Chapter 1

Examples of curious curves

While curves in the plane have intrigued mathematicians for centuries, the modern era erupted from Georg Cantor's investigations that began in the 1860's. Cantor showed that a line and a plane contain the same number of points by displaying a map from the *unit interval* I onto the *unit square* $U = I \times I$. Following this discovery, the natural question "Can you find a continuous map of I onto U?" drew the interest of many prominent mathematicians. Reference [Dauben (1970)] contains a fascinating discussion of Cantor's work, including, of course, the classical middle-thirds Cantor set that appeared in print as a note in 1883.

In this book a *curve* is the image of a nontrivial continuous map from I into the plane. (The image of a constant map is one point, a trivial curve.) The first example of a curve that contains a square, a *space filling curve*, appeared in 1890. Numerous interesting examples of space filling curves appear in [Sagan (1994)] with displays, photos, and commentary. A *simple curve* is the image of a one-to-one, continuous map from I into the plane.

Our central focus is the question "What can we learn about curves?" Elementary analysis is a basic tool for addressing this question. For example basic analysis results show that a curve is a compact, connected set in the plane and that a simple curve can contain no square. So U is a curve, but not a simple curve. Simple curves can be quite complex, possessing surprising properties related to area, Hausdorff dimension (which lets us compare sizes of curves that have area equal to zero), and smoothness. Geometry and Cantor sets are also basic tools for studying the subtleties of such curves.

Felix Klein said "Everyone knows what a curve is until he has studied enough mathematics to become confused through the countless number of possible exceptions." We aim to present a sampling of the classical results and introduce you to some fascinating new sets, some of which are not curves but are related. In most cases we illustrate the construction of the curves with computer generated drawings. In this chapter we begin with examples that illustrate particular features of curves and then introduce a class of Cantor sets that we use frequently. The interplay between curious curves and Cantor sets is a central theme of this book.

1.1 Variations of the Koch curve

The Koch curve was introduced by the Swedish mathematician Helge von Koch in 1904 as an example of a curve that does not have a tangent line at any point [Edgar (1993)]. It is also a classical example of a *self-similar curve*, in which every piece contains a portion similar to the whole figure.

1.1.1 *Koch curve*

The Koch curve is a self-similar, simple curve. There are several ways to describe the Koch curve. In this introduction, we construct the curve by an iterative replacement process beginning with the generator shown in Figure 1.1. We replace each of the four linear segments (parts) of the generator with a $1/3$-scaled copy of the generator as shown in Figure 1.2. When the process is repeated ad infinitum, the limiting image is defined to be the Koch curve K. A representation of the curve is shown in Figure 1.3. In Chapter 2 we will give the details of this construction and will define the function that gives K as a one-to-one continuous image of I. Note that the Koch curve is composed of four similar images of itself. Each similar image lies along a side of the generator. The Koch curve does not have a tangent at any point. Intuitively, the Koch curve has area equal to zero. We will later see simple curves that have non-zero area, which seems counter-intuitive (at this point).

The Koch snowflake (See Figure 1.4.) is obtained by replacing each of the three edges of an equilateral triangle with an outward facing Koch curve. The Koch snowflake is neither a self-similar curve nor a simple curve; it is a one-to-one, continuous image of a circle.

1.1.2 *Modified Koch curve*

Generalizations of the Koch curve begin with a generic generator $G = G(a, \theta)$ as shown in Figure 1.5, allowing the angle θ and the length a to vary. Note that the length L in this figure depends on a and θ. If $L < 1$ then G generates a self-similar curve by a replacement process similar to that for the Koch curve described above. In Figure 1.6 we show modifications for a few values of a with θ fixed at $\pi/3$.

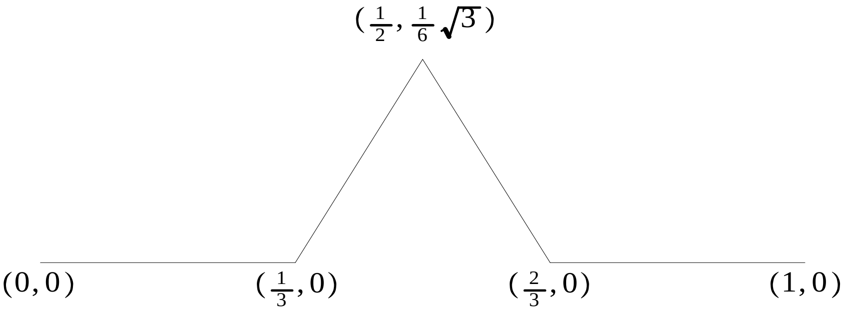


Fig. 1.1: Graph of the generator of the Koch curve.

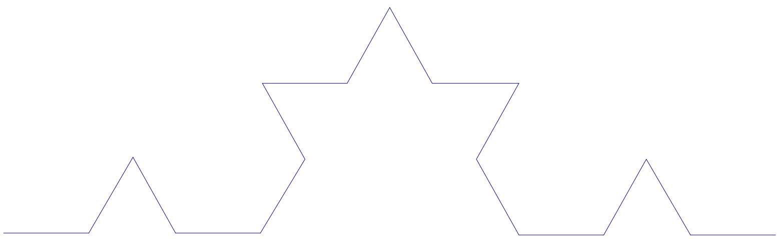


Fig. 1.2: Stage 2 in the generation of the Koch curve.

Note especially that for $a > 1/4$ the curve appears to be a simple curve. However, for values of $a < 1/4$, we see that the curve is self-intersecting. Because of this property, we designate $a = 1/4$ as the *pivotal value*. The curve generated with $a = 1/4$ displays the just-touching property. The self-intersection points of this curve are the vertices of equilateral triangles that are replicated throughout the curve because of the self-similarity property.

With the value of $\theta = \pi/4$, there is also a unique pivotal value which allows for the self-intersecting curve shown in Figure 1.7. We will show that the points of self-intersection in this curve form Cantor sets of points in the plane. These modified curves maintain many of the properties of the classic Koch curve while displaying intriguing differences. The details to verify that the modified Koch curve is a continuous image of I are given in Section 4.1.

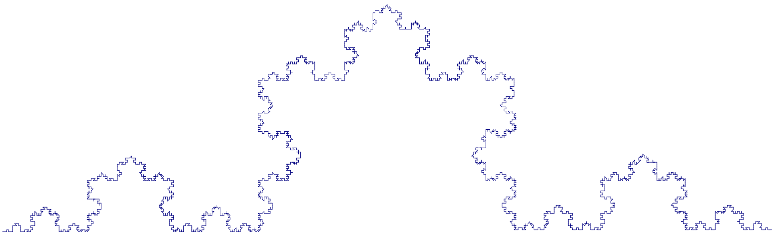


Fig. 1.3: The Koch curve.

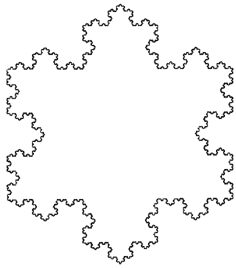


Fig. 1.4: The Koch snowflake.

1.1.3 *Basics of complex numbers*

As the Koch curve indicates, we extensively use rotations in the plane, conveniently described by multiplication of complex numbers. The set $\mathbb{C} = \{x + iy : x, y \in \mathbb{R}\}$ of complex numbers is closely related to the Euclidean plane

$$\mathbb{R}^2 = \{(x, y) : x, y \in \mathbb{R}\}.$$

We use the identification $x + iy \Leftrightarrow (x, y)$ and go between $\mathbb{R}^2$ and $\mathbb{C}$ without comment. Let $z = x+iy$ and $w = u+iv$ be in $\mathbb{C}$. Then addition and multiplication are defined on $\mathbb{C}$ by

$$z + w = (x + u) + i\,(y + v)$$

and

$$zw = (x + iy)\,(u + iv) = (xu - yv) + i\,(xv + yu) \tag{1.1}$$

respectively. In particular, $i^2 = -1$. Notice that addition and multiplication of real numbers are preserved. The modulus of z is the real number

$$|z| = |x + iy| = |(x, y)| = \sqrt{x^2 + y^2}.$$

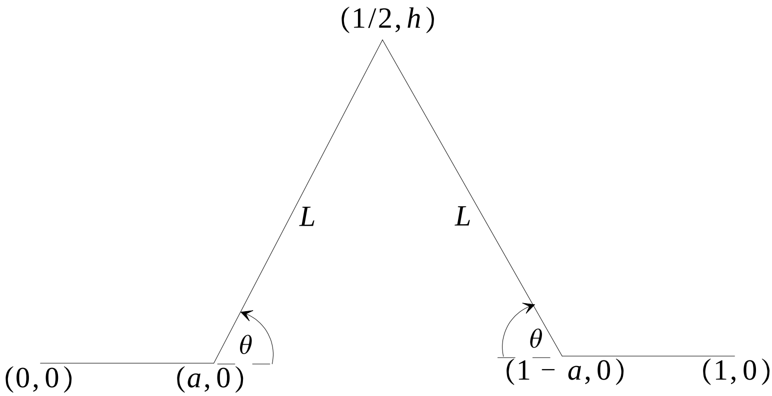
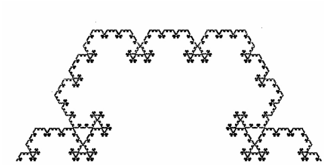
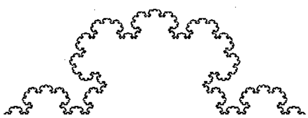


Fig. 1.5: Generic generator.

(a) $a < \frac{1}{4}$



(b) $a = \frac{1}{4}$



(c) $a > \frac{1}{4}$

Fig. 1.6: Generalized Koch curves.

Given $z \in \mathbb{C}$, it follows from trigonometry that there are numbers $\theta \in \mathbb{R}$ satisfying the equations $x = |z| \cos\theta$ and $y = |z| \sin\theta$. If $z = 0$, the set of solutions is $\mathbb{R}$. Otherwise, there is one solution, say ϕ, in $[0, 2\pi) = \{t \in \mathbb{R} : 0 \leq t < 2\pi\}$ and the other solutions are of the form $\theta = \phi + 2j\pi$, where j is any integer; the solutions comprise an equivalence class $\{\theta : \theta = \phi + 2j\pi,\ j \in \mathbb{Z}\}$.

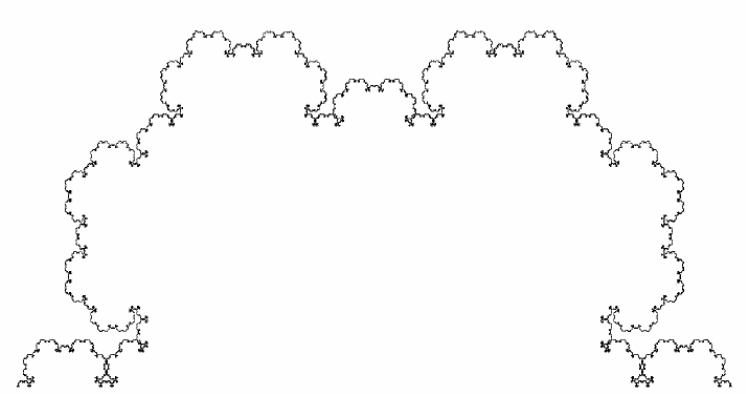


Fig. 1.7: Example of a self-intersecting curve with $\theta = \pi/4$.

We will usually suppress the equivalence class notation and simply let any number in the class stand for the whole class.

A complex number $(\cos\theta, \sin\theta)$ of modulus 1 is given the exponential notation $e^{i\theta}$. Complex exponentials behave like real exponentials. For example, if $z = x+iy = |z|(\cos\theta+i\sin\theta) = |z|e^{i\theta}$ and $w = u+iv = |w|(\cos\phi+i\sin\phi) = |w|e^{i\phi}$, using Equation 1.1 and the addition formulas for sine and cosine, we have

$$\begin{aligned} zw &= |z|\,e^{i\theta}\,|w|\,e^{i\phi} \\ &= |z|\,|w|\,e^{i\theta}e^{i\phi} \\ &= |z|\,|w|\,(\cos\theta + i\sin\theta)(\cos\phi + i\sin\phi) \\ &= |z|\,|w|\,((\cos\theta\cos\phi - \sin\theta\sin\phi) + i\,(\sin\theta\cos\phi + \cos\theta\sin\phi)) \\ &= |z|\,|w|\,(\cos(\theta+\phi) + i\sin(\theta+\phi)) \\ &= |z|\,|w|\,e^{i(\theta+\phi)}. \end{aligned}$$

In particular, multiplying z by $e^{i\phi}$ rotates the point z by an angle ϕ about the origin in a counter-clockwise direction when $\phi > 0$.

1.2 More examples of curious curves

1.2.1 *The unit square is a curve*

We define an initiator $f_0(t) = (t,t)$, the line segment in $U = [0,1]\times[0,1]$ connecting the points $(0,0)$ and $(1,1)$. The first replacement step $f_1(I)$, tracing the directed path from $(0,0)$ to $(1,1)$ following the numerical order of the squares

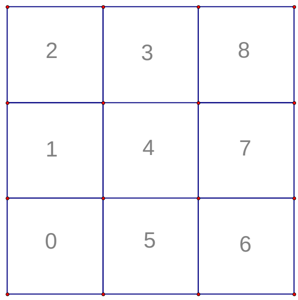


Fig. 1.8: Strategy for defining f_1.

in Figure 1.8, is shown on the left of Figure 1.9. If the replacement scheme is continued with rotations as suggested by $f_2(I)$ on the right of Figure 1.9, the limiting curve will fill the entire square. Other illustrative drawings and the proof that every element in U is the image of a point in I are deferred to Chapter 3. Note also that U is a curve with area one.

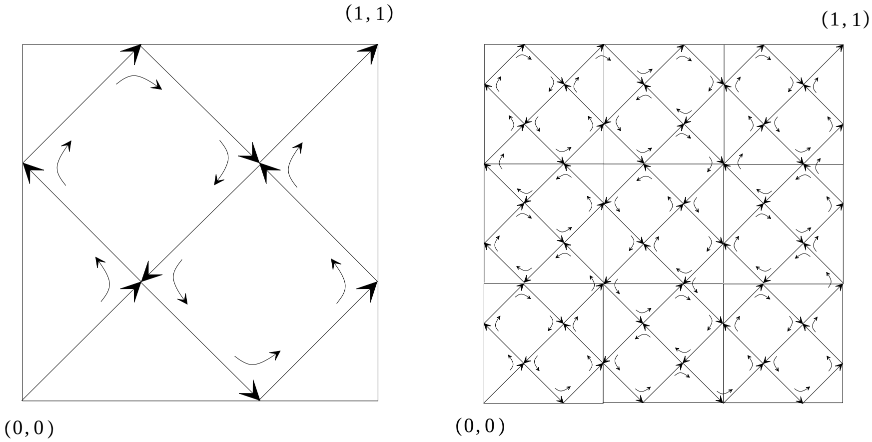


Fig. 1.9: Continuously mapping the unit interval onto the unit square.

1.2.2 *Iterated function systems produce curves*

The word *fractal*, coined by Mandelbrot [Mandelbrot (1983)], appears throughout our discussion of curves. Many fractals have the self-similarity property; that is, they are made up of parts which are, in some way, similar to the whole. For example, the Koch curve is made up of four scaled copies of itself. Fractal images can be generated by replacement strategies. A more complete discussion of this material is in Chapter 6, but we introduce some fundamental ideas here.

Let D be a closed subset of $\mathbb{R}^2$. A function f defined on D to $\mathbb{R}^2$ is called a **contraction map** on D if there exists a real number $e < 1$ such that $|f(s) - f(t)| \leq e\,|s - t|$ for each pair (s,t) in D.

Definition 1.1. An **iterated function system** (IFS) is a collection of contraction mappings $\{f_i,\ i = 1, \ldots, m\}$. A compact set F is **invariant** for the transformations $\{f_i\}$ if

$$F = \cup_{i=1}^{m} f_i\,(F)\,.$$

Invariant sets are often fractals.

In Chapter 6 we will see that an IFS has exactly one invariant set.

Example 1.1. The Koch curve K can be described as an invariant set. Let $f_1, f_2, f_3, f_4 : I \to \mathbb{R}^2$ be the four functions

$$f_1\,(z) = (1/3)\,z,\ f_2\,(z) = (1/3) + (1/3)\,e^{i\pi/3} z$$

$$f_3\,(z) = \left(1/2, \sqrt{3}/6\right) + (1/3)\,e^{-i\pi/3} z,\ f_4\,(z) = (2/3) + (1/3)\,z.$$

Then $f_1\,(I)\,, f_2\,(I)\,, f_3\,(I)$ and $f_4\,(I)$ are the four pieces of the generator in Figure 1.1. Furthermore, $f_1\,(K)\,, f_2\,(K)\,, f_3\,(K)$ and $f_4\,(K)$ are the four similar pieces of K so that

$$K = f_1\,(K) \cup f_2\,(K) \cup f_3\,(K) \cup f_4\,(K)\,.$$

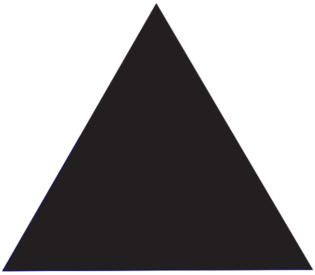
Thus K is invariant for the mappings of the IFS $\{f_1, f_2, f_3, f_4\}$.

In Chapter 2 we approach the Koch curve from a different perspective. There we generate a nested sequence $\{E_k\}$ of compact sets that converge to K.

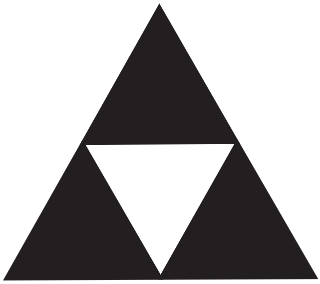
The previous example demonstrates that an IFS can produce a curve. Modifications of the Koch curve can also be written as iterated function systems that produce curves. Other examples of this process will be discussed in detail in Chapter 6 where we define the continuous function whose image is the invariant set. There we will develop the necessary definitions and topological and convergence properties. The culminating theorem for that discussion is

Theorem 1.1. *A connected invariant set (attractor) of an iterated function system is a curve.*

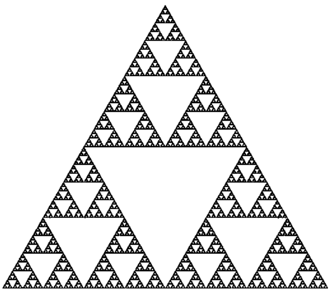
Example 1.2. The Sierpinski triangle is a curve. The Sierpinski triangle is the invariant set for three transformations in the plane. These three transformations define the process of replacing the equilateral triangle shown in Figure 1.10a with the three scaled equilateral triangles shown in Figure 1.10b. Repeating the replacement process produces an image represented in Figure 1.10c. By the theorem, the Sierpinski triangle is a curve. It is connected because each iterate is connected. (See Theorem 6.6.)



a. Equilateral triangle.



b. First step.

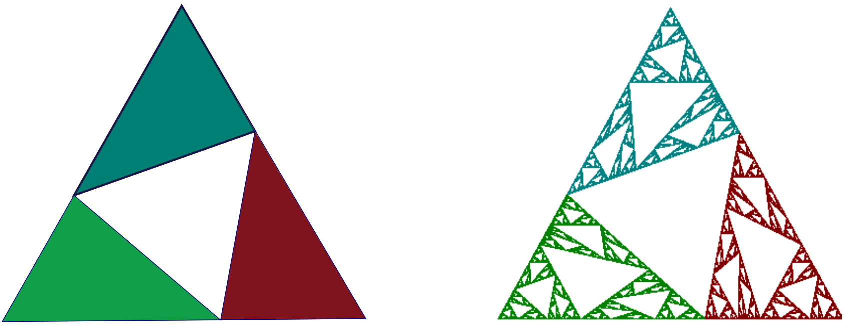


c. Sierpinski triangle.

Fig. 1.10: Construction of the Sierpinski triangle.

Example 1.3. We can modify the construction in the previous example and replace the original equilateral triangle with the image in Figure 1.11a. The result, another curve, is the irregular Sierpinski triangle represented in Figure 1.11b.

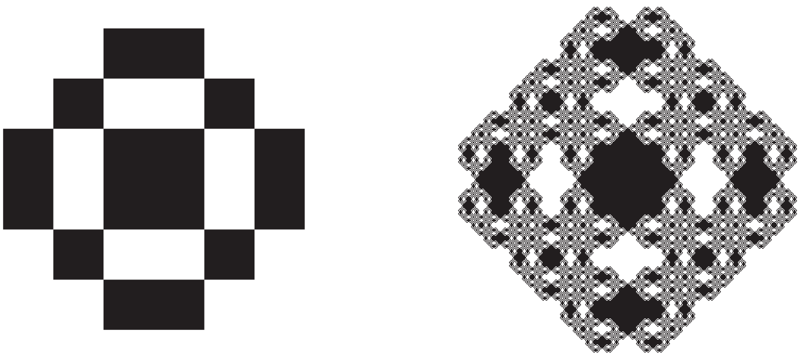
Example 1.4. Begin with the unit square and define 16 functions that will replace the square with the sixteen smaller squares shown in Figure 1.12a. Repeating the replacement process will produce the image shown in Figure 1.12b, another



a. Original triangle. b. Irregular Sierpinski triangle.

Fig. 1.11: Construction of the irregular Sierpinski triangle.

example of a curve. An exercise in Chapter 6 asks you to find the functions needed to produce the drawing. The limiting figure is called the "Badge and Hydrant" or "hydrant" curve.



a. First iteration. b. Badge and Hydrant curve.

Fig. 1.12: Construction of the Badge and Hydrant curve.

1.3 Construction of a family of Cantor sets

In this section we construct a family of Cantor sets. When such sets first appeared in the mathematical literature, they were categorized by some mathematicians as pathological and only of interest to those who studied mathematical oddities. They are now known to be appropriate tools for analyzing the behavior and structure of sets.

1.3.1 *Middle-thirds Cantor set*

Before constructing general Cantor sets, we construct the classical middle-thirds Cantor set. It is one of the best known and most easily constructed fractals, and it displays many typical fractal characteristics. We denote the classical middle-thirds Cantor set by C . This subset of $[0, 1]$ is obtained by successive deletion of middle-third open subintervals as follows:

$$\begin{aligned} I_0 &= [0,1] \\ I_1 &= [0,1/3] \cup [2/3,1] \\ I_2 &= [0,1/9] \cup [2/9,1/3] \cup [2/3,7/9] \cup [8/9,1] \\ &\vdots \end{aligned}$$

Note that I_n is I_{n-1} with the middle open third of each subinterval removed. For each $n \in \mathbb{N}$, I_n is the union of 2^n closed intervals, each of length $1/3^n$, and $I_0 \supset I_1 \supset I_2 \supset \cdots$. We define $C = \cap_{n \geq 0} I_n$. The construction is illustrated in Figure 1.13.

The Cantor set C can also be described as the invariant set for the IFS $\{f_1, f_2\}$, where

$$f_1(x) = \frac{1}{3}x; \; f_2(x) = \frac{1}{3}x + \frac{2}{3}.$$

Then $f_1(C)$ and $f_2(C)$ are just the left and right 'halves' of C, so that $C = f_1(C) \cup f_2(C)$. Thus C is invariant for the mappings f_1 and f_2. In Chapter 3 we will show the remarkable fact that any compact set in $\mathbb{R}^2$ is a continuous image of C.

Because we begin the construction of C with three subintervals of I, it is convenient to use base 3 representation of elements in I. Any number $x \in I$ has a base 3 expansion of the form $x = \sum_{j=1}^{\infty} x_j/3^j$, where $x_j \in \{0, 1, 2\}$; $x = 0.x_1x_2 \cdots$ base 3. For example, $1/3 = 0.1$ base 3. Normally, we use an underbar to denote continuing repetition of digits in a base 3 expansion; and so, using the properties of convergent geometric series, we see that $1/3 = 0.0\underline{2}$ base 3 also. Any base 3 representation of a number for which $x_j = 0$ for all j larger than some positive integer N is called a terminating expansion. A number with a terminating expansion has two base 3 representations. For example, $2/3 = 0.2\underline{0}$ base 3 $= 0.1\underline{2}$ base 3. In such cases you can use whichever expansion is convenient. With this exception it can be shown that every $x \in I$ has a unique base 3 representation. From the construction of C, we see that each element of C can be represented as a string of 0's and 2's. Furthermore, every string of 0's and 2's represents a point in C. For example $1/4 = 0.\underline{02}$ is an element of C.

Fig. 1.13: Initial steps in the construction of the Cantor set.

1.3.2 *Construction of generalized Cantor sets*

You will frequently be encouraged to refer to Figure 1.14 where we have labeled the x-coordinates of a few points that are referenced in the construction. You should be able to determine any unlabeled points. To construct the sets depicted in Figure 1.14 we begin with the unit square

$$U = [0,1] \times [0,1] = \{(x,y) : 0 \leq x \leq 1, 0 \leq y \leq 1\}$$

in the plane and remove a sequence of triangles from this square according to the following steps.

Example 1.5. Construction of a family of Cantor sets.

Step 1 Remove the inside of the triangle in Figure 1.14a with vertices $\left(\frac{1}{3},0\right), \left(\frac{1}{2},1\right)$ and $\left(\frac{2}{3},0\right)$ along with the open interval $(1/3, 2/3)$. Notice that we keep the two line segments $\left[\left(\frac{1}{3},0\right), \left(\frac{1}{2},1\right)\right]$ and $\left[\left(\frac{2}{3},0\right), \left(\frac{1}{2},1\right)\right]$ that compose the sides of the triangle and that nothing is removed from the interval $[(0,1),(1,1)]$ composing the top edge of U. In what follows we will simply write "Remove the triangle" in situations where we remove the interior and base leaving the base vertices of a triangle. Also, note that removing the interval $(1/3, 2/3)$ from the interval $[0,1]$ is the first step in the construction of the classical Cantor set. Let S_1, the part of Figure 1.14a shaded light gray, denote the part of the square U that remains after the triangle is removed.

Step 2 Remove 2 triangles with top vertices $\left(\frac{1}{4},1\right)$ and $\left(\frac{3}{4},1\right)$ and base segments $(1/9, 2/9)$ and $(7/9, 8/9)$, respectively. The two base segments are the open middle thirds of the two base intervals $\left[0,\frac{1}{3}\right]$ and $\left[\frac{2}{3},1\right]$ that remain after *Step 1*. Let S_2, shaded light gray in Figure 1.14b, denote the subset of S_1 that remains after these two triangles are removed. Again, nothing is removed from the interval $[(0,1),(1,1)]$.

Step 3 Remove the 2^2 triangles with top vertices $\left(\frac{2j-1}{2^3},1\right)$, $j=1,\ldots,2^2$, and corresponding base segments that are the open middle thirds of the four base intervals of length $\frac{1}{3^2}$ that remain after *Step 2*. Let S_3, shaded gray in Figure 1.14c, denote the subset of S_2 that remains after these 2^2 triangles are removed.

Step n Continue iteratively so that at *Step n* we remove 2^{n-1} triangles, each with a base segment of length $\frac{1}{3^n}$ (an open middle third of a base interval of length $\frac{1}{3^{n-1}}$ that remains after *Step n–1*) and a top vertex $\left(\frac{2j-1}{2^n},1\right)$, $j=1,\ldots,2^{n-1}$, in the middle of a corresponding interval of length $\frac{1}{2^{n-1}}$ on the top. A set S_n remains. *Notice that the points at the vertices remain.*

After completing a step for every positive integer n, we have removed an infinite collection of triangles from U. Denote the set that remains by S. For $0\le h\le 1$, denote by C_h the intersection of S with the horizontal line $y=h$, which is distance h above the base of the square. (The line $y=h$ is the set $\{(x,h):x\in\mathbb{R}\}$.) Figure 1.14c shows the line for $h=1/3$.

(1) When $h=0$, we have the part of the base of U which has not been removed; this set C_0 is the classical (standard, middle-thirds) Cantor set in the interval $[0,1]$.
(2) When $0<h<1$, C_h is a Cantor set on the interval $[(0,h),(1,h)]=\{(x,h):x\in I\}$.
(3) When $h=1$, we have the entire top edge of U.

1.3.3 *The length of the Cantor set C_0*

We begin by setting $h=0$ so that we will calculate the length of the standard middle-thirds Cantor set. At **Step 1**, a single segment of length $1/3$ is removed from the interval $[0,1]$ on the x-axis. At **Step 2** two segments of length $1/3^2$ are removed so that the total length of the removed segments is $1/3+2(1/3^{2)}=5/9$. In general, at **Step n**, 2^{n-1} segments of length $1/3^n$ are removed. The total length ℓ_0 of the segments removed is

$$\ell_0=\frac{1}{3}+2\frac{1}{3^2}+2^2\frac{1}{3^3}+\cdots=\frac{1}{3}\left(1+\frac{2}{3}+\left(\frac{2}{3}\right)^2+\cdots\right)=1.$$

Thus, the length of the Cantor set is

$$\mathrm{len}(C_0)=\mathrm{len}([0,1])-\ell_0=0.$$

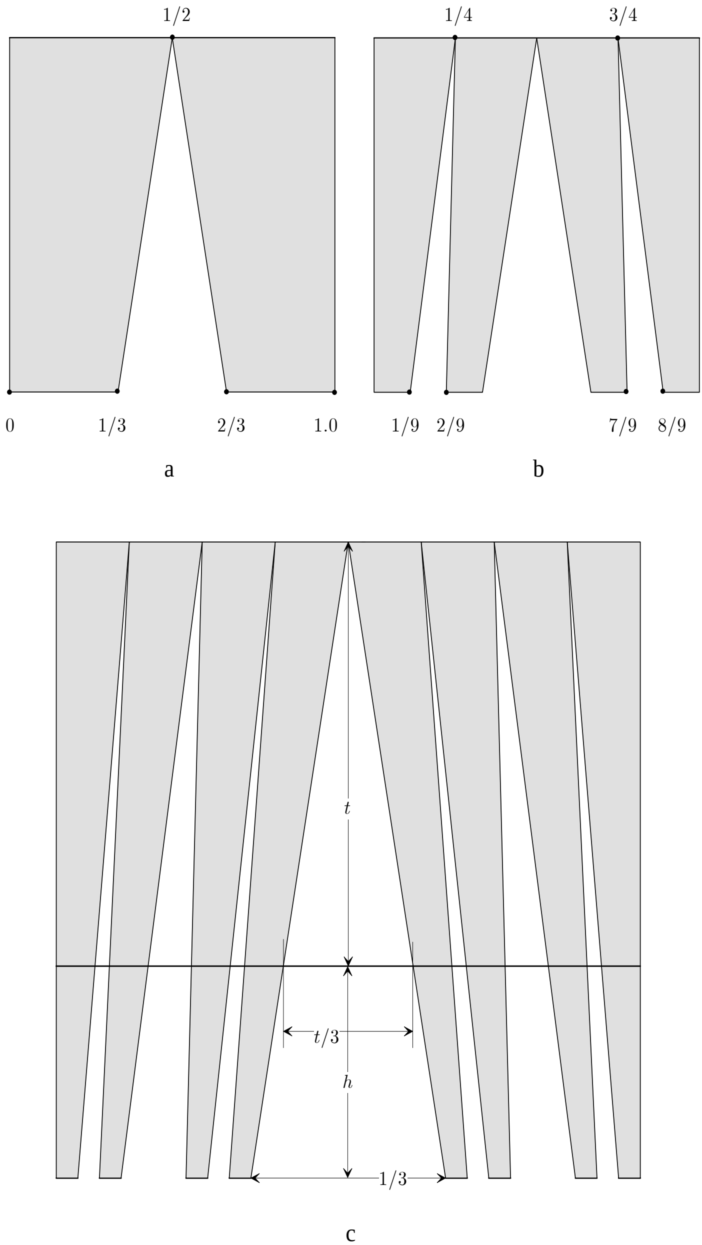


Fig. 1.14: A family of Cantor sets.

1.3.4 *The sets C_h*

Now put $t = 1 - h$ and examine Figure 1.14c which, coupled with the above calculations, should provide insights to the following discussion. At **Step n**, the length of each of the 2^{n-1} segments that are removed from the line $y = h$ is $t/3^n$. Thus (referring to the calculation given above for the case $t = 1$ or $h = 0$), t is the total length of the segments that are removed from the interval $[(0, h), (1, h)] \subset U$. Consequently, $\text{len}(C_h) = 1 - \ell_t = 1 - t = h$. In particular, $\text{len}(C_1) = 1$ which is not surprising since $C_1 = I$.

The set C_h has length h and contains the same number of points as I, but it contains no interval when $0 \leq h < 1$. We will use these properties of the sets C_h to construct curves in the plane. In Chapter 3 we further verify that $\text{area}\,(C_h \times C_h) = h^2$ and use this set product to construct simple curves with positive area.

1.4 What is not a curve?

The standard Cantor set $C = C_0$ shares several properties with the interval $[0, 1]$. Both sets are compact, and they contain the same number of points. However, they differ drastically with respect to connectedness; while I is connected, the only connected subsets of C are sets containing a single point. By Definition C.17, C is totally disconnected. The Cantor set is not a curve.

In Chapter 3 we discuss the following two examples and some striking modifications of them. Each example is a compact, connected set that is not a curve. The first (Example 3.3) is the geometric comb

$$GC = I \cup \{(0, y) : 0 \leq y \leq 1\} \cup \bigcup_{n \geq 0} \{(1/2^n, y) : 0 \leq y \leq 1\}.$$

(See Figure 1.15.)

A *generalized curve* is a continuous image of the closed half line $[0, \infty)$. We will demonstrate that GC is a generalized curve.

The second (Example 3.4) is the Cantor comb

$$CC = I \cup \bigcup_{x \in C} \{(x, y) : 0 \leq y \leq 1\} = I \cup (C \times I),$$

where the bases of the tines are the points in the Cantor set. We will show that the Cantor comb is not a generalized curve but as a seemingly minor modification of CC, shown in Figure 1.16, is a curve. Likewise, by the end of Chapter 3 you will be able to define a continuous map from I onto the curve illustrated by Figure 1.17. This curve is composed of the unit circle R_1 and a sequence of larger

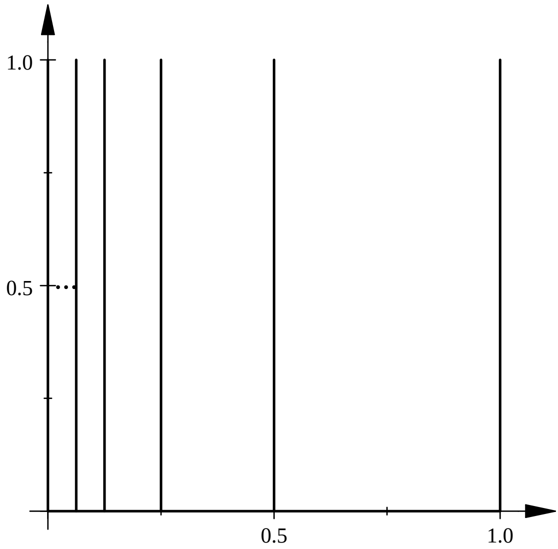


Fig. 1.15: The geometric comb is an example of a compact, connected set that is not a curve.

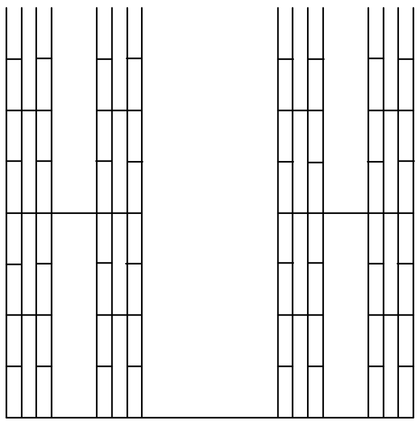


Fig. 1.16: Modified Cantor comb.

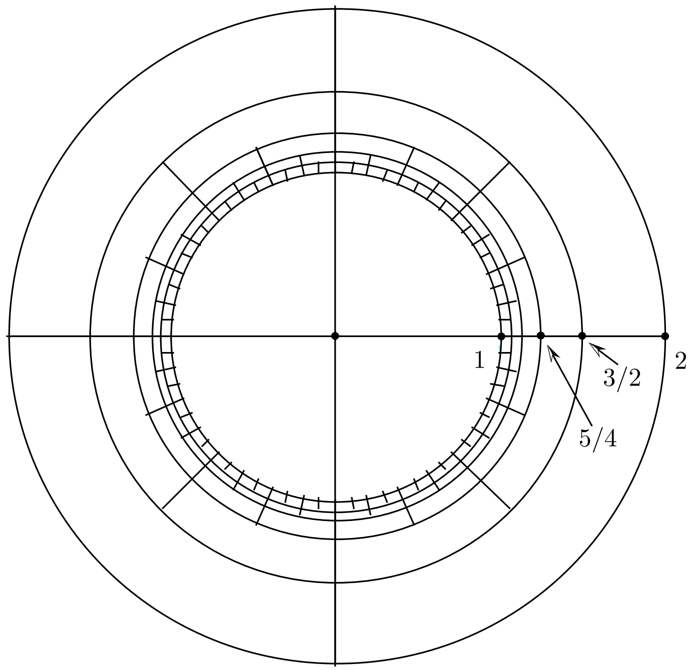


Fig. 1.17: A circular shaped curve.

concentric circles $R_{1+1/2^n}$ of radius $1 + 1/2^n$ for $n = 0, 1, 2, 3, 4, \cdots$, together with radial spokes connecting the unit circle to the larger circles. The first six circles and their spokes are displayed in Figure 1.17.