **"Parallel Machine".** We'll do this as follows:

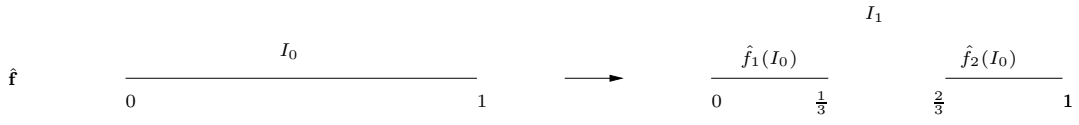
For any subset $S \in [0,1]$, define that following set-valued mapping,

$$\hat{\mathbf{f}}(S) = \hat{f}_1(S) \cup \hat{f}_2(S)\,. \tag{96}$$

In other words, the set-valued mapping is the union of the actions of the individual maps $\hat{f}_1$ and $\hat{f}_2$. Let's examine the action of $\hat{\mathbf{f}}$ on the set $I_0 = [0,1]$:

$$\begin{aligned} I_1 = \hat{\mathbf{f}}([0,1]) &= \hat{f}_1([0,1]) \cup \hat{f}_2([0,1]) \\ &= \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right]. \end{aligned} \tag{97}$$

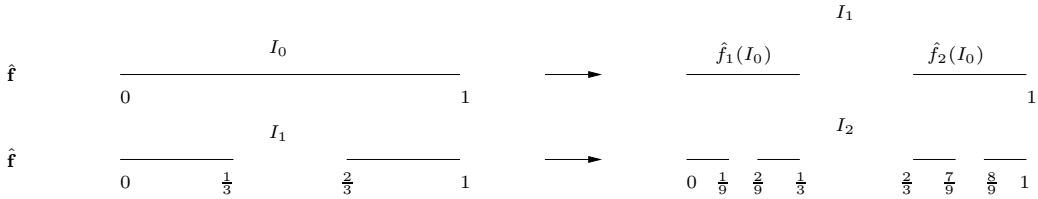
This is shown graphically below:



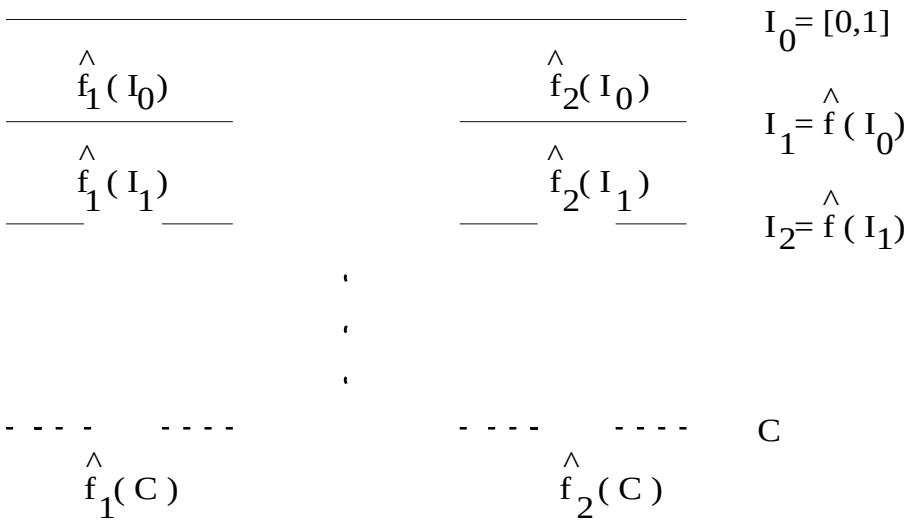
Now apply $\hat{\mathbf{f}}$ to the set I_1:

$$\begin{aligned} I_2 = \hat{\mathbf{f}}(I_1) &= \hat{f}_1(I_1) \cup \hat{f}_2([I_1) \\ &= \left(\left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right]\right) \cup \left(\left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]\right) \\ &= \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right]. \end{aligned} \tag{98}$$

This is shown graphically below:



We see that **the repeated application of the "parallel machine" $\hat{\mathbf{f}}$ performs the "middle-thirds dissection procedure" that was employed in the construction of the ternary Cantor set C in $[0,1]$:**



If we consider the following iteration process involving the parallel map $\hat{\mathbf{f}}$:

$$I_{n+1} = \hat{\mathbf{f}}(I_n) = \hat{f}_1(I_n) \cup \hat{f}_2(I_n)\,, \tag{99}$$

then it appears that

$$\lim_{n\to\infty} I_n = C\,, \quad \text{ternary Cantor set in [0,1]}\,. \tag{100}$$

But the story is not over! As shown in the figure above, **the ternary Cantor set C is a "fixed point" of the parallel operator $\hat{\mathbf{f}}$**, i.e.,

$$C = \hat{\mathbf{f}}(C) = \hat{f}_1(C) \cup \hat{f}_2(C)\,. \tag{101}$$

In other words, $\hat{\mathbf{f}}$ maps **the set** C to itself.

Recalling that the set-valued maps $\hat{f}_1$ and $\hat{f}_2$ "shrink" sets, we see that:

The ternary Cantor set C is a union of shrunken copies of itself.

We actually noticed this earlier, but we didn't really specify the maps that produced these copies. Now, we can identify the maps that produce the copies – the two maps $\hat{f}_1$ and $\hat{f}_2$ that comprise the "parallel operator" $\hat{\mathbf{f}}$.