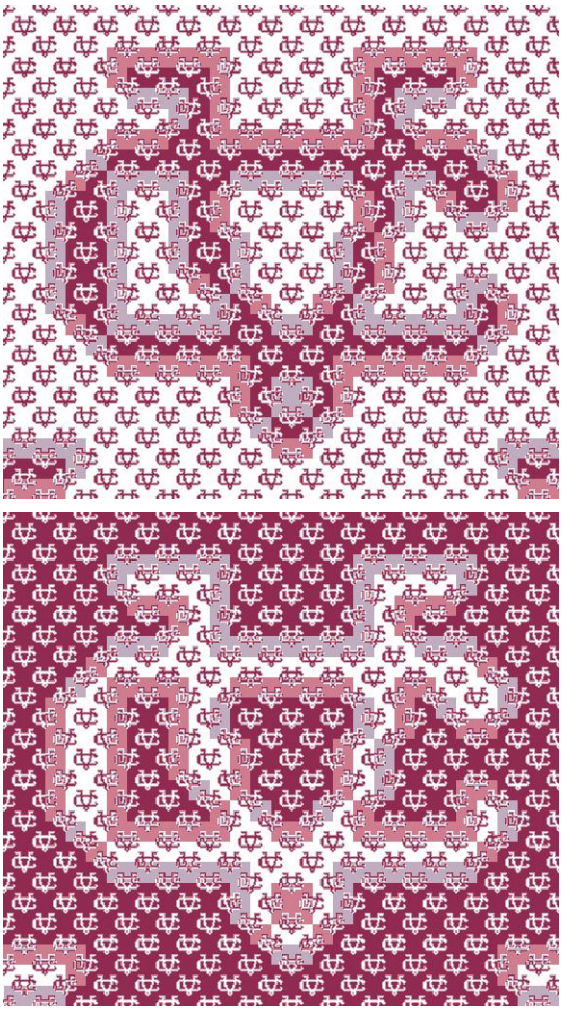
Lecture Notes in Fractal Geometry



May 4, 2020

Natalie Priebe Frank

Vassar College, Poughkeepsie, NY 12603
E-mail address: nafrank@vassar.edu

Contents

Chapter 1. Introduction 1
1.1. Classic examples 1
1.2. A geometric approach to transformations. 4
1.3. Collage maps: the building blocks of iterated function systems. 6
1.4. Affine transformations in two dimensions: a geometric approach 8
1.5. Collage maps in two dimensions 11
1.6. What is a fractal? 15
1.7. Exercises 16

Chapter 2. Hausdorff metric and the space of fractals 19
2.1. A tiny bit of point-set topology 19
2.2. $\mathcal{H}(X)$, the space of fractals. 21
2.3. Metric spaces 21
2.4. Hausdorff metric on $\mathcal{H}(X)$ 23
2.5. Exercises 28

Chapter 3. Iterated Function Systems 29
3.1. Collage maps as contractions on the space of fractals 29
3.2. Existence of Attractors for Iterated Function Systems 33
3.3. Two computer algorithms for IFS fractals 34
3.4. The Collage Theorem 36
3.5. Exercises 37

Chapter 4. Dimensions 39
4.1. Motivating examples, or, Fun with length and area. 39
4.2. The idea of fractal dimension 42
4.3. Similarity dimension 46
4.4. Box-counting dimension 51
4.5. When the dimensions are equivalent 54
4.6. Exercises 56

Chapter 5. Julia Sets 57
5.1. Basic example: dynamical systems in $\mathbb{R}$ 58
5.2. Classic example: dynamical systems in $\mathbb{C}$. 60
5.3. The Escape Time Algorithm 62
5.4. Exploring some more, algebraically and Mathematica-ally. 62

Chapter 6. Mandelbrot Sets 65
6.1. Using the escape-time algorithm to plot the Mandelbrot set. 66

Bibliography 69

CHAPTER 1

Introduction

In this course we will study the complicated geometric structures that arise from simple natural processes and are known as fractals. We will focus on fractals obtained in two distinct ways. One way, which is the subject of the first several sections of these notes, is when they arise as limit points of "iterated function systems". The other is when they arise as collections of points that exhibit certain behaviors under repeated iteration of functions. Examples of the former are shown next; the latter will be discussed in the Julia and Mandelbrot sections.

With mathematical precision we will discuss how fractals are constructed, and we will prove conditions that guarantee their existence. We will discuss geometric properties such as self-similarity and fractal dimensions. We will learn probabilistic algorithms that allow images to be generated efficiently on a computer, and we will spend time on the computer making our own pictures. To give a flavor of what sort of structures we will be discussing, we begin with three classic examples of fractals.

1.1. Classic examples

EXAMPLE 1.1. This example begins with the closed interval $C_0 = [0, 1]$ and proceeds in stages; the *middle-thirds Cantor set* $\mathcal{C}$ is the limit[a] of the process. In the first stage, the middle third is removed to obtain the set $C_1 = [0, 1/3] \cup [2/3, 1]$. In the second stage, the middle thirds of the remaining sets are removed to obtain the set $C_2 = [0, 1/9] \cup [2/9, 1/3] \cup [2/3, 7/9] \cup [8/9, 1]$. To obtain the set C_3, remove the middle thirds of the remaining intervals in C_2. The process is illustrated geometrically in figure 1.

It is clear that not all of the points in $[0, 1]$ are removed, so $\mathcal{C}$ is not empty. It is also true, but not necessarily clear, that there are no intervals in $\mathcal{C}$, but yet that it is uncountable. Interestingly, no point $c \in \mathcal{C}$ is 'isolated' from the rest of it: in any interval around c there are other points from $\mathcal{C}$.

Although he did not discover the set, Georg Cantor introduced it to the mathematical public in a paper in 1883 as an example of a set that is totally disconnected but has no isolated points. This general definition is what is meant by topologists when they throw around the term "Cantor set".

[a] We will make the notion of convergence of sets more precise in Chapter 2.

EXAMPLE 1.2. The *Koch curve* lives in the unit square

$$S = [0,1] \times [0,1] = \{(x,y) \in \mathbb{R}^2 \text{ such that } 0 \leq x \leq 1, 0 \leq y \leq 1\}.$$

The usual way to construct the curve is to begin with a line segment, then replace it with four line segments that are each 1/3 as long as the original and placed as in the figure marked K_1 below. To construct K_2, four copies of K_1 are made, scaled by 1/3 and placed in the same configuration as the segments that made K_1. At the next stage, four copies of K_2 are made, scaled by 1/3, and placed in the same configuration yet again. The first several steps appear in figure 2.

Helge von Koch introduced his curve in a paper in 1906 as an example of a curve that is everywhere continuous but nowhere differentiable. It can be parameterized in the form $x = f(t), y = g(t)$. We will see in Chapter 4 that it has infinite length yet zero volume, and a fractal dimension of $\ln 4/\ln 3$, which is strictly between 0 and 1, despite it being a parameterized curve.

EXAMPLE 1.3. The *Sierpinski triangle* lives in the set

$$\{(x,y) \in \mathbb{R}^2 \text{ such that } 0 \leq x \leq 1, 0 \leq y \leq 1, x+y \leq 1\}$$

and you can think about its construction in either the Cantor set or Koch curve ways. In the middle-thirds sense, one can see S_1 of figure 3 as obtained by removing the 'middle' triangle from S_0, to leave three triangles that are each half the size of the original. From S_1 we obtain S_2 by removing the middle triangles of the remaining triangles. We continue removing the middle triangles of the remainder ad infinitum to obtain the Sierpinski triangle $\mathcal{S}$.

Alternatively, we can think of constructing S_1 by making three copies of S_0, shrunk by a factor of two, and placed as shown. To make S_2, we shrink three copies of S_1 and place them as prescribed again. We continue to this process forever to obtain $\mathcal{S}$.

Waclaw Sierpinski introduced this triangle in 1915 using a construction as a curve that is distinct from the two methods discussed here. We will see that its fractal dimension is $\ln 3/\ln 2$, again between 1 and 2.

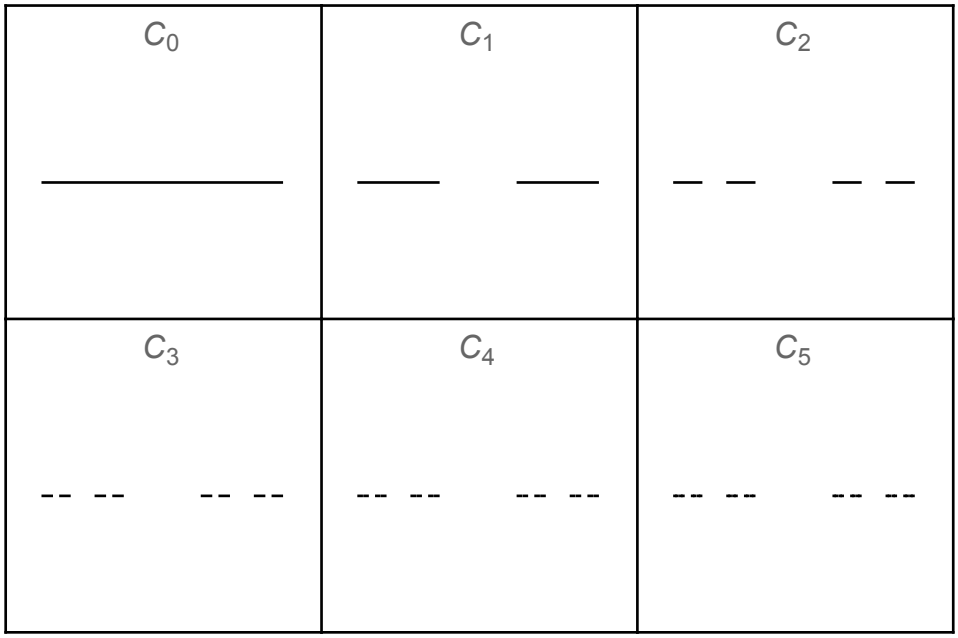


FIGURE 1. Constructing the middle-thirds Cantor set.

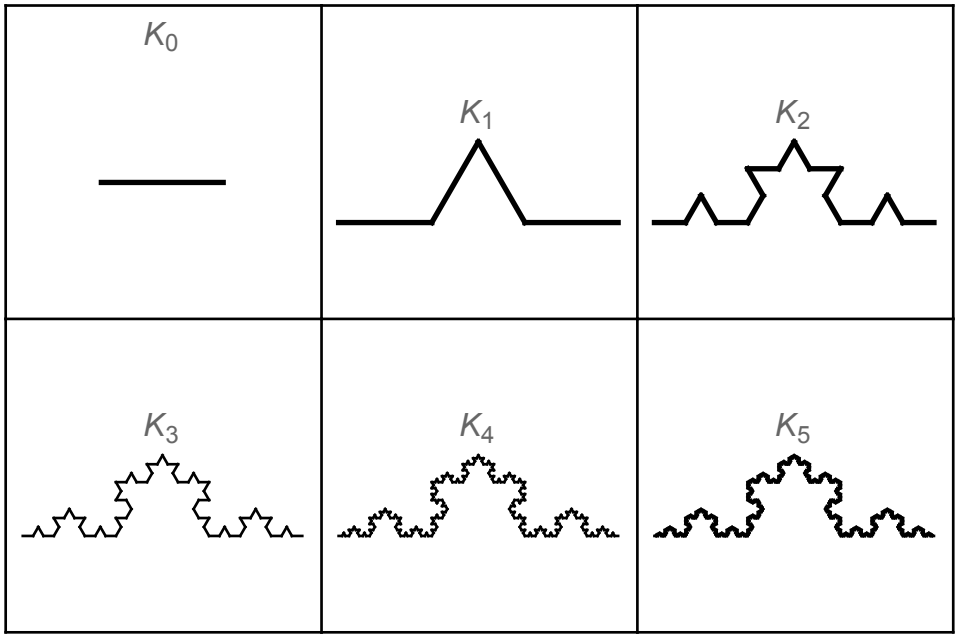


FIGURE 2. Constructing the Koch curve.

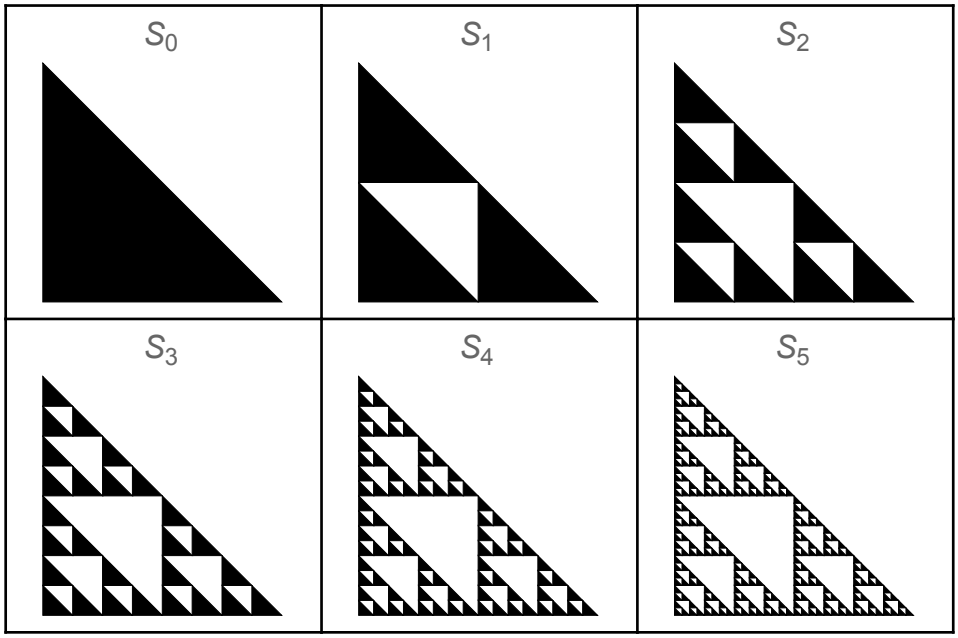


FIGURE 3. Constructing the Sierpinski triangle.

1.2. A geometric approach to transformations.

Transformations are essential in constructing fractals and are the fundamental building blocks of iterated function systems. In this section we give fundamental definitions to be used througout the course, and some one-dimensional examples are provided to help build your geometric intuition.

In general the symbol X will be used to denote a mathematical space where fractals live. Always X is a set; in the early going it will be always be $\mathbb{R}$ or $\mathbb{R}^2$ or $\mathbb{R}^n$ or $\mathbb{C}$ or a subset of those. Later on we will make X be a "compact metric space", but for now just imagine some subset of the real numbers. A transformation is a mapping defined from X to itself as follows.

DEFINITION 1.4. A *transformation* $T : X \to X$ is a function with domain X and codomain X. That is, T assigns to each element $x \in X$ an element $y = T(x) \in X$.

It is natural in fractal geometry to use certain notation and terminology from the world of dynamical systems, such as these next two definitions.

DEFINITION 1.5. Let $T : X \to X$ be a transformation and let $x \in X$. We define $T^2(x) = T \circ T(x) = T(T(x))$ and, in general,

$$T^n(x) = T \circ T \circ \cdots \circ T(x) = T(T(\cdots T(x)) \cdots)) \text{ (n times).}$$

We define the *orbit* of x to be the sequence

$$\mathcal{O}(x) = \{x, T(x), T^2(x), T^3(x), ...\}.$$

DEFINITION 1.6. Let $T : X \to X$ be a transformation. We say $x \in X$ is a *fixed point* of T if $T(x) = x$.

In the following exercise we are going to consider a transformation that has a geometric interpretation you've known for years: its graph is a line. However, we need a different geometric interpretation, which is to think of the transformation as taking points from the domain back into the domain and visualizing the orbits.

EXERCISE 1.7. Define the transformation $T : [0, 1] \to [0, 1]$ by $T(x) = \frac{-1}{2}x + \frac{3}{4}$.

(1) Compute the first four elements of the orbit of $x = 0$.
(2) Visualize the orbit as follows: draw yourself a fairly large copy of the unit interval $[0, 1]$ and place the points from the orbit on it. Connect each point to the one that follows it with an arrow in a follow-the-bouncing-ball sort of way. (Your arrows will live above or below your unit interval, depending on your artistic choices.)
(3) Repeat the previous parts using $x = 1$.
(4) Compute the fixed point(s) of T algebraically.[a]
(5) Reflect briefly on how the orbits seem to relate to the fixed point.

[a] The word "algebraically" tends to mean to do a computation or use mathematical symbols in some other way. Algebraic arguments tend to feel very precise and rigorous. I will also often ask you for "geometric" arguments; these can include graphs or sketches or they can be constructed from the language of geometric objects.

The transformation of exercise 1.7 is special in several ways, the most obvious being that its graph is a line. But it also has the property that it brings points closer together as it is applied.

DEFINITION 1.8. Let X be a subset of $\mathbb{R}, \mathbb{R}^n$, or $\mathbb{C}$.[a] The transformation $T : X \to X$ is called a *contraction* if there is some constant $c \in [0, 1)$ such that

$$|T(x) - T(y)| \leq c|x - y| \text{ for all } x, y \in X.$$

The number c, which is not unique, is called a *contractivity factor* for T.

[a] This definition applies to transformations on any metric space, so when we learn about those it will apply retroactively.

EXERCISE 1.9. Write as formal a proof as you can muster that the transformation from exercise 1.7 is a contraction. Note: in your "scratch work" to prepare for the proof, you should identify a good value for the contractivity factor c.

Another useful feature of the transformation we have been considering is that it can be broken down as the composition of two even simpler transformations: $T_1(x) = \frac{-1}{2}x$ and $T_2(x) = x + \frac{3}{4}$. Take a moment on scratch paper right now to verify that $T(x) = T_2 \circ T_1(x)$ algebraically. What this means is that you can see two separate geometric actions that move your point x from its location to the location specified by $T(x)$: first the size of x is halved and it is flipped across the origin, then it is moved over to the right by $\frac{3}{4}$. This insight might not seem like a big deal to you right now, but it is particularly useful when you are applying your transformation to entire subsets of X rather than just points.

DEFINITION 1.10. Let $T : X \to X$ be a transformation and let $A \subseteq X$. We define the *image of A under T* to be

$$T(A) = \{T(a) \text{ such that } a \in A\}.$$

Similarly, for $n = 2, \ldots$ we define *image of A under T^n* to be

$$T^n(A) = \{T^n(a) \text{ such that } a \in A\}.$$

Put another way, $T(A)$ is the **set** you get by applying T to every element of A. It is useful to be able to think of a transformation applied to a set geometrically as well as algebraically, as in this next exercise.

EXERCISE 1.11. Let T be as in exercise 1.7 and let $A = [0, 1]$. Compute and sketch $T(A)$ in the following two different ways, the first more algebraic and the second more geometric.

(1) Plug the endpoints of A into T. If that is enough for you to know what $T(A)$ is, write the answer now. If not, choose some other points $a \in A$ and compute $T(a)$. Try to express $T(A)$ in a mathematically familiar way.
(2) Second, think of T as the composition $T_2 \circ T_1$ as discussed in the paragraph before definition 1.10. Construct a sketch for what T_1 does to A, then make a sketch what T_2 does to that.
(3) Repeat the process (both ways) to compute $T^2(A)$ and $T^3(A)$.

1.3. Collage maps: the building blocks of iterated function systems.

In Chapter 3 we will give the formal definition of an iterated function system, but we can start working with them right away by defining the collage map given by a finite set of transformations. We can use the transformations together to make a mapping that takes **subsets of** X to **subsets of** X:

DEFINITION 1.12. Let $T_i : X \to X$ be a transformation for all $i = 1, 2, ..., k$ and let $A \subseteq X$. The *collage map* defined by these transformations is given by

$$\mathcal{T}(A) = \bigcup_{i=1}^{k} T_i(A) = T_1(A) \cup T_2(A) \cup \cdots \cup T_k(A)$$

We may write $\mathcal{T} = T_1 \cup T_2 \cup \cdots \cup T_k$.

Notice that since $A \subseteq X$ and each T_i takes X to itself it must be true that $T_i(A) \subseteq X$ for each i. Since $\mathcal{T}(A)$ is a union of such subsets of X it must also be a subset of X. That means that it is perfectly appropriate to define $\mathcal{T}^2(A) = \mathcal{T} \circ \mathcal{T}(A)$, and $\mathcal{T}^n(A) = \mathcal{T} \circ \mathcal{T} \circ \cdots \circ \mathcal{T}(A)$ (n times) as before.

EXAMPLE 1.13. Let $X = [0, 10]$, $T_1(x) = \frac{x}{5}$, and $T_2(x) = \frac{x}{5} + 5$, and define the collage map $\mathcal{T} = T_1 \cup T_2$. Suppose that $A = [5, 10]$. Then $T_1(A) = [1, 2]$ and $T_2(A) = [6, 7]$, hence $\mathcal{T}(A) = [1, 2] \cup [6, 7]$. The set $\mathcal{T}(A)$ is pictured at the top of figure 4.

To compute $\mathcal{T}^2(A) = \mathcal{T}([1, 2] \cup [6, 7])$, we need to consider what T_1 and T_2 each do to the set $\mathcal{T}(A) = [1, 2] \cup [6, 7]$, then union the result. Now $T_1([1, 2] \cup [6, 7]) = [1/5, 2/5] \cup [6/5, 7/5]$, and $T_2([1, 2] \cup [6, 7]) = [5 + 1/5, 5 + 2/5] \cup [5 + 6/5, 5 + 7/5]$. Thus $\mathcal{T}^2(A)$ is the union of four intervals, shown in the middle line of figure 4. In the bottom of figure 4 we show $\mathcal{T}^3(A)$, which consists of eight intervals. At each stage, the leftmost interval's left endpoint is getting closer to 0. Can you determine where some other endpoints seem to be tending?

EXERCISE 1.14. In general, the collage map technically isn't a transformation from a space X to itself even though its component maps T_i all are. Why is that? (Hint: given the set $\{x\}$ containing a single point $x \in X$, what kind of object is $\mathcal{T}(\{x\})$?)

The collage map is actually a transformation on the "space of fractals" $\mathcal{H}(X)$, which we will define properly in chapter 2.

EXERCISE 1.15. Let $X = [0, 1]$ and define the transformations $T_1(x) = \frac{1}{3}x$ and $T_2(x) = \frac{1}{3}x + \frac{2}{3}$. Consider the collage map given by $\mathcal{T} = T_1 \cup T_2$.

(1) Let $A = X$ (i.e. the whole unit interval). Compute the sets $\mathcal{T}(A)$ and $\mathcal{T}^2(A)$ algebraically.
(2) Make sketches of $A, \mathcal{T}(A)$, and $\mathcal{T}^2(A)$.
(3) Repeat with $A = \left\{\frac{1}{2}\right\}$ (i.e. a set with one point in it).
(4) In each case, consider whether $\mathcal{T}^n(A)$ seems to tend to a specific set A_0 as n goes to infinity. If so, how do the limit sets compare?

EXERCISE 1.16. Let $X = [0, 6]$, $T_1(x) = \dfrac{x}{2}$, and $T_2(x) = \dfrac{2x}{3} + 2$, and define the collage map $\mathcal{T} = T_1 \cup T_2$.

(1) Let $A = X$. Compute the sets $\mathcal{T}(A)$ and $\mathcal{T}^2(A)$ algebraically.
(2) Make sketches of $A, \mathcal{T}(A)$, and $\mathcal{T}^2(A)$.
(3) Repeat with $A = \{6\}$.
(4) In each case, consider whether $\mathcal{T}^n(A)$ seems to tend to a specific set A_0 as n goes to infinity. If so, how do the limit sets compare?

EXERCISE 1.17. Let $X = [0, 1]$, $T_1(x) = \dfrac{x}{4}$, and $T_2(x) = \dfrac{x}{2} + \dfrac{1}{4}$, and define the collage map $\mathcal{T} = T_1 \cup T_2$.

(1) Let $A = X$. Compute the sets $\mathcal{T}(A)$ and $\mathcal{T}^2(A)$ algebraically.
(2) Make sketches of $A, \mathcal{T}(A)$, and $\mathcal{T}^2(A)$.
(3) Repeat with $A = \{0\}$. For this one you should try to go to $\mathcal{T}^4(A)$.
(4) In each case, consider whether $\mathcal{T}^n(A)$ seems to tend to a specific set A_0 as n goes to infinity. If so, how do the limit sets compare?

EXERCISE 1.18. Let $X = [0, 10]$, $T_1(x) = \dfrac{x}{10} + 1, T_2(x) = \dfrac{x}{10} + 4$, and $T_3(x) = \dfrac{x}{10} + 7$, and define the collage map $\mathcal{T} = T_1 \cup T_2 \cup T_3$.

(1) Let $A = X$. Compute the sets $\mathcal{T}(A)$ and $\mathcal{T}^2(A)$ algebraically.
(2) Make sketches of $A, \mathcal{T}(A)$, and $\mathcal{T}^2(A)$.
(3) Repeat with $A = \{7\}$.
(4) In each case, consider whether $\mathcal{T}^n(A)$ seems to tend to a specific set A_0 as n goes to infinity. If so, how do the limit sets compare?

EXERCISE 1.19. Let $X = [0, 1]$, $T_1(x) = \dfrac{x}{2}$, and $T_2(x) = \dfrac{x}{2} + \dfrac{1}{2}$, and define the collage map $\mathcal{T} = T_1 \cup T_2$.

(1) Let $A = X$. Compute the sets $\mathcal{T}(A)$ and $\mathcal{T}^2(A)$ algebraically.
(2) Make sketches of $A, \mathcal{T}(A)$, and $\mathcal{T}^2(A)$.
(3) Repeat with $A = \left\{\dfrac{1}{3}\right\}$. For this one you should try to go to $\mathcal{T}^4(A)$, at least with sketches.
(4) In each case, consider whether $\mathcal{T}^n(A)$ seems to tend to a specific set A_0 as n goes to infinity. If so, how do the limit sets compare?

1.4. Affine transformations in two dimensions: a geometric approach

Fractals are more fun in two dimensions and so we need to get serious about studying $\mathbb{R}^2$-transformations. Although we have many types to choose from, as in the one-dimensional case we choose to stick with simple "affine" transformations. In one dimension, an affine transformation is any transformation of the form $T(x) =$

$ax + b$, in other words a linear transformation followed by a translation. Every example from the previous sections was of this form, and the shrink-and-move intuition we developed there will be useful in two and higher dimensions also. To make a precise definition of an affine map in two dimensions we first need to set some notation.

Let us denote elements of $\mathbb{R}^2$ in any of three ways, which we will use interchangeably:

$$\vec{x} = (x_1, x_2) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$$

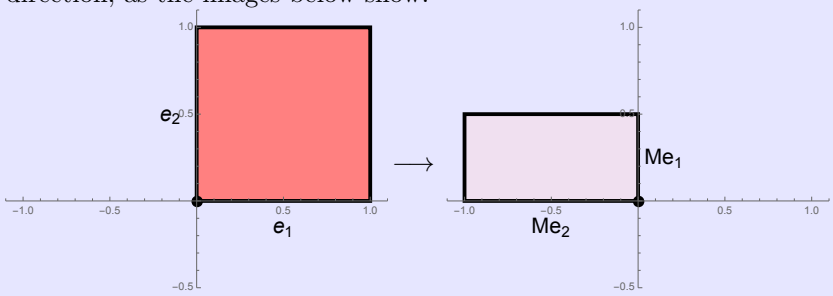
A *linear* transformation from $\mathbb{R}^2$ to itself is given by matrix multiplication and can be expressed in many ways, including these:

$$M\vec{x} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = (ax_1 + bx_2, cx_1 + dx_2)^1 \tag{1.1}$$

When thinking about linear transformations it is particularly useful to think about how they affect the standard basis vectors, which in this course we will denote by $\vec{e_1} = (1, 0)$ and $\vec{e_2} = (0, 1)$. Take a moment to use equation 1.1 to verify that $M\vec{e_1} = (a, c)$ and $M\vec{e_2} = (b, d)$. That is to say, the first column of M tells us what M does to the first standard basis vector, and the second column of M tells us what it does to the second.

A quick and dirty way to visualize matrix multiplication is to always do the following. Consider the unit square, which can be seen as having corners at the origin, $\vec{e_1}, \vec{e_1} + \vec{e_2}$, and $\vec{e_2}$. Then M sends the unit square to the parallelogram having corners at $M(0, 0), M\vec{e_1}, M(\vec{e_1} + \vec{e_2})$,and $M\vec{e_2}$. Take a moment to verify that if $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, then these corners are $(0, 0), (a, c), (a + b, c + d)$, and (b, d), respectively.

EXAMPLE 1.20. Suppose that $M = \begin{pmatrix} 0 & -1 \\ 1/2 & 0 \end{pmatrix}$. Then we can see that the unit square is rotated by $\pi/2$ counterclockwise and squished by half in one direction, as the images below show.



[1] If you have not seen matrix multiplication before you should take this as a definition.

EXERCISE 1.21. Consider the following matrices:

$$\begin{pmatrix} 0 & -1/2 \\ 1/2 & 0 \end{pmatrix} \quad \begin{pmatrix} 1/2 & -1/2 \\ 1/2 & 1/2 \end{pmatrix} \quad \begin{pmatrix} 0 & 1/3 \\ 1/3 & 0 \end{pmatrix} \quad \begin{pmatrix} 1/2 & 1/2 \\ 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1/2 & 1 \\ 0 & 1/2 \end{pmatrix}$$

(1) Sketch how each matrix acts on the unit square, like in example 1.20.
(2) Make a winking smiley face (or some other asymmetric image of your choice) inside your unit square and show how each matrix transforms it.
(3) For each matrix, determine whether the transformation given by $T(\vec{x}) = M\vec{x}$ is a contraction. If it is, give a contraction factor. If not, exhibit two vectors $\vec{x}$ and $\vec{y}$ that are not brought closer by the transformation.

We know by the remark immediately after definition 1.1 that the first column of a matrix is where it sends the first standard basis vector, and that the second column is where it sends the second. This means that if you know where a matrix sends the basis vectors then you know the matrix itself. That is, you can use geometric descriptions of linear transformations to make matrices for them. This next exercise uses words, but you can translate pictures into matrices using this principle also.

EXERCISE 1.22. In each of the following determine the matrix M that has the following effect on $\mathbb{R}^2$. **Find the columns by figuring out where $\vec{e_1}$ and $\vec{e_2}$ go.** You can verify your answer by multiplying your matrix by vectors of your choice and seeing if they go where they are supposed to.

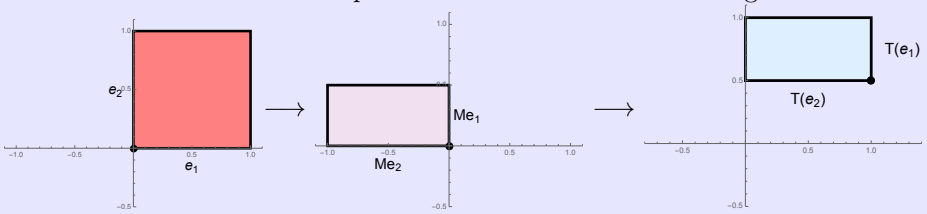
(1) M rotates vectors clockwise by $\pi/3$.
(2) M rotates vectors clockwise by $\pi/3$, then rescales the x direction by a factor of 2.
(3) M reflects across the line $y = -x$.
(4) M reflects across the line $y = -x$, then rescales the whole vector by a factor of $1/2$.

An *affine* transformation from $\mathbb{R}^2$ to itself is given by matrix multiplication followed by translation. It can be expressed in many ways, including these:

$$T(\vec{x}) = M\vec{x} + (e, f) = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} = (ax_1 + bx_2 + e, cx_1 + dx_2 + f) \tag{1.2}$$

Geometrically, affine transformations can be visualized by what they do to the unit square. This is best done in two stages: visualizing the matrix part of the transformation first, then translating.

EXAMPLE 1.23. Make an affine transformation out of the matrix $M = \begin{pmatrix} 0 & -1 \\ 1/2 & 0 \end{pmatrix}$ from example 1.20 by defining $T(\vec{x}) = M\vec{x} + (1, .5)$. Then the transformation of the unit square can be seen in these two stages:



EXERCISE 1.24. For each matrix in exercise 1.21, sketch the image of the unit square under the transformation $T(\vec{x}) = M\vec{x} + (1, .5)$.

EXERCISE 1.25.
(1) For each matrix from example 1.22 sketch the affine transformation $T(\vec{x}) = M\vec{x} + (0, .5)$. Decorate your unit square with an asymmetric image so that you can see how it transforms in each case.
(2) Which transformations are contractions?
(3) Which can be considered transformations on the unit square (as opposed to all of $\mathbb{R}^2$)?

1.5. Collage maps in two dimensions

In two dimensions we see the power and beauty of iterated function systems even better than in one dimension. In this section we will experiment with various choices of affine transformations and see how these choices affect the underlying fractal. By now you are probably beginning to ask two very important questions: (1) What exactly do we mean when we say "the limit under the collage map"? and (2) What exactly is a fractal? The answer to (1) is the subject of Chapter 2. The answer to (2) has not yet been agreed upon by the mathematical community. In section 1.6 we will address the key points of agreement on that question.

EXERCISE 1.26. Let $X = [0,1] \times [0,1]$ be the unit square in $\mathbb{R}^2$ and let $C(x,y) = (.5x, .5y)$. Define the following three transformations from X to itself:

$$T_1(x,y) = C(x,y) + (.5, 0) \tag{1.3}$$

$$T_2(x,y) = C(x,y) + (0, .5) \tag{1.4}$$

$$T_3(x,y) = C(x,y) + (.5, .5) \tag{1.5}$$

Given an initial set S_0, do the following four exercises on graph paper.

(1) Compute and sketch $T_1(S_0), T_2(S_0)$, and $T_3(S_0)$ in separate unit squares. Label your scale and important points.
(2) Let $\mathcal{T}$ be the collage map given by $T_1 \cup T_2 \cup T_3$. Define the set $S_1 = \mathcal{T}(S_0)$ and sketch it in a fresh unit square.
(3) Sketch $S_2 = \mathcal{T}(S_1)$ in a separate unit square from that of S_1.
(4) On your graph paper, make a really big unit square and sketch a nice picture of $S_3 = \mathcal{T}(S_2) = \mathcal{T}^3(S_0)$.

The choices for S_0 are as follows:

$(a) X$, $(b) [0,.5]\times[.5,1]$, $(c) [0,1]\times[0,.5]$, or $(d) \{(x,y) \in X$ such that $y \geq x\}$

In the examples from the previous section and in this example you may be noticing a pattern: it does not seem to matter what the initial set is. That is not a fluke, but rather a common trait that collage maps constructed from contractions share. The fractal associated with the IFS from exercise 1.26 is shown in figure **??**.

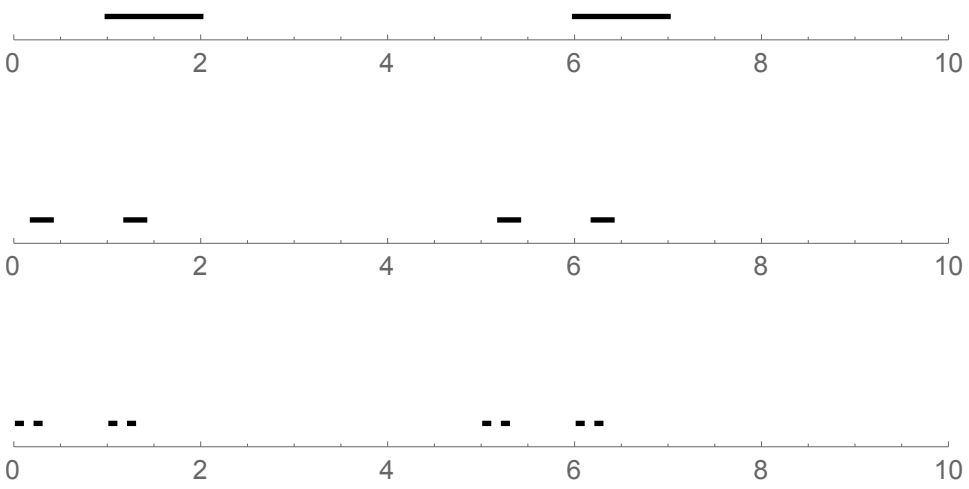


FIGURE 4. The collage map applied three times to A.

There are a few key things to notice about this fractal image, which we will call A. Each initial set S_0 was a subset of the unit square, as was $\mathcal{T}(S_0), \mathcal{T}^2(S_0)$, and all further images under the collage map. Naturally this makes A a subset of the unit square also. It is a very special subset in that it is a *fixed point of the collage*

map in the sense that $\mathcal{T}(A) = A$. Let's analyze the statement that $\mathcal{T}(A) = A$ geometrically with a detailed look at figure **??**.

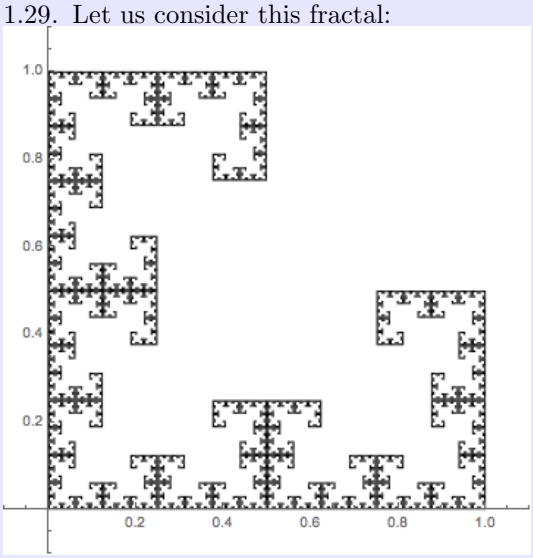
When T_1 is applied to A, it shrinks it by half and moves it to the right by $(.5, 0)$. You can see that A has a copy of itself in its lower-right corner, and that's the part of the collage map given by T_1. When T_2 is applied to A, it shrinks it by half and moves it up by $(0, .5)$. In that location you see that A has a copy of itself, which is the part of the collage map given by T_2. The map T_3 shrinks by half and moves diagonally by $(.5, .5)$; the third copy of A is there in the upper right. Thus when the collage map is applied to A, it makes three copies of itself whose union is A again. That is to say, $\mathcal{T}(A) = A$.

EXERCISE 1.27. Let X be the unit square in $\mathbb{R}^2$. Make up four affine transformations $T_i : X \to X, i = 1, ..., 4$ such that for their collage map $\mathcal{T}$ we have that $\mathcal{T}(X) = X$. Do you believe it is possible to do this in such a way that $T_i(X) \cap T_j(X) = \emptyset$ when $i \neq j$?

EXERCISE 1.28. Let $A = \{(1/2, 1/2)\}$ and consider the collage map you just made up for the previous question. Compute $\mathcal{T}^n(A)$ for some small values of n. Do you believe that it may be true (given proper definitions) that $\mathcal{T}^n(A) \to X$ as $n \to \infty$?

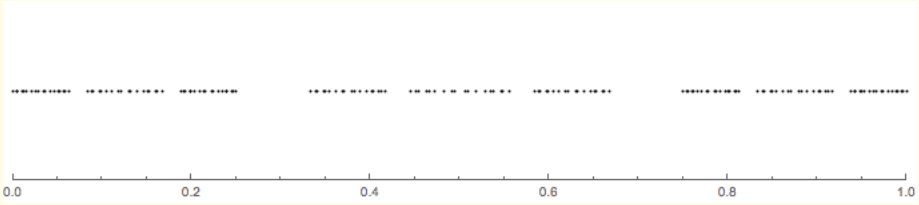
It is sometimes possible to look at a fractal such as the one in figure **??** and determine the collage it came from. To do this, one must parse the image into components that appear to be similar to the original. Then, for each piece of the image one must try to determine the affine transformation that takes the whole image into that piece.

EXAMPLE 1.29. Let us consider this fractal:



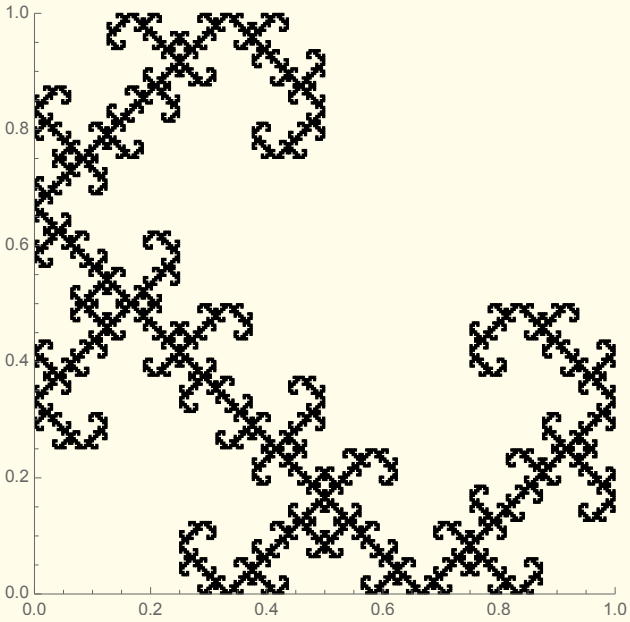
It is made up of three copies of itself, each half as big as the original. One copy contains the origin and is simply the image of the fractal under the map $T_1(\vec{x}) = \vec{x}/2$. The piece of the fractal between $x = .5$ and $x = 1$ is the entire image halved, rotated by $\pi/2$ counterclockwise, then moved to the right by $(1, 0)$. This corresponds to $T_2(\vec{x}) = \begin{pmatrix} 0 & -1/2 \\ 1/2 & 0 \end{pmatrix} \vec{x} + \vec{e_1}$. Finally the piece on the upper left, between $y = .5$ and $y = 1$, is the entire image halved, rotated by $\pi/2$ clockwise, and then moved up by $(0, 1)$. This corresponds to $T_3(\vec{x}) = \begin{pmatrix} 0 & 1/2 \\ -1/2 & 0 \end{pmatrix} \vec{x} + \vec{e_2}$. When we apply the collage map of these three transformations, we see that this fractal is fixed by it.

EXERCISE 1.30. Now it is your turn to try and decode the collage a fractal has come from, first in one dimension. Here we have a fractal A that is the fixed set of a collage map made from three affine transformations:



The fractal is contained in unit interval of the real line, but it is shown floating above the line so that you can see it. Find the three transformations $T_i : [0, 1] \to [0, 1]$ for which $\mathcal{T}(A) = A$. (You should be using the scale provided in the figure to help you determine precise contraction factors and translations).

EXERCISE 1.31. Now try to figure out the three affine transformations of the unit square that give this set A in the limit:



1.6. What is a fractal?

During the early study of fractals it became clear that a formal defintion was elusive, and indeed none of [**Bar08**, **Bar12**, **Fal06**, **PC09**] offer one. Instead, fractals are objects that are identified by two main properties: complicated geometric structure and self-similarity.

One way to describe a 'complicated geometric structure' is to describe what it is not: fractals are not simple objects like lines, circles, cones, or triangles. Rather, they are like clouds, trees, and coastlines: objects with obvious structure but no obvious way to describe or measure them. One way that this idea can be quantified is through dimension. In fact, originally Mandelbrot felt the definition of a fractal should depend on its fractal dimension.

The *box-counting dimension* Dim_B of a set is the foremost definition of a fractal dimension, and we will define it properly in chapter 4. For now, think of it as a number that captures how the mass of the set scales when it is expanded by a fixed amount. The box-counting dimension can be a number that is not an integer. By way of contrast, the *topological dimension* Dim_T of a set, which also has a technical definition we do not describe here, corresponds to our intuititve idea of dimension and can therefore only take integer values. In an early treatise on fractals [**Man77**], Mandelbrot brings us the following definition of what it means to be fractal:

> "The cases where $Dim_B = Dim_T$ include all of Euclid, and the cases where $Dim_B > Dim_T$ include every set I was ever tempted to call fractal... Hence, there is no harm in proposing the following definition:

> *A fractal will be defined as a set for which the Hausdorff-Besicovitch dimension*[2] *strictly exceeds the topological dimension.*"

In the fullness of time, it became clear that there were sets that failed this definition and yet still seemed to deserve to be called fractal [**Edg90**, p. 179]. However, sets with non-integer fractal dimension have the sort of complicated geometry that is often described as fractal.

The second main indicator of what it means for a set to be a fractal is *self-similarity*. The word 'similarity' can be construed in a number of ways, the strictest of which is the type taught in an elementary geometry course to describe things like similar triangles. In this definition two objects are similar if one is a rescaling of the other. A set is self-similar, then, if portions of the set are similar to the whole set. We see this version of self-similarity in all of the introductory examples of these notes.

However, it is useful to broaden the category of self-similarity to include sets for which portions are similar to some portion of the original, but perhaps not all of it. Moreover, we may wish to allow some flexibility in the word 'similar', perhaps allowing copies that are images under a contracting but nonlinear map. In chapters 5 and 6 we will learn about sets having this sort of self-similarity: Julia sets and the Mandelbrot set.

In [**Fal06**, p. xxv] an expansive list of properties is suggested, and we leave them as our final word on what it means to be fractal.

> "When we refer to a set F as a fractal, therefore, we will typically have the following in mind.
> (i) F has a fine structure, i.e. detail on arbitrarily small scales.
> (ii) F is too irregular to be described in traditional geometrical language, both locally and globally.
> (iii) Often F has some form of self-similarity, perhaps approximate or statistical.
> (iv) Usually, the 'fractal dimension' of F (defined in some way) is greater than its topological dimension.
> (v) In most cases of interest F is defined in a very simple way, perhaps recursively."

1.7. Exercises

EXERCISE 1.32. Let X be a subset of $\mathbb{R}, \mathbb{R}^n$, or $\mathbb{C}$, and let $T_1 : X \to X$ and $T_2 : X \to X$ be contraction mappings. Prove that $T_1 \circ T_2 : X \to X$ is also a contraction mapping.

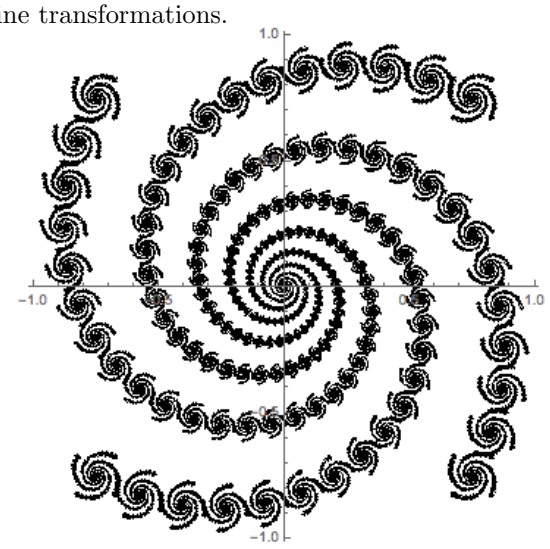
EXERCISE 1.33. Give an example of a contraction mapping $T : \mathbb{R} \to \mathbb{R}$ that has the property that $|T(x)| > |x|$ for some $x \in \mathbb{R}$. The ideal response will prove that T is a contraction, then exhibit a specific x that does not shrink under the transformation.

EXERCISE 1.34. Suppose that $T_1(\vec{x}) = (ax_1 + bx_2 + e, cx_1 + dx_2 + f)$ and $T_2 = (gx_1 + hx_2 + k, ix_1 + jx_2 + l)$ are affine transformations of $\mathbb{R}^2$. Prove that $T_1 \circ T_2$ is also an affine transformation of $\mathbb{R}^2$ and give its explicit formula.

[2] Pretend he said "box-counting dimension"

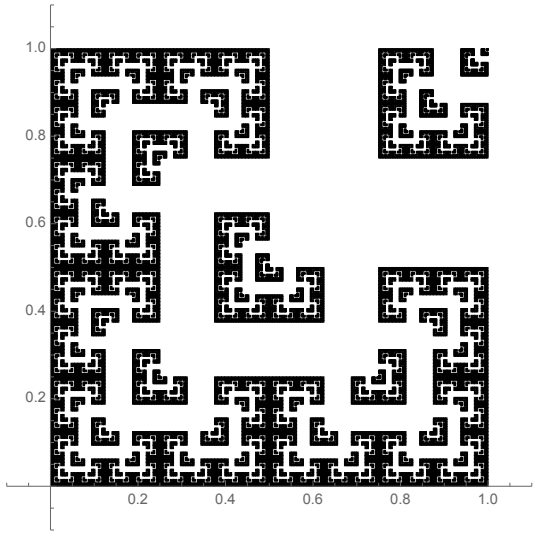
EXERCISE 1.35. (1) Find the formula of an affine transformation of $\mathbb{R}^2$ that takes the triangle with vertices at $(0,0),(1,0)$, and $(0,1)$ to the triangle with vertices at $(2,1),(3,3)$, and $(0,4)$.
(2) Find another affine transformation that accomplishes the same task.
(3) How many affine transformations are there in total that could do this?

EXERCISE 1.36. For this spiraling fractal do not attempt to find specific formulae for the affine transformations.



Instead, make a sketch of the domain $X = [-1,1] \times [-1,1]$ that shows your estimate of where each transformation takes X. (Hint: one of your transformations only contracts a little bit, and the rest contract a lot.)

EXERCISE 1.37. Find the affine transformations of the unit square whose collage map $\mathcal{T}$ gives this fractal:



CHAPTER 2

Hausdorff metric and the space of fractals

The goal of this chapter and the next is to rigorously prove conditions under which the fractal for a given collage map actually exists. That is, we will show when a collage map $\mathcal{T}$ admits a set A for which $\mathcal{T}(A) = A$. As an added bonus, we will prove that the set A is the limit of $\mathcal{T}^n(S_0)$ for any viable initial set S_0. That insight gives rise to an efficient algorithm that may be implemented on a computer to generate images of fractals.

In order to do all of this in a mathematically sound fashion it is necessary to build a foundation, which is the purpose of this chapter. The foundation we require is an understanding of "the space of fractals" $\mathcal{H}(X)$ as a metric space under the "Hausdorff metric" d_H. There are a number of technical details we will need to address in order to make this precise.

Until further notice we will use X to mean some subset of $\mathbb{R}^n$ or $\mathbb{C}$ along with its usual metric, which we will denote by $d(x, y)$ and call the *standard Euclidean metric*. To be precise, if $x = (x_1, x_2, ..., x_n)$ and $y = (y_1, y_2, ..., y_n)$ are elements of $\mathbb{R}^n$, then we will use the following definition and notations interchangeably:

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \cdots + (x_n - y_n)^2} = |x - y|.$$

If $z = a + ib$ and $w = c + id$ are elements of $\mathbb{C}$ the metric is basically the same and can be written as

$$d(z, w) = \sqrt{a - c)^2 + (b - d)^2} = \sqrt{(z - w)\overline{(z - w)}} = |z - w|.$$

2.1. A tiny bit of point-set topology

We cannot avoid learning a little bit of terminology that is commonly found in the realm of "analysis" (the branch of mathematics that calculus lives in). The main thing we need is the idea of a *compact set*, which we will give a simplified definition of here, and which you will learn/have learned about in your Real Analysis course. Compact sets in $\mathbb{R}^n$ or $\mathbb{C}$ are sets that are closed and bounded, so we must define those ideas now.

DEFINITION 2.1. We say that a set $A \subset X$ is *closed* if it contains its limit points. That is, if a sequence $\{a_n\} = \{a_1, a_2, a_3, ...\}$ in A has the property that $\lim_{n\to\infty} a_n = a$, then $a \in A$ also.[a]

[a]Of course, if you have not had Real Analysis you may not know the precise definition of the limit of a sequence. For our purposes it will suffice for you to use your intuitive idea of a limit. However, you may be curious about the official definition, so here it is.

DEFINITION 2.2. Let $\{x_n\}$ be a sequence in X. We say $\lim_{n\to\infty} x_n = L$ if and only if for every $\epsilon > 0$ there is an $N \in \mathbb{N}$ such that for every $n \geq N$, $d(x_n, L) < \epsilon$.

EXERCISE 2.3. Throughought your mathematical career your instructors have been referring to intervals like $[3, 7]$ as closed intervals. It turns out they have always meant it in the precise sense of Definition 2.1.

(1) Convince yourself that $[3, 7]$ satisfies the definition of being closed.
(2) Make an example of an interval that is not closed. Show, using the definition, exactly why it fails to be closed.

EXERCISE 2.4. Can a finite set fail to be closed?

EXERCISE 2.5. Give several examples of sets that are or are not closed, in $\mathbb{R}$ and $\mathbb{C}$. Try to make examples that differ in interesting ways.

The intuitive idea behind a set being bounded is that no portion of it heads off to infinity. In $\mathbb{R}^2$ or $\mathbb{C}$ that is equivalent to saying you can draw a big circle around it. In $\mathbb{R}$ it means that you can put upper and lower bounds on its elements. In higher dimensions it means you can enclose the set in a sufficiently large "n-ball" (interval, circle, sphere,...). For concreteness in the definition we choose to say that the n-ball is centered at the origin.

DEFINITION 2.6. We say that a set $A \subset X$ is *bounded* if there is some number $r > 0$ such that $|a| \leq r$ for all $a \in A$. That is to say, there is an $r > 0$ for which $d(a, 0) \leq r$ for all $a \in A$.

EXERCISE 2.7. Can a finite set fail to be bounded?

EXERCISE 2.8. Give several examples of sets that are or are not bounded, in $\mathbb{R}$ and $\mathbb{C}$. Try to make examples that differ in interesting ways.

What we really need to define the space of fractals is compactness. This is a profoundly useful property of sets that will be dealt with in detail in your real analysis course. What we are taking as a definition here is actually the celebrated Heine-Borel theorem, but that is something for you to tackle on another day.

DEFINITION 2.9. We say that a set $A \subset X$ is *compact* if it is closed and bounded.

There are a number of incredibly useful properties possessed by compact sets. One that is very important for fractal geometry is that it is possible to measure the distance between two compact sets unambiguously.

2.2. $\mathcal{H}(X)$, the space of fractals.

Let X be a compact subset of $\mathbb{R}^n$ (for some n) or $\mathbb{C}$, with the standard Euclidean metric d. The *space of fractals* is the set

$$\mathcal{H}(X) = \{A \subseteq X \text{ such that } A \text{ is compact}\}$$

An element of $\mathcal{H}(X)$ is therefore a compact subset of X. Take a moment to think about that carefully. Probably you will need several moments, because a "point" in the space of fractals $\mathcal{H}(X)$ is actually a "set" in the space X. Put another way, the space of fractals is the set of compact subsets of X.

EXERCISE 2.10. Let X be the unit interval $[0, 1]$. Give three specific examples of elements of $\mathcal{H}(X)$, complete with pictures. Try and make your examples as different from each other as possible.

EXERCISE 2.11. Give examples of subsets of $[0, 1]$ that are not elements of $\mathcal{H}(X)$. Draw pictures that explain why.

EXERCISE 2.12. Let X be the unit square in $\mathbb{R}^2$. Give three specific examples of elements of $\mathcal{H}(X)$, complete with pictures. Try and make your examples as different from each other as possible.

EXERCISE 2.13. Give examples of subsets of the unit square that are not elements of $\mathcal{H}(X)$. Draw pictures to help explain why.

Fractals are compact subsets of X, i.e. elements of $\mathcal{H}(X)$. They live in $\mathcal{H}(X)$ along with all the other compact subsets of X, but they satisfy special geometric properties that were discussed in the first chapter. We know that fractals can appear is as limit points of iterated function systems; next we need to develop the concept of metric spaces in order to make that idea precise.

2.3. Metric spaces

The only way to do geometry on a space is to first know how to measure the distance between points in that space. A property of an object in the space is considered "geometric" if it doesn't change when you move the object in a manner that preserves distances.

The domain of a metric on a space X is the set of all ordered pairs of elements of X. It takes the following notation and definition:

$$X \times X = \{(x, y) \text{ such that } x, y \in X\}.$$

DEFINITION 2.14. A metric on X is a function $d : X \times X \to \mathbb{R}$ satisfying the following conditions:

(1) $d(x, y) \geq 0$ for all $x, y \in X$,
(2) $d(x, y) = 0$ if and only if $x = y$,
(3) $d(x, y) = d(y, x)$ for all $x, y \in X$, and
(4) (triangle inequality) $d(x, y) \leq d(x, z) + d(z, y)$ for all $x, y, z \in X$.

Thus a metric is positive, symmetric, satisfies the triangle inequality, and distances between nonequal elements are never 0. An important detail to notice in this definition is the fact that the distance is not allowed to be ∞. That is because we have defined d with $d : X \times X \to \mathbb{R}$, and ∞ is not a real number.

One would hope that the standard Euclidean metric satisfies the official mathematical definition for a metric, and it does:

EXERCISE 2.15. Prove that when $X = \mathbb{R}$, the function $d(x, y) = |x - y|$ satisfies the definition of a metric. You may assume all standard properties of the absolute value function as you do your proof.

EXERCISE 2.16. For the standard Euclidean metrics in $\mathbb{R}$ and $\mathbb{R}^2$, determine conditions on x, y and z for which the triangle inequality is an equality.

EXERCISE 2.17. Consider the standard Euclidean metric in $\mathbb{R}^2$.

(1) Sketch a picture of $d(x, y)$ and explain it in terms of the Pythagorean Theorem.
(2) Sketch another picture in $\mathbb{R}^2$ that clearly shows why the triangle inequality is so named.

There are numerous reasons why it is useful in real life to have different definitions of a metric on the same space. For example, imagine a ruler that measures distance in the English system (inches) on one side and in the metric system (centimeters) on the other. If I ask you the distance between two points, your answer is going to depend on which side of the ruler you are using, although it would be natural for you to provide the units with your answer.

EXERCISE 2.18. Consider $d : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ given by the formula

$$d(x, y) = \begin{cases} 1 & x \neq y \\ 0 & x = y \end{cases}$$

Is d a metric on $\mathbb{R}^2$?

EXERCISE 2.19. For this exercise let $X = \mathbb{R}$.

(1) Make up an example of a function $d : X \times X \to \mathbb{R}$ such that condition (1) of a metric fails.
(2) Make up an example of a function $d : X \times X \to \mathbb{R}$ such that condition (1) of a metric holds but condition (2) fails.
(3) Make up an example of a function $d : X \times X \to \mathbb{R}$ such that conditions (1) and (2) of a metric hold but condition (3) fails.
(4) Make up an example of a function $d : X \times X \to \mathbb{R}$ such that conditions (1), (2), and (3) hold but (4) fails.

When you change the metric on a space you change its geometry in a fundamental way. Experiment with this idea using the metric(s) that you came up with here by considering *balls of radius ϵ centered at $x \in X$*, defined by and with the notation:

$$B(x, \epsilon) = \{y \in X \text{ such that } d(x, y) \leq \epsilon\}$$

The choice of metric determines the shape of the ball by determining which points fall into the ball and which do not. Put another way, the choice of the metric determines the geometry of the balls, and by extension the geometry of X. In the next few exercises you have the opportunity to see what happens when different metrics are applied.

EXERCISE 2.20.
(1) For the standard Euclidean metric in $\mathbb{R}$, compute $B(3, 1/2)$ and $B(2, -1)$.
(2) Let $d_\alpha(x, y) = |\alpha x - \alpha y|$ for some $\alpha > 0$. Compute $B(3, 1/2)$ in this metric.

EXERCISE 2.21. Try to make up an example of a metric on $\mathbb{R}^2$ that isn't the standard Euclidean metric by altering the definition of the standard Euclidean metric just a little bit. If you are successful, try to compute $B(x, \epsilon)$ for some choice of x and ϵ and see what shape the ball is.

2.4. Hausdorff metric on $\mathcal{H}(X)$

The Hausdorff metric is a way to measure the distance $d_H(A, B)$ between two compact sets $A, B \in \mathcal{H}(X)$. We're going to look at the definition two different ways, both of which depend on the standard Euclidean metric and which are equivalent. The first way depends on the idea of *ϵ-thickening* of sets. The second way depends on maximizing the minimum distance between elements of A and B. In both cases care needs to be taken to ensure the metric is symmetric, i.e. that $d_H(A, B) = d_H(B, A)$, and in both cases this will be done by taking a maximum.

2.4.1. Hausdorff metric definition using ϵ-thickening. .

Let A be a compact subset of X and let $\epsilon \geq 0$. The *ϵ-thickening* of A is the set

$$A_\epsilon = \{x \in X \text{ such that } d(x, a) \leq \epsilon \text{ for some } a \in A\}.$$

That is, A_ϵ is the subset of X that contains all the points in X that are within some element of A. Alternatively, one could imagine taking the union of all of the ϵ-balls $B(a, \epsilon)$, using each element a of A as a center.

EXAMPLE 2.22. Suppose $A \subset \mathbb{R}$ is given by $A = \{1, 2\}$ and let $\epsilon = 1/3$. Then $A_\epsilon = [2/3, 4/3] \cup [5/3, 7/3]$. This is because every point in $[2/3, 4/3]$ is within ϵ of $1 \in A$ and every point in $[5/3, 7/3]$ is within ϵ of $2 \in A$.

EXERCISE 2.23. Let $A = [1, 2] \subset \mathbb{R}$ and let $\epsilon = 1/4$. Find A_ϵ.

EXERCISE 2.24. Let $\mathcal{C} \subset [0, 1]$ be the middle-thirds Cantor set and let $\epsilon = 1/9$. Find $\mathcal{C}_\epsilon$. Repeat with $\epsilon = 1/27$.

EXERCISE 2.25. Let $A \subset \mathbb{R}^2$ be the unit circle and let $\epsilon = 1/2$. Give a precise description of A_ϵ. Repeat for $\epsilon = 2$.

EXERCISE 2.26. Let A be the line segment in $\mathbb{R}^2$ connecting the origin to $(1, 1)$ and let $\epsilon = .1$. Make a fairly precise sketch of the set A_ϵ.

EXERCISE 2.27. Let $A \in \mathcal{H}(X)$ and let $\epsilon = 0$. What is A_ϵ?

In order to find the Hausdorff distance between two compact sets $A, B \subset X$ you will need to be able to find the smallest ϵ for which $B \subset A_\epsilon$. That is to say, you will need to be able to find the minimum amount of thickening A needs in order to cover all of B. Let's look at a few concrete examples first and then define what we mean by this precisely.

EXAMPLE 2.28. Let $A = \{1, 2\} \subset \mathbb{R}$ and let $B = [.75, 1.25]$. The smallest ϵ for which $B \subset A_\epsilon$ is .25. That's because $A_{.25} = [.75, 1.25] \cup [1.75, 2.25]$ and if ϵ is any smaller than .25, the interval from A_ϵ that intersects B is too small to contain it.

EXERCISE 2.29. For the A and B in the previous example, what is the smallest ϵ for which $A \subset B_\epsilon$?

Let's be really precise about our usage of the word "smallest" by defining what it means to be the minimum value in a set of real numbers.

DEFINITION 2.30. Let $A \subset \mathbb{R}$. The *minimum element of* A, if it exists, is denoted $\min A = \min\{x \text{ such that } x \in A\}$ and is defined to be the element $a \in A$ for which $a \leq x$ for all $x \in A$.

It is possible for the minimum element of a set not to exist, for instance the interval $(0, 1)$ has no minimum element.[1] In the space of fractals we do not have to worry about this problem because compact sets of real numbers always have a minimum element.

Exercises 2.28 and 2.29 show that one needs to consider two minimum epsilons: the one for which $B \subset A_\epsilon$ and the one for which $A \subset B_\epsilon$. If we do not consider both we run the risk of the metric we define using ϵ-thickenings to fail to be symmetric.

DEFINITION 2.31. Let $A, B \in \mathcal{H}(X)$. The *Hausdorff distance* between A and B is given by

$$d_H(A, B) = \min\{\epsilon \text{ such that } A \subset B_\epsilon \text{ and } B \subset A_\epsilon\} \tag{2.1}$$

It is possible to rewrite this definition as the maximum of two minimums:

$$d_H(A, B) = \max\{\min\{\epsilon \text{ such that } A \subset B_\epsilon\}, \min\{\epsilon \text{ such that } B \subset A_\epsilon\}\} \tag{2.2}$$

Although that may look more complicated, it may the more useful because you will calculate each ϵ separately and then just take the larger of the two. To see that it is equivalent, consider the ϵ defining the minimum in equation 2.1. That ϵ is greater than or equal to each of the minimums of equation 2.2 and thus is greater than or equal to their maximum. On the other hand, the maximum of the two epsilons from 2.2 is certain to be an epsilon for which both $A \subset B_\epsilon$ and $B \subset A_\epsilon$, so it is greater than or equal to that from 2.1. When two numbers are greater than or equal to each other they must be equal.

EXERCISE 2.32. Let $A = \{1, 2\} \subset \mathbb{R}$ and let $B = [.75, 1.25]$. Find $d_H(A, B)$.

EXERCISE 2.33. In this exercise we consider distances between compact subsets A and B of $\mathbb{R}^2$. Please make sketches to illustrate your answers.

(1) Let A be the unit square $[0, 1]\times[0, 1]$and let $B = \{(x, y)$ such that $x^2 + y^2 = 1\}$ (the unit circle). Find $d_H(A, B)$.
(2) Let A be the unit square and let B be the disk of radius $1/2$ centered at $(1/2, 1/2)$. Find $d_H(A, B)$.
(3) Let A be the unit square and let B be the line segment connecting $(0, 1)$ to $(1, 0)$. Find $d_H(A, B)$.
(4) Let A be the unit square and let B be the line segment connecting $(-1, 2)$ to $(0, 2)$. Find $d_H(A, B)$.
(5) Let A be the line segment from the origin to $(1, 0)$ and let B be the line segment from the origin to $(0, 1)$. Find $d_H(A, B)$.

[1] There is a related mathematical notion called the *infimum* of a set, which is the largest number that is not greater than any element of the set. The infimum of $(0, 1)$ is 0.

EXERCISE 2.34. Let $A = [0, 1/2]$ and consider the Cantor set collage map $\mathcal{T} = T_1 \cup T_2$ on $\mathcal{H}([0, 1])$ given by $T_1(x) = 1/3x$ and $T_2(x) = 1/3x + 2/3$. Compute $d_H(A, \mathcal{T}(A))$ and $d_H(\mathcal{T}(A), \mathcal{T}^2(A))$.

EXERCISE 2.35. In exercise 1.26 you were assigned a set S_0 and asked to apply the collage map $\mathcal{T}$ to it two times. For that collage map and your particular S_0, compute $d_H(S_0, \mathcal{T}(S_0))$ and $d_H(\mathcal{T}(S_0), \mathcal{T}^2(S_0))$.

2.4.2. Hausdorff distance via the maximum of the minimum distances. We must build up to Hausdorff metric in stages in this method also. We begin by measuring the distance from a point $x \in X$ to a a set $A \in \mathcal{H}(X)$.

Consider the Euclidean distance $d(x, a)$ for each element of A; the minimum is defined to be *the distance from x to A* and we write

$$l(x, A) = \min\{d(x, a) \text{ such that } a \in A\}$$

As before, we don't have to worry about whether this minimum exists since A is a compact set.

EXERCISE 2.36. Let A be the unit circle in $\mathbb{R}^2$ and let $x = (1, 1)$. Find $l(x, A)$.

EXERCISE 2.37. Let $\mathcal{H}(X)$ be any space of fractals and let $A \in \mathcal{H}(X)$. If $x \in A$, what is $l(x, A)$?

A particularly nice consequence of the compactness of A is that not only is this minimum distance guaranteed to exist, it must be realized as the distance between x and at least one specific point in A.

LEMMA 2.38. *For any $A \in \mathcal{H}(X)$ and $x \in X$ there is an element $\hat{y} \in A$ for which $l(x, A) = d(x, \hat{y})$.*

For the proof of this lemma see [**Bar12**, p. 29].

EXERCISE 2.39. In each of the previous two exercises find $\hat{y}$.

Next we define the distance from one set $A \in \mathcal{H}(X)$ to another set $B \in \mathcal{H}(X)$ by considering all distances $l(a, B)$ over all $a \in A$. This maximum is, like the minimum, guaranteed to exist because of compactness. We define *the distance from A to B* as

$$l(A, B) = \max\{l(a, B) \text{ such that } a \in A\}$$

EXERCISE 2.40. Let $A = [0, 2]$ and let $B = [1, 1.5]$. Find $l(A, B)$.

There is a similar lemma saying that the maximum distance is attained by elements of A and B which we state here. The existence is a consequence of compactness.

LEMMA 2.41. *For any $A, B \in \mathcal{H}(X)$ there exists $\hat{x} \in A$ and $\hat{y} \in B$ such that $l(A, B) = d(\hat{x}, \hat{y})$.*

EXERCISE 2.42. Let $A = [0, 2]$ and let $B = [1, 1.5]$. Find $\hat{x}$ and $\hat{y}$ that satisfy the lemma. Are they unique?

So the function l seems like progress toward defining a metric on $\mathcal{H}(X)$. We immediately see that it is nonnegative. Consider the following exercise and then decide about conditions (2) and (3) of a metric.

EXERCISE 2.43. Let $A = [0, 2]$ and let $B = [1, 1.5]$. Find $l(B, A)$.

So we don't quite have a metric yet. But it turns out that you can fix both of the issues this example presented in a very simple way by defining the *Hausdorff metric* to be

$$d_H(A, B) = \max\{l(A, B), l(B, A)\} \tag{2.3}$$

It is a technical exercise to prove that this version of the definition is equivalent to the one given in terms of ϵ-thickenings. A viable strategy for the proof is to consider ϵ to be the value given by the first definition and ϵ' the value given by the second. Then you would prove that $\epsilon \leq \epsilon'$ and $\epsilon' \leq \epsilon$. This shows they are equal.

EXERCISE 2.44. A corollary to Lemma 2.41 is that there is an $\hat{x} \in A$ and a $\hat{y} \in B$ such that $d_H(A, B) = d(\hat{x}, \hat{y})$. Prove this corollary.

EXERCISE 2.45. Let $A = [0, 2]$ and let $B = [1, 1.5]$. Find $d_H(A, B)$ and find the $\hat{x}$ and $\hat{y}$ that represent this distance.

EXERCISE 2.46. Let A be the disk of radius 2 centered at $(3, 0)$ and let B be the rectangle with corners at $(2, -2), (3, -2), (3, 4)$, and $(2, 4)$.

(1) Make a sketch that uses Lemma 2.41 to show $l(A, B)$. Be sure to label $\hat{x}$ and $\hat{y}$.
(2) Repeat the previous part for $l(B, A)$.
(3) Find $d_H(A, B)$.

EXERCISE 2.47. Let X be the unit square and let $\mathcal{T}$ be the collage map of the transformations $T_i, i = 1, .., 4$ defined as follows.

$$T_1(x) = x/2 \qquad T_2(x) = x/2 + (1/2, 0)$$
$$T_3(x) = x/2 + (0, 1/2) \qquad T_4(x) = x/2 + (1/2, 1/2)$$

Let $A \in \mathcal{H}(X)$ be given by $A = \{(0, 0)\}$.

(1) Find $d_H(A, X), d_H(\mathcal{T}(A), X)$, and $d_H(\mathcal{T}^2(A), X)$.
(2) Find a formula for $d_H(\mathcal{T}^n(A), X)$.
(3) We can consider the sequence $\{\mathcal{T}^n(A)\}_{n=0}^{\infty}$ of elements of $\mathcal{H}(X)$. Discuss the evidence for the existence of the limit $\lim_{n\to\infty} \mathcal{T}^n(A)$.

2.5. Exercises

EXERCISE 2.48. Consider the subset of $\mathbb{R}^2$ given by $A = \{(x, \sin(\pi/x)) \text{ such that } x \in (0, 1)\}$.

(1) Make a fairly accurate sketch of this subset of $\mathbb{R}^2$.
(2) Is this set bounded?
(3) A is not a closed set. Find the limit points of A that are not in A.
(4) For two of the limit points of A that you found in the previous part, exhibit a sequence of elements of A that converge to it.

EXERCISE 2.49. The "taxicab metric" is a natural metric to use in $\mathbb{R}^2$ and is defined by:

$$d_t(x, y) = |x_1 - y_1| + |x_2 - y_2|$$

(1) Prove that d_t satisfies the conditions to be a metric.
(2) Explain in words, perhaps using a drawing to explain your thinking, why the word "taxicab" has been chosen to describe the metric.

EXERCISE 2.50.
(1) For the metric $d_s(x, y) = |x^3 - y^3|$ on $\mathbb{R}$, compute $B(0, 1/8)$ and $B(3, 1/8)$. Compare and contrast to each other and to the balls you would get using the standard Euclidean metric.
(2) For the taxicab metric in $\mathbb{R}^2$, calculate $B((0, 0), 1)$. Compare and contrast to what you get for $B((0, 0), 1)$ using Euclidean metric.

EXERCISE 2.51. Prove that if $\epsilon \geq 0$, then $B \subseteq B_\epsilon$.

EXERCISE 2.52. Prove, using whichever definition of d_H you like, properties (1), (2), and (3) of a metric.

EXERCISE 2.53. Prove that if $B \subset A$, then $l(B, A) = 0$.

EXERCISE 2.54. Let $X = \mathbb{R}^n$ and consider A and B to be compact subsets of X. Prove the following two facts:

(1) $A \cup B$ is a compact subset of X.
(2) $A \cap B$ is a compact subset of X.

CHAPTER 3

Iterated Function Systems

Throughout this chapter, X will denote a compact subset of $\mathbb{R}^n$ or $\mathbb{C}$ and $\mathcal{H}(X)$ will be its space of fractals, endowed with the Hausdorff metric. The reason for this stipulation is so that the space of fractals $\mathcal{H}(X)$ is itself a compact set under the Hausdorff metric, which is a fact that we present without proof. The most important thing this implies for our purposes is the following.

LEMMA 3.1. *If $\{A_n\}$ is a sequence in $\mathcal{H}(X)$ such that $\lim_{n\to\infty} A_n = A$, then $A \in \mathcal{H}(X)$. In particular, a convergent sequence of nonempty compact subsets of X converges to a nonempty compact subset of X.*

3.1. Collage maps as contractions on the space of fractals

For concreteness let us write down the definition of a contraction mapping in the setting of the space of fractals. In this definition consider $\mathcal{T}$ to be any transformation on $\mathcal{H}(X)$, most often but not always a collage map.

DEFINITION 3.2. We say $\mathcal{T} : \mathcal{H}(X) \to \mathcal{H}(X)$ is a *contraction mapping* if there is a $c \in [0, 1)$ such that for all $A, B \in \mathcal{H}(X)$,

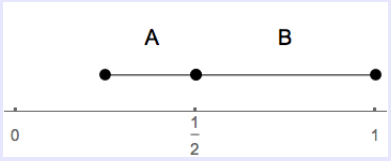
$$d_H(\mathcal{T}(A), \mathcal{T}(B)) \leq c\, d_H(A, B)$$

An elementary contraction mapping on $\mathcal{H}(X)$ is one obtained by a contraction on X. To be precise, let $T : X \to X$ be a contraction mapping with contraction factor $c \in [0, 1)$. Then by definition 1.10, T defines a transformation on $\mathcal{H}(X)$ and we will soon prove that it is a contraction on $\mathcal{H}(X)$ with contraction factor c.

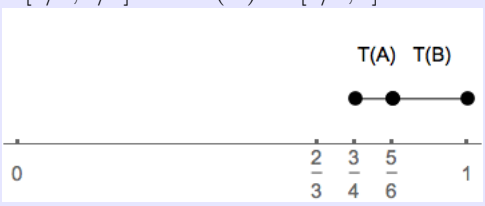
First we should note that whether or not $T : X \to X$ is a contraction, as long as it is continuous it extends to a transformation on $\mathcal{H}(X)$. To see this, consider $T(A)$ for some element $A \in \mathcal{H}(X)$. It is nonempty since A is. The fact that it is compact is a real analysis fact: we all learn the mantra "The continuous image of a compact set is compact". If T is a contraction on X, then it is automatically continuous and thus extends to a transformation of the space of fractals.

EXAMPLE 3.3. Let $X = [0,1]$, $A = [1/4, 1/2]$ and $B = [1/2, 1]$. Suppose that $T : X \to X$ is given by $T(x) = x/3 + 2/3$. We will prove that $d_H(T(A), T(B)) \leq 1/3\, d_H(A, B)$.

First let us compute $d_H(A, B)$. With this handy figure:



we can see that the furthest point in A from B is $1/4$, while the furthest point in B from A is 1, making the largest distance $1/2$ and so $d_H(A, B) = 1/2$. Next, we can compute $T(A) = [3/4, 5/6]$ and $T(B) = [5/6, 1]$. This next handy figure:



lets us compute $d_H(T(A), T(B)$ by noticing that the furthest point in $T(A)$ from $T(B)$ is $3/4$ while the furthest point in $T(B)$ from $T(A)$ is 1, making the largest distance $1/6$. Thus we see that $d_H(T(A), T(B)) = 1/6 = 1/3\, d_H(A, B)$.

Based on this evidence it would seem that the contracting factor for $\mathcal{H}(X)$ seems like it is going to equal the contracting factor on X itself, and that turns out to be true. The previous example is a special case of what happens for a general contraction on X.

LEMMA 3.4. *Suppose T is a contraction on X with contraction factor $c \in [0, 1)$. Then the transformation induced by T on $\mathcal{H}(X)$ is also a contraction mapping with contraction factor c.*

EXERCISE 3.5. The proof of Lemma 3.4 uses the following outline. Once you have figured out how all the steps work, write them into a formal proof.

(1) Consider any $A, B \in \mathcal{H}(X)$ and suppose $d_H(A, B) = \delta$. Explain why for every $a \in A$ there is a $b \in B$ for which $d(a, b) \leq \delta$.
(2) Compute an upper bound on the Euclidean distance between $T(a)$ and $T(b)$.
(3) Come up with an argument for why the previous step gives you a bound on the size of the ϵ-thickening of $T(B)$ required so that $T(A) \subseteq T(B)_\epsilon$. (Spoiler alert: that ϵ should be $c\delta$.)
(4) Make the symmetric argument that computes an epsilon for which $T(B) \subseteq T(A)_\epsilon$.
(5) Put the previous steps together to show that $d_H(T(A), T(B)) \leq c\delta = c\, d_H(A, B)$.

Of course when we are making fractals what we really care about is collages made from contractions. We would like to obtain a lemma similar to the previous lemma that tells us that such a collage is also a contraction, and what the contraction factor is. The following lemmas will help us get started.

LEMMA 3.6. *Let A, B, C, and D be elements of $\mathcal{H}(X)$. Then $d_H(A \cup B, C \cup D) \leq \max\{d_H(A, C), d_H(B, D)\}$.*

PROOF. Let $d_H(A, C) = \delta$ and $d_H(B, D) = \gamma$, and let $\nu = \max\{\delta, \gamma\}$. By definition of $d_H(A, C)$ we know that $A \subseteq C_\delta \subseteq C_\nu$; we also know that $C_\nu \subseteq C_\nu \cup D_\nu$. By Exercise 3.22 we know that $C_\nu \cup D_\nu = (C \cup D)_\nu$, and thus we have that $A \subseteq (C \cup D)_\nu$. Similarly, we see that $B \subseteq D_\gamma \subseteq D_\nu \subseteq C_\nu \cup D_\nu = (C \cup D)_\nu$. Putting these facts together we have that $A \cup B \subseteq (C \cup D)_\nu$.

In the opposite direction we know that $C \subseteq A_\delta \subseteq A_\nu \subseteq (A \cup B)_\nu$, and that $D \subseteq B_\gamma \subseteq B_\nu \subseteq (A \cup B)_\nu$. This shows us that $\nu \geq d_H(A \cup B, C \cup D)$ by the definition of Hausdorff metric using ϵ-thickening. □

This result extends to any finite unions of elements of $\mathcal{H}(X)$ as the next corollary shows. It can be proved by induction on the number of elements in the union.

COROLLARY 3.7. *Let $A_i, B_i \in \mathcal{H}(X)$ for $i = 1, 2, ..., n$ and let $\nu = \max_{1 \leq i \leq n} \{d_H(A_i, B_i)\}$. Then*

$$d_H\left(\bigcup_{i=1}^{n} A_i\,,\, \bigcup_{i=1}^{n} B_i\right) \leq \nu.$$

EXERCISE 3.8. Let $T_1 : X \to X$ be a contraction with contraction factor c_1 and let $T_2 : X \to X$ be a contraction with contraction factor c_2. In this exercise, fill in the blanks to complete a formal proof that the collage map $\mathcal{T} = T_1 \cup T_2$ is a contraction on $\mathcal{H}(X)$.

PROOF. Let $A, B \in \mathcal{H}(X)$. Then $\mathcal{T}(A) =$ _____ and $\mathcal{T}(B) =$ _____ by _____ . For $i = 1, 2$ we know that $d_H(T_i(A), T_i(B)) \leq$ _____ by Lemma _____. Applying Lemma _____ we know that

$$d_H(\mathcal{T}(A), \mathcal{T}(B)) \leq \max\{\text{_____}\}$$
$$\leq \max\{\text{_____}\}$$

This means that $d_H(\mathcal{T}(A), \mathcal{T}(B)) \leq c\, d_H(A, B)$, where c is defined to be _____. Since $c \in [0, 1)$ we have completed the proof that $\mathcal{T}$ is a contraction mapping on $\mathcal{H}(X)$. □

EXERCISE 3.9. Consider the transformations in Exercise 1.16, Exercise 1.17, Exercise 1.30, and Exercise 1.37. Using analysis similar to the proof in Exercise 3.8, determine the contraction factor for each collage map. Write the proof that your collage is a contraction following the form in Exercise 3.8; there are small details that will need to change (for instance, "Lemma" might become "Corollary" in a key place, depending on your number of maps).

Your work in the previous two exercises would replicate nearly exactly to prove the general case, stated next. Note that while in most of our examples the contraction maps have been affine, that is not needed. All that is required is that the transformations be contractions.

THEOREM 3.10. *Let $T_i : \mathcal{H}(X) \to \mathcal{H}(X)$ be a contraction mapping with contraction factor $c_i \in [0, 1)$ for all $i = 1, 2, ..., n$. Then the collage map $\mathcal{T} = \bigcup_{i=1}^{n} T_i$ is also a contraction mapping on $\mathcal{H}(X)$, with contraction factor $c = \max\{c_i \text{ such that } i = 1, 2, ..., n\}$.*

DEFINITION 3.11. Let (X, d) be a compact metric space. A *(hyperbolic) iterated function system (IFS)* is given by $\{X; T_1, T_2, ..., T_K\}$, where each T_j is a contraction from X to itself with contractivity factor c_j. The *contractivity factor* of the IFS is given by $c = \max\{c_1, c_2, ..., c_K\}$.

3.2. Existence of Attractors for Iterated Function Systems

We are ready to prove the existence theorem for fractals generated by contraction maps on $\mathcal{H}(X)$, which includes those given by collages of contractions (i.e., iterated function systems). To prove the theorem we will rely on a lemma from real analysis that we state without proof. It says that if $\{B_i\}_{i=1}^{\infty}$ is a nested sequence of nonempty compact subsets of X, then the limit of the sequence is a nonempty compact subset of X also.

LEMMA 3.12. *Let $B_1, B_2, ...$ be a sequence of elements of $\mathcal{H}(X)$ such that $B_i \subseteq B_{i-1}$ for all $i = 1, 2, ...$ Then $\lim_{i\to\infty} B_i$ is an element of $\mathcal{H}(X)$.*

Since $B_1 \supseteq B_2 \supseteq B_3...$ it is clear that the sets in the sequence are getting smaller; the key fact to note is that this lemma says that they cannot disappear entirely (although they certainly could converge to a set containing only a point!).

EXERCISE 3.13. Let $X = [-10, 10]$. Make up an example of a nested sequence of elements of $\mathcal{H}(X)$, and find its limit.

And now for the big theorem. It is a special case of the *contraction mapping theorem*, which in a broad range of cases proves that a contraction map has a unique fixed point to which the orbit of every point is attracted.

THEOREM 3.14 (Existence theorem for fractals). *Let X be a compact subset of $\mathbb{R}^n$ or $\mathbb{C}$ and let $\mathcal{H}(X)$ be its space of fractals endowed with the Hausdorff metric. Let $\mathcal{T} : \mathcal{H}(X) \to \mathcal{H}(X)$ be the collage of contractions $T_1, T_2, ..., T_K$ on X. Then*

1. *There exists a unique $A \in \mathcal{H}(X)$ for which $\mathcal{T}(A) = A$, and*
2. *For any $B \in \mathcal{H}(X)$,*

$$\lim_{n\to\infty} \mathcal{T}^n(B) = A \tag{3.1}$$

Before we launch into the proof, notice a few things. First, that part (1) promises us that a fixed point A exists, and that it is unique. The notation A is used because it is the "attractor" in the sense of (2): that the orbit of every $B \in \mathcal{H}(X)$ converges to it.

PROOF. Consider the sequence of elements of $\mathcal{H}(X)$ given by $\{\mathcal{T}^n(X)\}_{n=1}^{\infty}$. It is a nested sequence since $\mathcal{T}(X) \subseteq X$, and hence $\mathcal{T}^2(X) \subseteq \mathcal{T}(X)$, implying that $\mathcal{T}^3(X) \subseteq \mathcal{T}^2(X)$, and so on. By Lemma 3.12 we know that $\lim_{n\to\infty} \mathcal{T}^n(X)$ is an element of $\mathcal{H}(X)$ which we will call A. We must show that A is a fixed point of T, that it is unique, and that all orbits are attracted to it.

Since $\mathcal{T}$ is a contraction mapping, it is a continuous function and so

$$\mathcal{T}(A) = \mathcal{T}(\lim_{n\to\infty} \mathcal{T}^n(X)) = \lim_{n\to\infty} \mathcal{T}(\mathcal{T}^n(X)) = \lim_{n\to\infty} \mathcal{T}^{n+1}(X) = A,$$

and thus A is a fixed point of $\mathcal{T}$. The proof that A is unique is the next exercise. This will prove A is the only fixed point of $\mathcal{T}$ and thus completes the proof of part (1).

EXERCISE 3.15. Consider a set $C \in \mathcal{H}(X)$ for which $\mathcal{T}(C) = C$ and let $c \in [0, 1)$ represent a contraction factor for $\mathcal{T}$, which exists by Theorem _____. Then $d_H(\mathcal{T}(A), \mathcal{T}(C)) \leq$ _____. But A and C are both _____ by $\mathcal{T}$ and so that inequality becomes _____ $\leq c\, d_H(A, C)$. This is only possible if $d_H(A, C) =$ ___ . Since d_H is a _____ on $\mathcal{H}(X)$, we know that this is only possible if _____, proving that A is the unique fixed point.

The proof of equation (3.1) is the next exercise.

EXERCISE 3.16. Let $B \in \mathcal{H}(X)$ and let c be a contraction factor of $\mathcal{T}$. Prove equation (3.1) by proving that $\lim_{n\to\infty} d_H(\mathcal{T}^n(B), A) = 0$. As part of the proof, use induction to establish that $d_H(\mathcal{T}^n(B), A) \leq c^n\, d_H(B, A)$.

□

To summarize our findings, we now know the following.

(1) If $\{X; T_1, T_2, ..., T_K\}$ is an iterated function system with contraction factor $c = \max\{c_1, c_2, ..., c_K\}$, then it has a unique fixed point $A \in \mathcal{H}(X)$.
(2) That fixed point looks like $\mathcal{T}^n(X)$ for a large enough value of n, and so
(3) In chapter 1, the images for C_5, K_5, and S_5, which were the 5th iterations of X, look pretty much like the fractal does.

3.3. Two computer algorithms for IFS fractals

3.3.1. Deterministic algorithm. The theory is pretty clear: start with any initial compact set B and plot $\mathcal{T}^n(B)$ for as large of an n as your computer can handle. Here is some "pseudocode" for how you might actually do this; details will depend on the software you are using.

- Input the functions $T_1, T_2, ..., T_K$ into the computer.
- Input some initial set B_0, which could be as simple as a single point.
- Decide how many iterations you want to do and call it something like *numits*.
- Compute $\mathcal{T}^{numits}(B_0)$.
- Display $\mathcal{T}^{numits}(B_0)$.

For this course we will carry out the algorithm in Mathematica, and you will get some sample code for that soon.

3.3.2. Probabilistic Algorithm. The probabilistic algorithm for making fractals on the computer is quicker and more efficient than the deterministic one. It capitalizes on the fact that any set converges to A by applying the individual maps at random instead of using the entire collage map. We claim (but do not prove)

that the orbit of any point under a random application of the T_j's will trace out the entire fractal A. This may be believable since the collage map applied to any one-element set converges to A. Let us present the algorithm and then revisit this issue and the issue of relative efficiency.

Suppose we have the iterated function system $\{X; T_1, T_2, ..., T_K\}$. The algorithm begins with any initial point $x_0 \in X$ and computes its orbit a little differently than in Chapter 1, since $\mathcal{T}$ is not a map on X and we need to account for using all the T_js. For each $n \in 1, 2, 3, ...$ we select an index $j_n \in \{1, 2, ..., K\}$ randomly and apply that transformation to compute $x_n = T_{j_n}(x_{n-1})$. We compute and plot the entire orbit of x_0 up to $numits2$ iterations, where $numits2$ will be an extremely large number.

In advance of writing code there are two decisions to be made. It is necessary to decide on probabilities $p_1, p_2, ..., p_K$ to use to select the transformations; of course we need $\sum_{j=1}^{K} p_j = 1$. Higher probabilities mean more visits to the region that transformation covers, so the probabilities do affect the final images. That can be used as an advantage, for instance by giving smaller regions lower probabilities.

The other decision is what initial point x_0 to use. If the initial point is chosen from inside of A (for instance if it is the fixed point of a T_j), then the entire orbit of x_0 will be in A as you will prove in exercise 3.29. However it is not necessary to choose x_0 carefully since after a few iterations it must be close to A anyway, as you will prove in exercise 3.30.

Here is some pseudocode for the probabilistic algorithm.

- Input the T_j's, p_j's, and $numits2$.
- Generate a sequence $\{j_n\}_{n=1}^{numits2}$ of elements of $\{1, 2, ..., K\}$ according to the probabilities using your software's random number generator.
- Input x_0 and compute the set $\{x_0, x_1, x_2, ..., x_{numits2}\}$ according to $x_n = T_{j_n}(x_{n-1})$.
- Plot the subset $\mathcal{O} = \{x_{nmin}, x_{nmin+1}, x_{nmin+2}, ..., x_{numits2}\}$, where $nmin$ ensures the orbit is close enough to A. We may use $nmin = 10$ or 20.

One thing that is clear is that all of the points in $\mathcal{O}$ are close to points in A, as long as $nmin$ is chosen to be moderately large. You can decide on the size of $nmin$ by looking at the contraction factor of $\mathcal{T}$ and the maximum distance M in X. If you want the points to be within ϵ of A, simply choose $nmin$ so that $c^{nmin}M \leq \epsilon$. If you take ϵ to be pixel size and compute the corresponding $nmin$, then you can be sure that the points in $\mathcal{O}$ are, up to computer-visible resolution, points in A.

The harder thing to justify rigorously is that $\mathcal{O}$ fills up A in a representative fashion. In fact it will not if any of the p_is are set to be zero. Otherwise $\mathcal{O}$ should visit each T_js region about $p_j \cdot numits2$ times unless a highly improbable sequence of indices was selected by the random number generator (which is possible, of course). Ordinarily, then, we'd expect it to fill in each T_js part of the image of A pretty well. The experimenter can adjust the p_js if they feel the image is not representative.

3.3.3. A comparison of the two algorithms. The probabilistic algorithm is more efficient than the deterministic algorithm for getting a good picture of the attractor. It is beyond the scope of this course to prove that, but let's try and get a sense of why it is true. In everything that follows let $(X; T_1, ..., T_K)$ be an IFS with contraction factor $c = \max\{c_1, ..., c_K\}$.

EXERCISE 3.17. Suppose you are using the deterministic algorithm on the initial set $\{x_0\}$ for some $x_0 \in X$.

(1) What is the maximum number of elements in the set $\mathcal{T}^{numits}(\{x_0\})$?
(2) Let's make a really rough approximation by calling an "operation" the application of a single T_j to a single element of X. How many operations will the computer have to perform in order to compute $\mathcal{T}^{numits}(\{x_0\})$?

So you can see that *numits* has to be kept fairly small to make sure that your computer doesn't crash and can do all those operations in a timely fashion. However, if *numits* is small, $\mathcal{T}^{numits}(\{x_0\})$ may not be particularly close to A.

EXERCISE 3.18. Compute an upper bound on $d_H(\mathcal{T}^{numits}(\{x_0\}), A)$. (It is convenient to let M represent the maximum Euclidean distance between any two points in X.) Under what circumstances might this distance not be particularly small?

Now let's compare those results with what happens in the probabilistic algorithm.

EXERCISE 3.19. Now suppose that you are beginning with an arbitrary $x_0 \in X$ and using the probabilistic algorithm to compute the orbit out to $numits2$.

(1) Using the same rough definition of an operation as before, how many operations does it take to compute the orbit of x_0 up to $numits2$?
(2) If you cut off the first $nmin - 1$ elements of the orbit and consider the set $\mathcal{O} = \{x_{nmin}, x_{nmin+1}, x_{nmin+2}, ..., x_{numits2}\}$, give an upper bound on $\min\{\epsilon \text{ such that } \mathcal{O} \subset A_\epsilon\}$.
(3) Suppose $nmin$ is greater than the number of iterations $numits$ used in the deterministic algorithm, and suppose $numits2$ is about the same size as the number of points in $\mathcal{T}^{numits}(\{x_0\})$. Do you believe $\mathcal{T}^{numits}(\{x_0\})$ ought to be a better approximation of A than $\mathcal{O}$ is, or is $\mathcal{O}$ the better approximation of A? Explain.

3.4. The Collage Theorem

Suppose that you have an image L that you would like to store in as little space as possible. Or maybe you just want to be able to replicate L using an iterated function system for fun. The collage theorem, proved by Barnsley in 1985 and appearing on page 94 of [**Bar12**], gives you a way to measure how close the attractor of an iterated function system will be to your target image L.

THEOREM 3.20 (The Collage Theorem). *Let X be a compact subset of $\mathbb{R}, \mathbb{R}^d$, or $\mathbb{C}$ and let $L \in \mathcal{H}(X)$ be given. If $(X; T_1, ..., T_K)$ is an IFS with contraction factor c and attractor A, then*

$$d_H(L, A) \leq \frac{d_H(L, \mathcal{T}(L))}{1 - c}.$$

How is this theorem used? It certainly does not tell you how to get an IFS for which $d_H(L, \mathcal{T}(L))$ is small. You basically have to experiment with that, and there is literature out there containing strategies. You have full control over how many maps you want to use, their contraction factors, etc. But no matter what IFS you come up with from that process, the closeness of A to your target image L is bounded by how close L was to $\mathcal{T}(L)$.

EXERCISE 3.21. Will an IFS with a large contraction factor or a small contraction factor do a better job of approximating L? Does the number of T_is you use matter? Explain.

3.5. Exercises

EXERCISE 3.22. Suppose $\mathcal{H}(X)$ is a space of fractals with $A, B \in \mathcal{H}(X)$ and $\delta \geq 0$. Prove that

$$(A \cup B)_\delta = A_\delta \cup B_\delta \tag{3.2}$$

EXERCISE 3.23. Suppose A and B are elements of some space of fractals $\mathcal{H}(X)$ and let $\epsilon > 0$. Prove or give a counterexample to:

$$(A \cap B)_\epsilon = A_\epsilon \cap B_\epsilon$$

EXERCISE 3.24. Suppose $\mathcal{T} : \mathcal{H}(X) \to \mathcal{H}(X)$ is a collage of the maps $T_i : X \to X$, $i = 1, 2, ..., K$. Prove that if $A \subset B$ for $A, B \in \mathcal{H}(X)$, then $\mathcal{T}(A) \subset \mathcal{T}(B)$.

EXERCISE 3.25. Give an example of a transformation $\mathcal{T} : \mathcal{H}(X) \to \mathcal{H}(X)$ for which there are $A, B \in \mathcal{H}(X)$ with $A \subset B$, yet $\mathcal{T}(A) \not\subset \mathcal{T}(B)$.

EXERCISE 3.26. Make an IFS with three transformations in $\mathbb{R}^2$ that we haven't seen in class, and put it into the deterministic algorithm in mathematica. Please turn in your affine maps and your image of the attractor.

EXERCISE 3.27. Consider the spiral for Exercise 1.36. Here's a fun fact: the number of arms does not depend on the central transformation that has a large contraction factor. Instead it depends on the number of highly contractive transformations you put around the outside. Adapt the IFS for Exercise 1.36 to have three arms, and put your answer into the probabilistic mathematica code and see what the attractor looks like. Adjust until you are happy with your image.

EXERCISE 3.28. In this example you are going to play with probabilities when making the Sierpinski triangle. Compare and contrast the number of iterations necessary to produce a "good" view of the triangle when you use the following probabilities:

(1) $p_1 = .33, p_2 = .33, p_3 = .34$.

(2) $p_1 = .2, p_2 = .46, p_3 = .34$.
(3) $p_1 = .1, p_2 = .56, p_3 = .34$.

EXERCISE 3.29. Let $(X; T_1, ..., T_K)$ be an IFS and let $\{j_n\}_{n=1}^{numits2}$ be a sequence of indices from $1, 2, ..., K$ as would be selected for the probabilistic algorithm. Suppose a is an element of the attractor A of the IFS. Prove that the orbit of a in the algorithm is contained in A.

EXERCISE 3.30. Let $(X; T_1, ..., T_K)$ be an IFS with attractor A and let $\{j_n\}_{n=1}^{\infty}$ be a sequence of indices from $1, 2, ..., K$ as would be selected for the probabilistic algorithm (except infinite, theoretically). Suppose $x_0 \in X$ and compute its orbit $\{x_n\}_{n=1}^{\infty}$ as in that algorithm. Prove that $l(x_n, A) = \min\{d(x_n, a)$ such that $a \in A\}$ tends to 0 as $n \to \infty$.

CHAPTER 4

Dimensions

4.1. Motivating examples, or, Fun with length and area.

Let's consider this question: How big is the Koch curve? We can estimate its length by computing the lengths of the curves K_n introduced in Example 1.2 and shown in Figure 2.

Recall that we began with K_0 as a line segment of length 1, for concreteness suppose it is on the x-axis from $x = 0$ to $x = 1$. To construct K_1, four affine transformations with contraction factor $1/3$ are used. That means K_1 is the union of four segments of length $1/3$ and thus has length $4/3$. The union of four copies of K_1, scaled by $1/3$, make up K_2. Each scaled copy has length $4/9$, but there are four of them end-to-end and so K_2 has length $16/9$. At the next stage, four copies of K_2, which have been scaled by $1/3$ and thus have length $16/27$, union together to become K_3, which then has length $64/27$. In general we see that K_n will have length $4^n/3^n$. Since the Koch curve is $\lim_{n\to\infty} K_n$ (where the limit of course is in the Hausdorff metric), we can see that its length is $\lim_{n\to\infty} 4^n/3^n = \infty$.

We can consider this result to mean that we are not really measuring the size of the Koch curve correctly. It's a nice compact set, and to call it's size infinite seems not to give a lot of geometric information about the curve. We'll be able to get finer information using the idea of fractal dimension.

EXERCISE 4.1. Consider the middle-thirds Cantor set introduced in Example 1.1 and shown in Figure 1. Compute its length with the method used for the Koch curve, i.e. by finding the lengths of $C_0, C_1, C_2, ...$ and taking their limit.

EXERCISE 4.2. Consider the Sierpinski triangle introduced in Example 1.3 and shown in Figure 3. Compute its area with the method used for the Koch curve, i.e. by finding the areas of $S_0, S_1, S_2, ...$ and taking their limit.

Since the Koch curve has infinite length in a finite area, maybe we should have tried to compute its area rather than its length. However, it seems fairly clear that the area is zero. You might try to argue that point as follows: The area of any K_n is clearly 0, so it stands to reason that the area of the limit is 0 also. However, in a moment we will see an example of a "space-filling curve", where each approximating set has zero area but the limit is a set with nonzero area.[1]

[1]Notice that this calls into question the length and area computations we've done so far.

A different way to approximate the area of the Koch curve, and one which will generalize to our study of fractal dimension, involves covering each K_n with rectangles, or 'boxes', and compute those areas instead. Of course this will overestimate the area, but we're going to get 0 anyway and the example is instructive.

So, let's be ridiculous and cover K_0 with the box $[0,1]\times[-.5,.5]$, which is shown on the left of figure 1. The area of the box is one and that is an upper bound on the area of K_0.

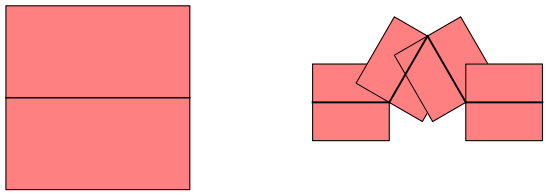


FIGURE 1. The first and second approximations of the Koch curve by boxes.

To construct a box-covering of K_1, simply apply the collage to the box that covers K_1. We see in the figure that K_1 is covered by four copies of this box, scaled *in length* by 1/3, but that means that the *area* of each rescaled box is 1/9. Again we shall be ridiculous and fail to account for the overlap between the boxes, getting an overestimate of the area of K_1 by adding the areas of the four boxes together. This gives us that the area of K_1 is less than 4/9.

EXERCISE 4.3.
(1) Using the same technique, give an upper bound on the area of K_2.
(2) If necessary, repeat for K_3, K_4, etc. until you can generalize your answer to an upper bound on the area of K_n.
(3) Determine the area of the Koch curve.

EXAMPLE 4.4. Here is an iterated function system called the Heighway Dragon (also known as the dragon curve). We begin with the vertical line segment connecting the origin to $(0, 1)$, and use the affine maps

$$T_1\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \begin{pmatrix}.5 & -.5\\ .5 & .5\end{pmatrix}\begin{bmatrix}x\\y\end{bmatrix} \qquad T_2\left(\begin{bmatrix}x\\y\end{bmatrix}\right) = \begin{pmatrix}-.5 & -.5\\ .5 & -.5\end{pmatrix}\begin{bmatrix}x\\y\end{bmatrix} + \begin{bmatrix}0\\1\end{bmatrix}$$

This IFS may seem simpler than that of the Koch curve, but the result looks really different. The reason is that this curve folds back to touch itself repeatedly. Figure 2 shows the first several applications of $\mathcal{T}$, and figure 3 shows H_{13} and the fixed set.

Suppose we were to overestimate the area of the Heighway dragon by covering H_0 with a rectangle. We would not have a sequence of covers that converge to 0. What would happen instead is more subtle, and we delay discussing it for now.

Side note: You can make iterations of the Heighway Dragon using a long, thin strip of paper. If you fold the paper in half and open it up again, you have a scaled version of H_1. If you fold it twice, being careful to fold the same way both times, then when you open it up so that the folds are at right angles you get H_2. Doing a bunch of iterations is a fun party trick.

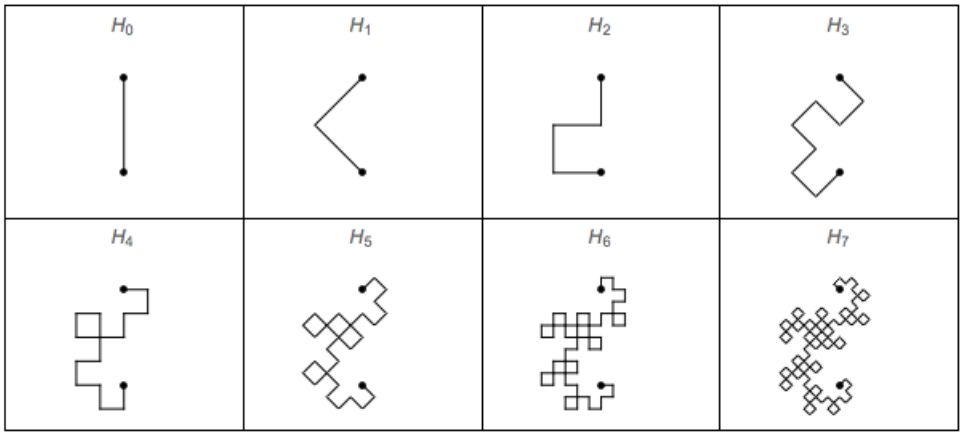


FIGURE 2. The first several iterations of the collage map. Points at the origin and at $(0, 1)$ are marked for reference.

EXERCISE 4.5. Make a really good paper Heighway dragon to show the class. Explain the questions you had or problems you solved to do it.

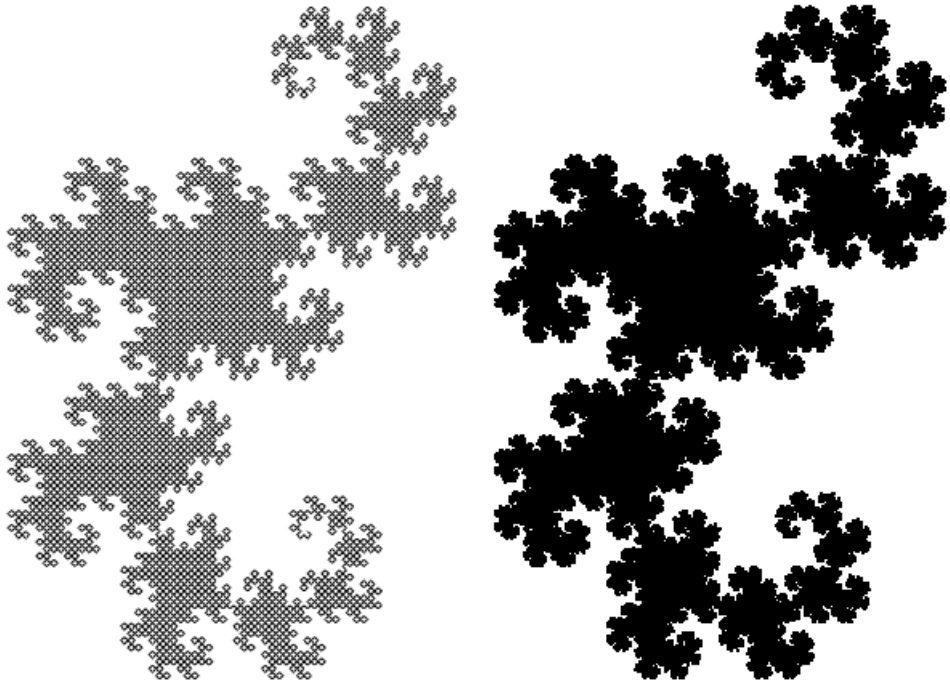


FIGURE 3. The left is H_{13}; the right is the fixed set of the IFS.

4.2. The idea of fractal dimension

Everyone has an intuitive idea of dimension: one-dimensional objects look like (possibly deformed) line segments; two-dimensional objects look like (possibly deformed) pieces of planes; three-dimensional space is the space we live in, where there are three independent directions one can move. In linear algebra a mathematical definition is given for vector spaces: it is the number of basis vectors required to define the space. In a topology class, objects are n-dimensional if they are locally homeomorphic images of $\mathbb{R}^n$ in a technical sense.

In calculus we learn an excellent strategy for calculating the length, area, or volume of an object we will call A. To compute the length $\mathcal{L}(A)$, we approximate it with line segments of length Δx; the length of A is approximately the sum of those segments, i.e. $\mathcal{L}(A) \approx \sum_{\#\text{ segments}} (\Delta x)^1$. We let $\Delta x \to 0$ and if the limit exists in a certain sense it becomes an integral that represents the length of the object. This is where the arclength formulae that you learn in calculus come from.

If we want to estimate the area of A, we can cover it with squares of side length Δx; the area of each square is $(\Delta x)^2$ and so the area $\mathcal{A}(A)$ is approximately $\mathcal{A}(A) \approx \sum_{\#\text{ squares}} (\Delta x)^2$. Again in a calculus class we would let $\Delta x \to 0$, and if the limit existed in a certain sense we'd get the area of A.

The same process works to estimate the volume of A, covering it with cubes of side length Δx. The volume of such a cube is $(\Delta x)^3$, so the volume $\mathcal{V}(A)$ is approximately $\mathcal{V}(A) \approx \sum_{\#\text{ cubes}} (\Delta x)^3$.

The general term we will use for a line segment, square, cube, or an n-dimensional analogue thereof will be a *box*. The n-dimensional *volume* of such a box will be it's side length to the nth power. Thus to estimate the n-dimensional volume of an object A, we approximate the set with boxes with n-volume $(\Delta x)^n$, arriving at the formula

$$\mathcal{V}_n(A) \approx \sum_{\#\text{ boxes}} (\Delta x)^n$$

Clearly, **the dimension of a box appears as the exponent of its volume calculation**, with length corresponding to exponent 1, area to exponent 2, ordinary volume to exponent 3, and n-dimensional volume to exponent n.

There is nothing to stop us from trying to use the calculus procedure to compute the n-dimensional volume of an object whether or not it is n-dimensional. Indeed, calculating the n-volume of an object that isn't fundamentally n-dimensional yields predictable results that we are beginning to understand.

EXAMPLE 4.6. Let A be the circle $x^2 + y^2 = 1$. We use the above process to compute the length of A. Let us estimate A using an inscribed regular k-gon, so that the circle is approximated by k line segments. As we let $k \to \infty$, the approximation will converge in Hausdorff metric to the circle. A k-gon is pictured for $k = 11$ on the left of figure 4.

Each edge of the k-gon has length Δx_k and can be seen as the short side of an isosceles triangle with long sides 1 and angle $2\pi/k$, which means by that $\Delta x_k = 2\sin(\pi/k)$ and so the circumference is approximately $\sum_{\#\text{ segments}} (\Delta x_k)^1 = 2k\sin(\pi/k)$, and from calculus we know (and you will verify in a homework exercise) that the limit as $k \to \infty$ is 2π.

An alternative trick to determine Δx_k, perhaps a bit circular, is to use the fact that the arc subtended by an angle $2\pi/k$ in the unit circle has length $2\pi/k$. Since Δx_k is approximately the length of that arc it follows that $\Delta x_k \approx 2\pi/k$, with the approximation becoming more and more accurate as $k \to \infty$. There are k sides to the k-gon, so the length of the circle is approximately $\sum_{\#\text{ segments}} (\Delta x_k)^1 = k\Delta x_k \approx k(2\pi/k) = 2\pi$. Either way, as $k \to \infty$, our approximation approaches 2π, as it should.

EXERCISE 4.7. Let's now cover the circle with squares whose diagonals are the sides of the regular k-gons above. We picture the covering for $k = 11$ on the right of Figure 4.

(1) Compute the approximate length of the side of each square using the second approximation of the diagonal, which was $2\pi/k$. Use this to approximate the area of each square.
(2) Compute the approximation of the area of A using your squares.
(3) Letting $k \to \infty$, show that the area of A is 0.

There was no particular reason to choose the covering of the circle by squares to look like the left side of figure 4. We could instead have had each edge of the k-gon be a side of the square, or the midline of the square like we did for the Koch curve.

EXERCISE 4.8. Suppose we did the covering of the circle using the midline version of the square covering. Compute the area of A again using this covering.

When doing box-related calculations like these it is very important that the answer be independent of the placement of the boxes. There exist bizarre examples that fail to have this independence, but we will not encounter any. It is also important that the shape of the boxes not matter; that is, if we approximated the shape circles, or rectangles, or sets with different dimensions, we'd always get the same answer as the diameters of the sets went to 0. In the wild woolly world of analysis there exist examples for which this is not true. Such examples will not concern us in this course.

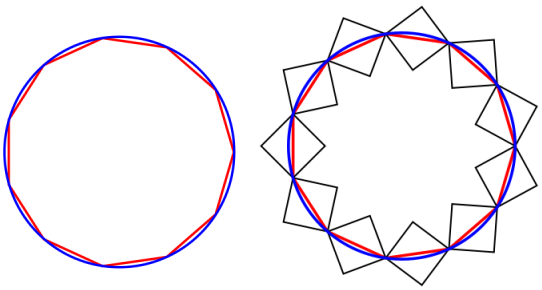


FIGURE 4. Approximating the circle with boxes of dimension 1 and 2.

EXERCISE 4.9. Consider the line segment L connecting the origin to the point $(1,1,1)$ in $\mathbb{R}^3$. For integers k let $\Delta x_k = \sqrt{3}/k$. Calculate the length, area, and volume of L using the calculus method. Namely, for $n = 1,2,3$, cover L with n-boxes of side length $\sqrt{3}/k$, compute the estimate $\sum_{\#\text{ boxes}} (\Delta x_k)^n$, and take the limit as $k \to \infty$.

EXERCISE 4.10. Now let A be the unit square lying in the $x-y$ plane of $\mathbb{R}^3$. Compute the length, area, and volume of A using boxes of side length $\Delta x_k = 1/k$. Note: your estimate of the square using line segments will leave a lot of the square uncovered, but in the Hausdorff metric limit the approximation by line segments will converge to the square.

In each one of the examples we have seen there were three possible outcomes: 0, a finite number, or ∞. We got 0 when we overestimated the dimension of A, we got ∞ when we underestimated the dimension of A, and we got a finite number when we got the right dimension. Except for the Koch curve: we got 0 for 2 dimensions and ∞ for one dimension, suggesting that the correct dimension is somewhere between 1 and 2.

EXERCISE 4.11. Suppose we are covering the approximations K_j of the Koch curve with boxes of side length Δx_k, except that we imagine the boxes to be s-dimensional in the sense that they have volume $(\Delta x_k)^s$.

(1) We cover K_0 by a box with side length 1 as before. The s-volume of that box is then $1^s = 1$. When we apply $\mathcal{T}$ to this box, we get four boxes, where the side length of each is now $\Delta x_1 = 1/3$. Thus our s-dimensional volume approximation for K_1 is then $\sum_{\#\text{ boxes}} (\Delta x_1)^s = 4(1/3)^s = 4/3^s$.
(2) The side length of each box making up the cover for K_2 has length $1/9$. Compute $\sum_{\#\text{ boxes}} (\Delta x_2)^s$.
(3) Find the formula for the s-volume of the cover of K_j, for $j \geq 2$.
(4) We want the limit of that s-volume to be a finite positive number C. Solve for s. (Hint: set your answer from (3) approximately equal to C and solve for s.)

4.2.1. Four tenets of a good notion of dimension. A really good definition of fractal dimension ought to have a number of properties that make sense. The two definitions we will make will satisfy some or all of them in some or all types of examples. These four tenets are described in the lecture notes [FN], p. 102. Let's suppose we have called our candidate for fractal dimension Dim.

- (FAMILIARITY) The dimension of $\mathbb{R}^n$ should be n. Moreover, line segments and vector spaces of linear algebraic or topological dimension 1 should also have a fractal dimension of 1. Similarly a square should have fractal dimension 2, a cube dimension 3, and so on.
- (MONOTONICITY) If $A \subset B$, then $Dim(A) \leq Dim(B)$. In particular this means that if $A \subset \mathbb{R}^n$, then $Dim(A) \leq n$.
- (STABILITY) The process of taking unions of sets should not affect the fractal dimension unpredictably. Thus we prefer our definition of dimension to satisfy $Dim(A \cup B) = \max\{Dim(A), Dim(B)\}$.
- (INVARIANCE) Suppose $T : X \to X$ is an isometry like rotation, reflection, or translation, or some other 'nice' map like a similarity. Then although T might affect the *size* of A, it should not affect its fractal dimension. Thus we prefer that our definition of fractal dimension satisfy, for such a map T, that $Dim(T(A)) = Dim(A)$.

4.3. Similarity dimension

We begin with the simplest type of fractal dimension to compute: the similarity dimension. An advantage is that it can be computed for any IFS once the contraction factors of its transformations are known, but a disadvantage is that it is accurate[2] only in some of those cases. Fortunately, a many of the examples we have been considering are ones for which the similarity dimension makes sense.

[2]Ideally, a set's similarity dimension would agree with its other kinds of dimensions in a wide swath of examples.

4.3.1. Similarity dimension: one scaling factor. We know that if we scale a figure by 1/2, then its length scales by 1/2, but its area scales by 1/4, and its volume scales by 1/8, and by extension its volume in dimension s scales by $(1/2)^s$. If we know that a set such as the Koch curve is made up of a certain number of copies of itself, all scaled by the same amount, we can use that fact to solve for s exactly. This is the foundation for our definition of similarity dimension.

EXAMPLE 4.12. Let us denote the Koch curve by $\mathcal{K}$ and its s-volume by $V_s(\mathcal{K})$. We know that $\mathcal{K}$ is the fixed point of a collage map $\mathcal{T}$ composed of four maps, each with contraction factor 1/3, i.e., that $\mathcal{T}(\mathcal{K}) = T_1(\mathcal{K}) \cup T_2(\mathcal{K}) \cup T_3(\mathcal{K}) \cup T_4(\mathcal{K})$. We also know that $V_s(T_i(\mathcal{K})) = V_s(1/3(\mathcal{K})) = (1/3)^s V_s(\mathcal{K})$. This implies that

$$V_s(\mathcal{K}) = 4(1/3)^s V_s(\mathcal{K})$$

This is an equation we can solve for s. We see quickly that $3^s = 4$, which means that $s \ln 3 = \ln 4$, and so $s = \ln 4 / \ln 3$.

We have now arrived at the dimension $\ln 4 / \ln 3$ for the Koch curve using both parts of the Existence Theorem: in Exercise 4.11 we did it by seeing $\mathcal{K}$ as the limit of sets under repeated iteration of $\mathcal{T}$, and in the previous example we did it using the fact that $\mathcal{K}$ is the fixed point of $\mathcal{T}$. We can be fairly confident that this number is representative of the fundamental 'size' of the Koch curve.

EXERCISE 4.13. Use the technique of example 4.12 to compute the similarity dimension of the middle-thirds Cantor set $\mathcal{C}$.

EXERCISE 4.14. Use the technique of example 4.12 to compute the similarity dimension of the Sierpinski triangle $\mathcal{S}$.

Computation of similarity dimension is appropriate for sets $A \in \mathcal{H}(X)$ that are unions of rescaled copies of themselves, and the scaling factors don't all have to be the same. However, the transformations have to be *similarities* in the sense of your high school geometry course. Let us take a moment to review them.

4.3.2. Similarity transformations. Although in this course we focus on compact subsets of Euclidean space, the following definition of similarity holds in any metric space.

DEFINITION 4.15. A transformation $T : X \to X$ is a *similarity* if there is a positive $c \in \mathbb{R}$ such that for every $x, y \in X$, $d(T(x), T(y)) = c\, d(x, y)$. We call c the *similarity ratio* of T.

In Euclidean space, transformations are similarities if they are certain kinds of affine maps. In one dimension, any affine map will do as long as it is invertible.

EXERCISE 4.16. Prove that $T : \mathbb{R} \to \mathbb{R}$ is a similarity if and only if there are constants $a, b \in \mathbb{R}$ with $a \neq 0$ such that $T(x) = ax + b$.

In two dimensions, similarity transformations are the ones that you actually learned about in high school geometry: they take triangles to similar triangles (that is, they preserve angles). There are several ways to identify when an affine transformation is a similarity, of which we give three. The first two are geometric and the third comes from linear algebra.

One way is to look at what the transformation does to the standard basis vectors $\vec{e_1}$ and $\vec{e_2}$. If it sends them to vectors that are orthogonal and that have the same length, then the transformation is a similarity. This trick works to identify a similarity in any dimension.

Another way is to know all possible *isometries* of $\mathbb{R}^2$, that is, all possible maps that don't change distance between points. Those are simply the translations, rotations, and reflections. If your affine map is a scaling factor times one of these isometries, then it is a similarity.

A linear algebraic way to determine if a transformation in any $\mathbb{R}^d$ is a similarity is to determine the eigenvalues and eigenvectors of the underlying linear transformation. The matrix must be diagonalizable in that it must have a basis of (possibly complex) eigenvectors, and its eigenvalues must all have the same magnitude. Under those conditions the eigenvalues stretche the basis vectors by the same amount, and thus stretches every vector in $\mathbb{R}^d$ by the same amount.

EXERCISE 4.17. Consider the transformations in the collage for Barnsley's fern. Determine which, if any, are similarities. If it is a similarity, determine the scaling factor.

Notice that a similarity transformation is a contraction whenever $c \in [0, 1)$.

4.3.3. Similarity dimension: multiple scaling factors. Now let us consider the fractal dimension of the attractor of the IFS $(X; T_1, ..., T_n)$. If T_i is a similarity of X for $i = 1, 2, ...n$ then it is an appropriate system on which to make a definition of similarity dimension. The similarity dimension $Dim_S(A)$ will be defined via the scaling factors of its contraction maps in a manner quite similar to our discussion of the Koch curve.

When we computed the dimension of the Koch curve in example 4.12, we needed to solve the equation

$$(1/3)^s + (1/3)^s + (1/3)^s + (1/3)^s = 1,$$

which we of course simplified to solving $4(1/3)^s = 1$. In exercise 1.30 we saw a Cantor-set-like example made of transformations of different scaling factors that makes a good test case for similarity dimension.

EXAMPLE 4.18. Let $X = [0,1]$, $T_1(x) = x/4$, $T_2(x) = x/3 + 1/3$, and $T_3(x) = x/4 + 3/4$. The attractor A of the IFS $(X; T_1, T_2, T_3)$ is pictured in exercise 1.30. If we have the correct dimension s of A calculated, then its volume $V_s(A)$ must satisfy

$$V_s(A) = V_s(1/4A) + V_s(1/3A) + V_s(1/4A),$$

giving rise to the equation

$$V_s(A) = (1/4)^s V_s(A) + (1/3)^s V_s(A) + (1/4)^s V_s(A).$$

Thus the dimension s must be the solution to

$$(1/4)^s + (1/3)^s + (1/4)^s = 1.$$

Mathematica finds the approximation $s \approx 0.856738$ using the expression[a]

FindRoot[(1/4)^s + (1/3)^s + (1/4)^s - 1, {s, 1}]

[a] I told it to look for the root near $s = 1$ since s should be somewhere between 0 and 1.

So we see that when the scaling factors are different we arrive at the equation

$$c_1^s + c_2^s + \cdots + c_n^s = 1.$$

We need to know that that equation has a unique solution, which is the content of this lemma that we state without proof.

LEMMA 4.19 (see [**Edg90**], p. 105). *Suppose $c_1, c_2, ..., c_n$ are numbers in $[0,1)$ for all i. Then there is a unique number $s \geq 0$ such that $c_1^s + c_2^s + ... + c_n^s = 1$. The number s is 0 if and only if there is only one nonzero value for c_i.*

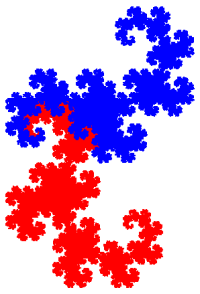
An unfortunate but unavoidable fact is that in general there is no way to solve for s unless the c_i's are related to each other in some way. However it is very easy to get your computer or calculator to give you an approximation that is accurate to as many digits as you like. Here is the official definition of similarity dimension.

DEFINITION 4.20. Let A be the attractor for an iterated function system $(X, T_1, T_2, ..., T_n)$ for which each T_i is a similarity with similarity ratio $c_i \in [0,1)$. The *similarity dimension* of A, denoted $Dim_S(A)$, is defined to be the solution s to the equation

$$c_1^s + c_2^s + ... + c_n^s = 1.$$

EXERCISE 4.21. Compute the similarity dimension of the spiral fractal from exercise 1.36. Give the equation that it solves and obtain a numerical estimate of its value.

EXERCISE 4.22. The Heighway dragon is the union of two copies of itself under the two transformations that make up its collage map, given in example 4.4. Compute its similarity dimension.



EXERCISE 4.23. Compute the equation for the similarity dimension of the fractal in exercise 1.37 and give an numerical estimate for its dimension.

There is a clever way to solve for the similarity dimension in the last exercise exactly that capitalizes on the fact that $1/4 = (1/2)^2$.

EXERCISE 4.24. In the equation obtained for the similarity dimension in the previous exercise, let $x = (1/2)^s$. Rewrite the equation in terms of x and solve using the quadratic formula. This gives a solution for x that can be solved for s exactly. Give the exact solution for s and approximate it with a calculator value.

Here is a simple example where the similarity dimension does not do so well.

EXAMPLE 4.25. Let $X = [0, 1]$, $T_1(x) = 2x/3$, and $T_2(x) = 2x/3 + 1/3$. Each of T_1 and T_2 are similarities and so the definition of similarity dimension applies and we find that $Dim_S(A)$ is the solution s to $2(2/3)^s = 1$. Thus $Dim_S(A) = \ln 2/\ln 1.5 > 1$, a troubling result. Because of the amount of overlap, it turns out that $\mathcal{T}(X) = X$ and so the attractor A of $\mathcal{T}$ is all of $[0, 1]$. By the tenet of familiarity, we should therefore have arrived at the dimension of A as being 1. We also failed the tenet of monotonicity: the dimension certainly should not have exceeded 1 since A is a subset of $\mathbb{R}$.

The problem evident in this example is that its IFS is *overlapping*. Later in this chapter we will define what if means for an IFS to be *totally disjoint* or *just-touching*. The similarity dimension of an IFS satisfying either of those two conditions will be seen to have an 'accurate' similarity dimension.

So we see some obvious benefits and obvious drawbacks to our definition of similarity dimension. The main benefits are that it is quite easy to compute, at least numerically, and that it is quite natural for the examples for which it is defined. There are a few drawbacks. One is that it is not defined for iterated function systems such as Barnsley's fern, where the collage contains transformations that are not similarities. Another is that even when it is defined, there are situations such as our previous example where the result is misleading. The good news is that it is possible to write down which situations give us trouble, and that the number obtained from the equation for similarity dimension has meaning even in cases like

the fern. To discuss this further we need to think about a more general definition of dimension.

4.4. Box-counting dimension

Let $A \subset \mathbb{R}^d$ be a compact set. A key idea necessary to compute dimension of A is how many boxes of a given side length are needed to cover A. Let us co-opt the notation we used earlier for balls of radius ϵ and let

$$B(x, \epsilon) = \text{box of side length } \epsilon \text{ centered at } x$$

In this definition we mean that a box in $\mathbb{R}$ is an interval of length ϵ with x at its center; in $\mathbb{R}^2$ it is a square of side length ϵ with x at its center, in $\mathbb{R}^3$ it is a cube, and so on. So the boxes are easy to picture even when their s-volume is not.

DEFINITION 4.26. Let $A \in \mathbb{R}^d$ be compact and let $\epsilon > 0$. The smallest positive integer N for which $A \subset \bigcup_{n=1}^{N} B(x_n, \epsilon)$ is called the *ϵ-covering number* of A and is denoted by $\mathcal{N}(A, \epsilon)$.

Let us take a moment to parse this definition. For any integer, say K, we can choose any K points in X to call $x_1, x_2, ..., x_K$. Those points can be taken to be the centers of closed boxes of radius ϵ; in that case the union of all of those boxes will be some subset $\bigcup_{n=1}^{K} B(x_n, \epsilon)$ of X. Maybe our set A is in that union and maybe it is not. Maybe we should have made a more judicious choice for our centers. What is clear is that there are infinitely many choices of K and then (mega-uncountably) infinitely many choices for $x_1, x_2, ..., x_K$ we could make for the centers. In this definition we consider all of them simultaneously, focusing in particular on choices where $A \subset \bigcup_{n=1}^{K} B(x_n, \epsilon)$. We ask ourselves the question, what is the smallest possible value of K for which $A \subset \bigcup_{n=1}^{K} B(x_n, \epsilon)$? The answer to that is $\mathcal{N}(A, \epsilon)$.

EXAMPLE 4.27. Let $A \in \mathbb{R}^2$ be the line connecting the origin to $(1, 1)$ and suppose that $\epsilon = 1$. Then $\mathcal{N}(A, \epsilon) = 1$ because we can take x_1 to be the midpoint $(1/2, 1/2)$. The box of side length ϵ will equal the unit square and thus contains A.

Slightly more interesting might be to take $\epsilon = 1/2^n$. In this case we can space out our centers along A to see that we need 2^n boxes to cover A.

EXERCISE 4.28. Verify the example for $n = 2$ and $n = 3$. In both cases make a sketch of $\bigcup_{n=1}^{N} B(x_n, \epsilon)$. Write a formula for placing the centers for a general n.

Of course in the previous examples the ϵ was selected to be particularly nice relative to the set A. The result was that we were able to cover A without needing much overlap of the boxes. In general the fit won't be quite as nice, as you see in the next exercise.

EXERCISE 4.29. Continue with example 4.27, this time letting $\epsilon = e^{-1}$ and $\epsilon = \pi^{-1}$.

EXAMPLE 4.30. Let A be a square of side length 1 in $\mathbb{R}^2$. It doesn't matter which one, but for concreteness center it at $(1/2, 1/2)$. Consider $\epsilon = 1/4$. Then $\mathcal{N}(A, \epsilon)$ is 16 because we need 16 squares of that side length to completely cover the A.

EXERCISE 4.31. In the previous example, consider $\epsilon = 1/8$ and determine $\mathcal{N}(A, \epsilon)$. Compare and contrast to your answer for the same epsilon for the line connecting the origin to $(1, 1)$.

The fact that $\mathcal{N}(A, \epsilon)$ is finite for any $A \in \mathcal{H}(X)$ is a consequence of compactness. (Indeed, the formal definition of compactness is that any open cover contains a finite subcover.)

Notice that unless A is a finite set it will be true that $\mathcal{N}(A, \epsilon) \to \infty$ as $\epsilon \to 0$. The real question is, *how* does it go to infinity? To give this question meaning, compare and contrast examples 4.27 and 4.30. It is clear that they will go to infinity at quite different rates, and this is related to the fact that they are fundamentally of different dimensions. The main diagonal is one-dimensional, whereas the square is two-dimensional; this fact appears in the covering numbers for various epsilons.

EXERCISE 4.32. Let $\mathcal{C}$ be the middle-thirds Cantor set and consider the subset of $\mathbb{R}^2$ given by

$$A = \{(x, y) \text{ such that } x \in \mathcal{C} \text{ and } y \in [0, 1]\}$$

The set A is pictured in figure 5 below. Let $\epsilon_n = 1/3^n$. Compute $\mathcal{N}(A, \epsilon_n)$ for $n = 1, 2$ and give a formula for a general n.

EXERCISE 4.33. Let A be a compact subset of $\mathbb{R}^d$ and let $\epsilon > 0$. Suppose that you are given $\mathcal{N}(A, \epsilon)$ and let $V_s(A)$ denote the s-volume of A.

(1) Given an approximate equation of $V_s(A)$ in terms of $\mathcal{N}(A, \epsilon)$. That is, fill in the right side of the expression $V_s(A) \approx$ _____
(2) Solve your approximate equation for s, treating your $\approx$ sign as an equals sign.
(3) Your approximate equation must be true for all $\epsilon > 0$, and in fact becomes increasingly accurate as $\epsilon \to 0$ since the approximation by boxes becomes more accurate. What expression do you get for s in the limit?

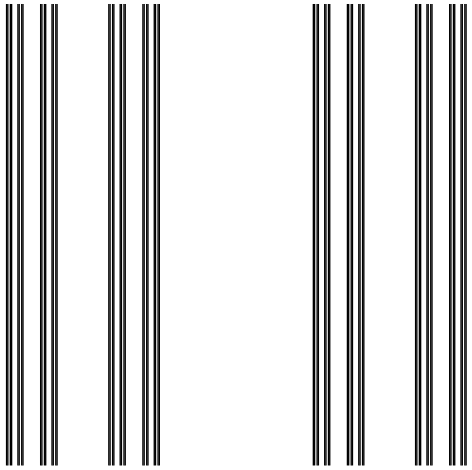


FIGURE 5. The set A for exercise 4.32.

This exercise leads us directly to another definition of dimension, simply called fractal dimension in [**Bar12**, p. 173] and box-counting dimension in [**FN**, p. 118].

DEFINITION 4.34. Let A be a compact subset of $\mathbb{R}^d$. The *box-counting dimension* of A is defined to be

$$Dim_B(A) = \lim_{\epsilon \to 0} \left(\frac{\ln(\mathcal{N}(A, \epsilon))}{\ln(1/\epsilon)} \right),$$

provided this limit exists.

The limit does not always exist, in which case it can be convenient to look at the "upper" and "lower" box dimensions instead, which are defined in terms of lim sup and lim inf. We will not pursue that issue further here, but the interested reader can refer to [**PC09**, p. 87] or [**Fal06**, p. 41] for more details.

EXERCISE 4.35. Let X be the unit square and let $A = \{a, b, c\}$ be any set of three nonequal points in X. Compute $Dim_B(A)$.

It can be difficult to compute $\mathcal{N}(A, \epsilon)$ for general ϵ in many cases, making the limit in our definition of box-counting dimension intractable to use. If the limit exists, there are a number of equivalent ways of computing it, some of which are summarized in [**Fal06**, p. 43] and in the *Box Counting Theorem* of [**Bar12**, p. 175].[3] Of particular interest to us is the fact that instead of letting $\epsilon \to 0$ continuously it suffices to choose a constant $M > 1$ and let $\epsilon_n = 1/M^n$ in our computations of the box-counting dimension. We have the following lemma that makes the computation convenient.

[3] I should probably expand the boxes-in-a-grid idea that is used for approximation of the dimension for later versions of these notes.

LEMMA 4.36. *Let A be a compact subset of $\mathbb{R}^d$ and let $M > 1$. If the box-counting dimension of A exists, then*

$$Dim_B(A) = \lim_{n\to\infty}\left(\frac{\ln(\mathcal{N}(A, 1/M^n))}{n\ln(M)}\right)$$

EXERCISE 4.37. Fill in the blanks to prove the lemma.

PROOF. Let us assume that $Dim_B(A)$ exists. If that is the case, then since the limit exists as ϵ goes to 0 it must also exist when we let $\epsilon = 1/M^n$ and let $n \to \infty$ because in that case $1/M^n$ _____. [a] Thus we see that by definition $Dim_B(A) = \lim_{\epsilon\to 0}\left(\text{_____}\right)$, which by substituting $\epsilon = 1/M^n$ becomes $Dim_B(A) = \lim_{n\to\infty}\left(\text{_____}\right)$. By applying _____, the denominator becomes _____, which finishes the proof. □

[a] The interplay between continuous and sequential limits requires rigorous treatment in an analysis class, but in this situation there is no logical problem. A logical problem could arise if we tried to argue that a limit existing for $\epsilon = 1/M^n$ as $n \to \infty$ implied that it existed for $\epsilon \to 0$. We're not doing that here. Our sequence of ϵs is going to 0 along with the rest of them.

EXERCISE 4.38. Compute the box-counting dimension of the Koch curve.

EXERCISE 4.39. Compute the box-counting dimension of the example in exercise 4.32.

4.5. When the dimensions are equivalent

Box-counting and similarity dimension agree when the set A is the attractor of an iterated function system that not only is made of similarities, but also satisfies one of the first two conditions given in this definition.

DEFINITION 4.40. Let $(X; T_1, ..., T_n)$ be an iterated function system and let A be its attractor. The iterated function system is said to be *totally disconnected* if $T_i(A) \cap T_j(A) = \emptyset$ for all $i \neq j$ with $i, j \in \{1, 2, ..., n\}$. It is said to be *just-touching* if there is an open set[a] $\mathcal{O} \subset A$ for which

(1) $T_i(\mathcal{O}) \cap T_j(\mathcal{O}) = \emptyset$ for all $i \neq j$ with $i, j \in \{1, 2, ..., n\}$, and
(2) $\mathcal{T}(\mathcal{O}) \subset \mathcal{O}$. If the IFS is neither totally disconnected nor just-touching it is said to be *overlapping*.

[a] A set is open if its complement $X/\mathcal{O}$ is closed.

EXAMPLE 4.41. The middle-thirds Cantor set satisfies the definition of totally disconnected since $T_1(\mathcal{C}) \cap T_2(\mathcal{C}) = \emptyset$.

EXERCISE 4.42. Determine whether the Sierpinski triangle and Koch curve are totally disconnected. If they are not, give elements of A that are in more than one image of A.

The just-touching definition is a bit more subtle but isn't so bad when you understand it. Basically what it is saying in a mathematically precise way is that the attractor is allowed to overlap just on its "boundary points". The points you gave in the last exercise are such boundary points. In such a situation, the set $\mathcal{O}$ can be taken to be the part of the attractor that is on the "inside" in a way we will see in the next example. When the collage is applied to it, it stays on the inside and the images of the individual maps in the collage do not overlap. That is enough, it turns out, to ensure that the dimension computation is accurate.

EXAMPLE 4.43. We construct the set $\mathcal{O}$ for the Sierpinski triangle as follows. Here X is the unit square and $\mathcal{S}$ denotes the Sierpinski triangle. Let $P \subset X$ denote the right triangle connecting the origin to $(1, 0)$ and $(0, 1)$, which is the boundary of $\mathcal{S}$. Then let $\mathcal{O} = \mathcal{S}/P$, i.e., $\mathcal{O}$ is the set of points in $\mathcal{S}$ that are not in this boundary triangle. Note that $\mathcal{O}$ is open since its complement in X is just P, which contains its limit points and is therefore closed.

Now we must verify that it satisfies the two conditions. The first condition seems relatively clear, since the only way for the images to overlap is on the boundary, which we have removed. The second condition also checks out, since $\mathcal{T}(\mathcal{O})$ does not contain any portion of P and thus must be inside $\mathcal{O}$.

EXERCISE 4.44. Determine a set $\mathcal{O}$ for the Koch curve to show that it is just-touching.

EXAMPLE 4.45. The iterated function system in example 4.25 is overlapping.

THEOREM 4.46. [**Bar12**, p. 183] *Let $(X; T_1, ..., T_n)$ be an iterated function system, let $c_1, c_2, ..., c_n$ be the similarity ratios of $T_1, T_2, ..., T_n$ respectively, and let A be the attractor of the IFS. If the IFS is totally disconnected or just-touching then the box-counting dimension is the similarity dimension of A. If the IFS is overlapping, then $Dim_B(A) \leq Dim_S(A)$.*

4.6. Exercises

EXERCISE 4.47. Go back through your introductory calculus notes or book and give the proof that $\lim_{k\to\infty} 2k\sin(\pi/k) = 2\pi$. You may assume the identity $\lim_{h\to 0} \frac{\sin(h)}{h} = 1$.

EXERCISE 4.48. Use the methods of section 4.2 and especially problem 4.11 to obtain a value s for the dimension of the middle-thirds Cantor set.

EXERCISE 4.49. Let $A \subset \mathbb{R}$ be the set given by $A = \{1/n$, such that $n = 1, 2, 3, ...\}$. Find the box-counting dimension of A.

EXERCISE 4.50. Let $A \subset \mathbb{R}$ be the set given by $A = \{1/n^2$, such that $n = 1, 2, 3, ...\}$. Find the box-counting dimension of A.

CHAPTER 5

Julia Sets

Our primary source for fractals up to this point has been iterated function systems. We have seen that a collection of contraction maps $T_i : X \to X$, acting together as a collage, have a unique fixed set A. That set A is a collection of points in our original space X that have some or all of the properties identified in section 1.6: a complicated, "fine" structure that is too detailed to be described in traditional geometric terms; that structure is often a form of self-similarity; and the dimension is often not an integer.

Like the attractors A we've been studying, Julia sets are subsets of X that behave in a certain way relative to a transformation. The transformations will no longer be affine,they aren't contractions, and the behavior we study is a little different. We can already find interesting examples when we think about polynomial transformations in the complex plane in section 5.2.

We know from experimentation on iterated function systems that seemingly minor alterations to the T_i's can have unpredictable effects on the set A. Similarly, by changing a single parameter in the transformation we can change its Julia set dramatically. To pique your interest, figure 1 shows two Julia sets that come from the transformation $f(z) = z^2 + \lambda$.

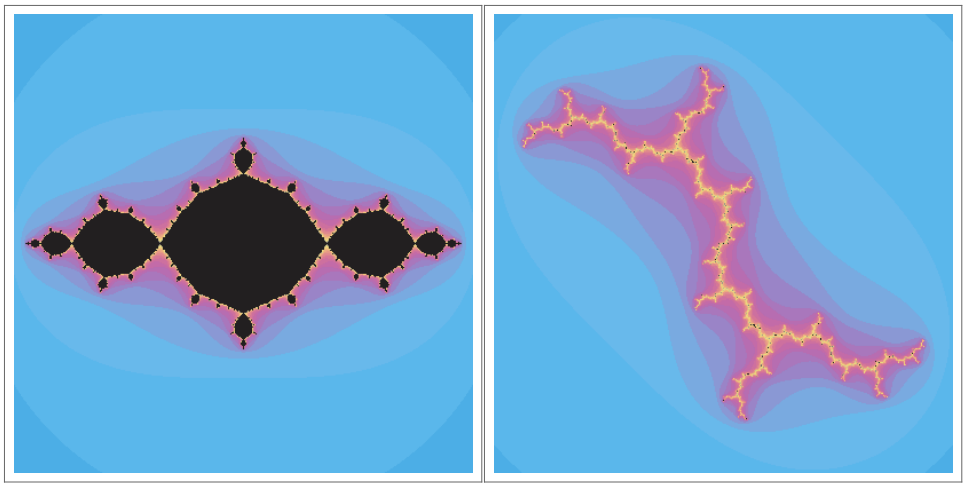


FIGURE 1. The filled Julia sets for $\lambda = 1$ (left) and $\lambda = i$ (right).

5.1. Basic example: dynamical systems in $\mathbb{R}$

We recall some of the notation and terminology from chapter 1 and, with $X \subset \mathbb{R}$, consider some transformation $f : X \to X$. We define $f^2(x)$ to be $f \circ f(x) = f(f(x))$, not $(f(x))^2$, and in general $f^n(x)$ is defined to be $f \circ f \cdots \circ f(x)$ (n times). For $x \in X$ consider the sequence

$$\mathcal{O}(x) = \{x, f(x), f^2(x), f^3(x), ...\}$$

We've called $\mathcal{O}(x)$ the *orbit* of x and we use the notation (X, f) to denote the *dynamical system* defined by f acting on X.

EXERCISE 5.1. Let $A = \mathbb{R}$ and let $f(x) = x^2$. Compute the orbits under f of $x = 2, x = -1$, and $x = 1/2$. Make some sort of graphical depiction of these orbits. Try to give a full qualitative[a] description of the possible behavior of points in this system.

[a] "Qualitative" is used here in contrast to "quantitative". In a qualitative description you give a general description of the behavior that depend on qualities of the initial points. A relevant quality could be something like "falls into this set" or "is at least that big" or some such.

You almost certainly noticed that there are two fixed points of the system in the previous example. One of them counts as attracting and the other repelling. The behavior of orbits near fixed points plays an important role in dynamical systems theory. Another thing you will have noticed is that some of the orbits stay bounded and others go off to infinity.

DEFINITION 5.2. Let $X \subset \mathbb{R}, \mathbb{R}^d$, or $\mathbb{C}$ and let $f : X \to X$ be a polynomial transformation. The *filled Julia set* F_f of f is the set of all points in X whose orbits under f stay bounded. The *Julia set* J_f of f is the boundary[a] of F_f.

[a] The boundary of a set A is the set of all points $x \in X$ such that $B(x, \epsilon)$ contains points in A and not in A for all $\epsilon > 0$. In straightforward examples, the boundary of a set is what you'd expect it to be.

Note that there are more technical characterizations of Julia sets that hold for larger classes of functions. A definition for rational functions, which are quotients of polynomials, is found on [**Bar12**, p. 278], and a definition of 'analytic' functions (ones with convergent Taylor series) appears in [**Fal06**, p. 219] and [**FN**, p. 334]. For the case of polynomial transformations these technical definitions coincide with ours.

EXERCISE 5.3. Identify F_f and J_f for exercise 5.1.

In general, it turns out to be relatively easy to conceptualize F_f and to write computer code to generate images of it. It is more difficult to work with the Julia set definition, which requires us to have an analysis-level comprehension of the definition of boundary points footnoted below. For that reason, in this class we will restrict our attention to filled Julia sets.

EXERCISE 5.4. Consider the function $f : \mathbb{R} \to \mathbb{R}$ given by $f(x) = x^2 - 2$.

(1) Pick several choices for x and investigate their orbits. Try to find examples of different behavior.
(2) From your experimental points, determine the filled Julia set F_f.

Of course your answer to part (2) isn't a proof, and it would require some more work to make a proof. One way to do it would be to show that if $|x| \geq$ some number, then $|f^n(x)| \to \infty$. This is where the idea of the "escape region" comes in. You can think of an escape region as being all points that are at least K in magnitude and for which $|f(x)|$ is at least as big as $M|x|$, for some $M > 1$. Formally,

DEFINITION 5.5. Let $X \subset \mathbb{R}, \mathbb{R}^d$, or $\mathbb{C}$ and let $f : X \to X$. A region of the form $V = \{x \in X \text{ such that } |x| > K\}$ is an *escape region* for f if there is an $M > 1$ for which $|f(x)| \geq M|x|$ for all $x \in V$.

EXAMPLE 5.6. Let's try and find an escape region using, say, $M = 2$ for exercise 5.4. We need to find a K that works for the definition. So we need to solve $|f(x)| \geq 2|x|$. Noticing the symmetry that $f(x) = f(-x)$, we work with positive values of x and try to solve $x^2 - 2 \geq 2x$. A little bit of the quadratic formula later we see that if $x \geq 1 + \sqrt{3}$, then $|f(x)| \geq 2|x|$. So an escape region for $f(x) = x^2 - 2$ is the set $V = \{x \in \mathbb{R} \text{ such that } |x| > 1 + \sqrt{3}\}$.

You may feel like we worked a little bit backwards in that example, but choosing K first and trying to find M was a bit difficult algebraically. For our purposes the precise value of M doesn't matter as much as figuring out what the K is for *some* value of M. Any $M > 1$ will be good enough for the escape-time algorithm, as we shall see soon. For now, though, let us notice a handy way to use escape regions to rule out a point being in F_f.

EXERCISE 5.7. Let $f : X \to X$ and suppose that you have been given an escape region V along with the M and K from the definition.

(1) Show that if $x \in V$, then $f(x) \in V$ also.
(2) Find a lower bound on $|f^n(x)|$ for $x \in V$.
(3) Deduce that if $x \in V$ then $|f^n(x)| \to \infty$ as $n \to \infty$.

This exercise showed that if the orbit of a point ever enters an escape region, then it is destined to wander off to infinity and thus cannot be in the filled Julia set.

EXERCISE 5.8. Investigate the dynamics and compute the filled Julia sets for the following examples.

(1) $f(x) = x + 7$
(2) $f(x) = x/2$
(3) $f(x) = 2x$
(4) $f(x) = x^2/7$

5.2. Classic example: dynamical systems in $\mathbb{C}$.

5.2.1. Review of complex numbers. The complex plane is visualized in much the same way as $\mathbb{R}^2$, with the point $(x, y) \in \mathbb{R}^2$ corresponding to $z = x + iy \in \mathbb{C}$. The horizontal axis consists of the purely real numbers $z = x + 0i \in \mathbb{R}$ and the vertical axis consists of the purely imaginary numbers $z = 0 + iy$. The addition of complex numbers algebraically is $(a+ib)+(c+id) = (a+c)+i(b+d)$. Geometrically this follows the usual parallelogram rule for adding vectors (ak, b) and (c, d) in $\mathbb{R}^2$.

The operation that sets $\mathbb{C}$ apart from $\mathbb{R}^2$ is multiplication. It is defined using the ordinary distributive law

$$(a + ib)(c + id) = (ac - bd) + i(ad + bc)$$

and is therefore quite natural to compute.

EXERCISE 5.9. Let $z_1 = 1 + i, z_2 = 1 + \sqrt{3}i$, and $z_3 = -2i$. Compute $z_i z_j$ for all possible combinations of i and j.

The geometric interpretation of the product of two complex numbers is very interesting and it is essential that you keep it in mind throughout this chapter. To understand it, we must think of z as being written in polar coordinates, so $z = a+ib = r(\cos\theta+i\sin\theta)$, where $r = |z|$ and θ is measured in radians counterclockwise from the positive real axis. In the next exercise you may start to see what happens geometrically with the products you did in the previous exercise.

EXERCISE 5.10. For the complex numbers in example 5.9,

(1) Compute r_i and θ_i for $i = 1, 2, 3$, and
(2) Compute the length and angle of $z_i z_j$ for an assortment of i's and j's of your choosing, looking for a relationship to the lengths and angles of z_i and z_j. Graph z_i, z_j, and $z_i z_j$ on the same set of axes.

The pattern that appears in part (2) is universal and allows us to quickly visualize the product of two complex numbers. All we have to do to get the product is multiply the lengths and add the angles. Let's make that precise.

EXERCISE 5.11. Show that if $z = r(\cos\theta + i\sin\theta)$ and $w = s(\cos\phi + i\sin\phi)$, then

$$zw = rs(\cos(\theta + \phi) + i\sin(\theta + \phi))$$

Our work with Julia sets will primarily use polynomial transformations $f(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_1 z + a_0$, where $a_0, ..., a_{n-1} \in \mathbb{C}$. Having a geometric understanding for the terms will be quite helpful to you.

EXERCISE 5.12. Explain geometrically the relationship between z^n and z for any given complex number z.

EXERCISE 5.13. Let $z = \cos(\pi/6) + i\sin(\pi/6)$. Make a sketch of z^n for $n = 0, 1, 2, 3, ...$

EXERCISE 5.14. Let $z = \cos(3) + i\sin(3)$ (notice that $\theta = 3$ measured in radians is not a rational multiple of π but it is nonetheless a perfectly good angle). Sketch z, z^2, z^3, and think about what happens to z^n as $n \to \infty$.

5.2.2. The classic example in $\mathbb{C}$. The original transformations Mandelbrot performed computer calculations on in the 1970's was $f(z) = z^2 + \lambda$ [**FN**, Ch. 7], and we will follow in his footsteps. The question he asked was, for which values of λ is the Julia set connected? This is a topological question beyond the scope of our course, but it turns out that the answer can be understood in much more simple language that we will investigate this more in Chapter 6. For now, let's experiment with a few λs.

EXERCISE 5.15. Let $f : \mathbb{C} \to \mathbb{C}$ be given by $f(z) = z^2$.

(1) Compute and make a sketch of the orbits of $z = 0, z = \cos(\pi/6) + i\sin(\pi/6), z = \cos(3) + i\sin(3)$, and $z = 2i$.
(2) What is the filled Julia set for this transformation?
(3) Find an escape region for f, showing your choice of K and M.

EXERCISE 5.16. For $\lambda = 1$,

(1) Compute the first few elements of the orbits of $z = 0, z = i$, and $z = 1$.
(2) Find the fixed points of f, if any. At least those are in the filled Julia set.

EXERCISE 5.17. For $\lambda = -1$,

(1) Compute the first few elements of the orbits of $z = 0, z = i$, and $z = 1$.
(2) Try and find an element of the filled Julia set that isn't a fixed point.

There is a fairly simple way to find an escape region for $f(z) = z^2 + \lambda$ for all values of λ. Let $K = \max\{|\lambda|, 2.1\}$. Let $V = \{z \in \mathbb{C} \text{ such that } |z| > K\}$. Let us prove this is an escape region for f.

PROOF. We are interested in showing that $|z^2 + \lambda| \geq M|z|$ for some $M > 1$, where $z \in V$. You can use the triangle inequality to prove that $|z^2 + \lambda| \geq |z^2| - |\lambda|$. Since $|z| \geq |\lambda|$ this implies that $|z^2 + \lambda| \geq |z^2| - |z|$. By our discussion of the multiplication of complex numbers we know that $|z^2| = |z|^2$, so $|z^2 + \lambda| \geq |z|^2 - |z| = |z|(|z| - 1)$. Now we can use the fact that $z \in V$ by noticing that $|z| > K$ implies that $|z| - 1 \geq 2.1 - 1 = 1.1$. This means that $|f(z)| = |z^2 + \lambda| \geq 1.1|z|$ for all $z \in V$ and the definition of an escape region is satisfied. □

Notice that any number of the form $2 + \epsilon$ with $\epsilon > 0$ would have worked in the previous proof. In that case $M = 1 + \epsilon$. As it turns out, you can actually use $K = \max\{|\lambda|, 2\}$ as an escape region without disruption to the escape-time algorithm even though it doesn't quite satisfy our definition.

EXERCISE 5.18. Prove that $|z^2 + \lambda| \geq |z^2| - |\lambda|$ using the triangle inequality.

5.3. The Escape Time Algorithm

Here is the algorithm that was used to create the images in Figure 1. The basic idea is that we test values of z by looking at whether $f^n(z)$ is in our escape region V or not. If we let n be large enough and find that $f^n(z) \notin V$, we surmise (possibly incorrectly) that $z \in F_f$. This algorithm will certainly be accurate about finding points that are *not* in F_f, since we know that as soon as an orbit enters V it is not bounded. We will probably see some false positives, though, especially if n is small, but we are just getting an image and can only be accurate up to the size of a pixel no matter what.

Let's refine this idea more, with the goal of arriving at an algorithm we can put in the computer. The basic way the algorithm works is that it cycles through a grid of points, applies f a given number *numits* times to each point in the grid, and determines whether the result falls into the escape region. If it does, that point is certainly not in F_f and is given a color representing the number of iterations it took to get into the escape region. If the point has not entered the escape region by the time *numits* iterations has been done, it is considered to be in F_f and is colored black.

Here are the elements we need in order to run the algorithm to graph F_f for a given $f : \mathbb{C} \to \mathbb{C}$.

- An escape region V, which we compute beforehand using pencil and paper.
- A viewing window of the form $[a, b] \times [c, d]$, i.e. $x + iy \in \mathbb{C}$ with $a \leq x \leq b$ and $c \leq y \leq d$. The viewing window does not need to include large portions of the escape region since we already know F_f isn't in there.
- The number of points from your window you wish to sample. It's convenient to decide on a grid size *numgrid* and sample z at increments of $(b - a)/numgrid$ horizontally and $(d - c)/numgrid$ vertically.
- Computer code that takes each z from your grid and assigns it a number $j \in \{0, 1, ..., numits\}$ that is either the smallest number for which $f^j(z) \in V$, or it is simply $j = numits$ if the orbit of z does not reach the escape region in *numits* tries.
- A way to have the computer turn this $numgrid \times numgrid$ array of js into colors.

We will see how these elements are coded into Mathematica for the special case of $f(z) = z^2 + \lambda$ using the (slightly cheating) escape region $V = \{z \in \mathbb{C}$ such that $|z| > 2\}$ we computed in the previous section. Note that upping the sizes of *numgrid* and *numits* will slow the computer down, but we can manipulate them to get trustworthy images.

5.4. Exploring some more, algebraically and Mathematica-ally.

We continue our investigation of the filled Julia sets of $f(z) = z^2 + \lambda$. We have seen that when $\lambda = 0$, the filled Julia set is just the unit disc and the Julia set itself is the unit circle. The unit circle has fractal dimension 1. In our experiments on the computer we have seen that moving λ away from zero produces interesting

Julia sets that at least appear to have a larger fractal dimension. Are there any other values of λ for which the Julia set is simple enough to have fractal dimension equal to one? Where should we expect the Julia sets to lie, and for which λ is the Julia set "interesting" in some way?

EXERCISE 5.19. In this problem we start to think about where the filled Julia set is located in $\mathbb{C}$. Restrict your attention to real values of λ.

(1) The filled Julia set is never completely empty because there is always at least one element of $\mathbb{C}$ whose orbit is bounded. Find a formula in terms of λ for such a point.
(2) Find a condition on λ that guarantees that the filled Julia set contains a number that is not in $\mathbb{R}$.

EXERCISE 5.20. Let $\lambda = -2$.

(1) Find an interval of real numbers that lies in the filled Julia set.
(2) For $z = a+ib$, find a condition on a and b that guarantees that $f(z)$ lies in the escape region $V = \{z \in \mathbb{C} \text{ such that } |z| > 2\}$. (Hint: look at the size of the real part of $f(z)$ and see if the imaginary part adds enough length to get $f(z)$ into V.)
(3) Using your previous answer, make a sketch of $\mathbb{C}$ indicating all values of z whose orbits are sure to escape to infinity.
(4) In mathematica, enlarge the viewing window to include $[-2, 2] \times [-2, 2]$, and make the Julia set for as large of *numits* and *numgrid* as you need to satisfy yourself that you are seeing the true Julia set. Explain where your answer to part (3) fits in the mathematica image.

EXERCISE 5.21. Let's investigate the appearance of the Julia sets for λs in the interval $[.2, .3]$.

(1) Using a calculator or other technology (I found the calculator to be fast), compute the first several points in the orbit of 0 for $\lambda = .2, .25$, and $.3$. Make a conjecture about whether the origin is in the filled Julia set.
(2) To lend credence to your findings in the previous part it helps to think about the fixed points of f. For each case, compute them.
(3) Because λ is real, you know that $f(z) \in \mathbb{R}$ whenever $z \in \mathbb{R}$. That means that you can sketch the graph of $y = f(x)$ to try to understand the orbit of 0. Make a careful sketch of the graph for each case, and include on your sketch the line $y = x$, which should intersect at the fixed points. Use these graphs to verify or refute your conjecture that the origin is in F_f.

EXERCISE 5.22. Let's focus in on the region around $\lambda = .25$ and let mathematica compute some images.[a]

(1) Alter the mathematica code to plot the Julia sets for the the λs in the previous exercise, going up from $\lambda = .2$ in increments of .0125. What do you notice about how the Julia sets change? Experiment with *numits*. At what level of *numits* does your set of images seem to stabilize?
(2) Choose a smaller range of λs around .25 and repeat with a smaller step size. Experiment with *numits*. At what level of *numits* does your set of images seem to stabilize?
(3) Choose a teeny weeny range around .25 and repeat.

[a]The presenter(s) for this one will need to send me their images before class so I can get them onscreen for everyone.

CHAPTER 6

Mandelbrot Sets

For historical and aesthetic reasons, we have focused our attention to a parameterized class of transformations $f : \mathbb{C} \to \mathbb{C}$ given by $f(z) = z^2 + \lambda$. We will continue to do that in this section, thinking of λ as our parameter. However, it is interesting to note that the game that we are playing with the escape-time algorithm, Julia sets, and various λs, can be played with all sorts of different functions f. You are encouraged to experiment on the computer with generalizations as simple as $f(z) = z^3 + \lambda$ or as crazy as $f(z) = \lambda \sin(z) - 1$.

Mandelbrot's original question, on which he performed some of the first computer experiments, was the question: "For which values of λ is the Julia set connected?" This changes our way of thinking about Julia sets. The escape-time algorithm was able to plot the filled Julia set for a specific value of λ. This question is asking us to consider *every* $\lambda \in \mathbb{C}$, picture its Julia set, and determine if it is connected. The result is a subset $\mathcal{M} \subset \mathbb{C}$ of λs that share a common property: namely, that their Julia sets are connected.

It turns out that there is a way to understand the Mandelbrot set without needing to perform a topological analysis of the Julia set for each value of λ. What we are giving here as a definition is given elsewhere as a theorem, sometimes known as the "fundamental theorem of the Mandelbrot set" [**Fal06**, p. 225].

DEFINITION 6.1. The Mandelbrot set is the set of all $\lambda \in \mathbb{C}$ such that 0 is in the filled Julia set of $f(z) = z^2 + \lambda$. That is,

$$\mathcal{M} = \{\lambda \in \mathbb{C} \text{ such that the orbit of } 0 \text{ under } f(z) = z^2 + \lambda \text{ is bounded}\}$$

In order to begin understanding this definition you can try some values of λ to determine whether or not they are in $\mathcal{M}$ as in the next exercise. Use a calculator, earlier problems, and/or what you know about escape regions to make the computations simpler.

EXERCISE 6.2. Determine whether $\lambda \in \mathcal{M}$ for $\lambda = 0, 1, i, -2, 1 + i$, and .25.

This next example provides a general sense of where the Mandelbrot set lies in the complex plane.

EXERCISE 6.3. Prove that if $|\lambda| > 2$, then $\mathcal{O}(0)$ is unbounded by filling in the blanks on this proof.

PROOF. Let $\lambda \in \mathbb{C}$ with $|\lambda| > 2$. We know from the proof and discussion on page _____ that an escape region for $f(z) = z^2 + \lambda$ takes the form $V =$_____, where $K =$ _____. So we need only show that there is some n for which $|f^n(0)| >$_____.

We compute $f(0) =$_____ and $f^2(0) =$_____, and so $|f^2(0)| =$ _____, which simplifies to _____. We can apply the triangle inequality result of exercise _____ to see that therefore $|f^2(0)| \geq$ _____. Since $|\lambda| > 2$, this means that _____ $>$ ___ and therefore $|f^2(0)| >$ _____, proving that it is in the escape region and therefore the orbit of 0 is unbounded. □

Let us consider a sample of λs of the form $a + bi$, where a ranges from $-.5$ to .5 and b ranges from 0 to 1. Figure 1 shows the filled Julia sets for λs in this range, sampled at increments of .25 in both a and b.

Figure 1 gives us a sense of which values of λ may fall into $\mathcal{M}$. The origin, at center bottom, is definitely in. Looking from there to the right, we see (and have already proved) that .25 is in $\mathcal{M}$ and any real $\lambda > .25$ is not. We have shown that $\lambda = i$ is not in $\mathcal{M}$; its Julia set appears in the center top and looks like it might be connected, but our calculations in exercise 6.2 have proved it is not.

6.1. Using the escape-time algorithm to plot the Mandelbrot set.

We know from exercise 6.3 any λ in the Mandelbrot set lies in the circle of radius 2 around the origin. Figure 2 shows a Mathematica-generated depiction of the Mandelbrot set generated using an adaptation of the escape-time algorithm. The viewing area is $[-2.1, 2.1] \times [-2.1, 2.1]$, the Mandelbrot set is colored black, and the coloring of the exterior represents the number of iterations before the origin enters the escape region.

Many of the elements of the escape-time algorithm for plotting Julia sets are reused to plot the Mandelbrot set. Here are the ingredients:

- We can use the escape region $V = \{z \in \mathbb{C}$ such that $|z| > 2\}$ since the only λ under consideration satisfy $|\lambda| \leq 2$.
- A viewing window of the form $[a, b] \times [c, d]$, i.e. $x + iy \in \mathbb{C}$ with $a \leq x \leq b$ and $c \leq y \leq d$. This time the viewing window is for values of λ.
- We use *numits* and *numgrid* the same way as before, sampling our λs in a *numgrid* by *numgrid* array. The key difference now is that for each λ, we only need to test whether the orbit of 0 enters the escape region within *numits* iterations.
- A way to have the computer turn this *numgrid* $\times$ *numgrid* array of js into colors.

So our Mandelbrot escape-time algorithm is going to cycle through a grid of λs of our choosing. For each λ, we compute the first *numits* elements of the orbit of 0 and record when, if ever, the origin enters the escape region. If it hasn't by *numits* iterations, we proclaim λ to be in the Mandelbrot set. If not, we assign a color to λ in terms of how long the origin took to enter the escape region. As before, we will have no false negatives but may see false positives. Upping *numits* helps

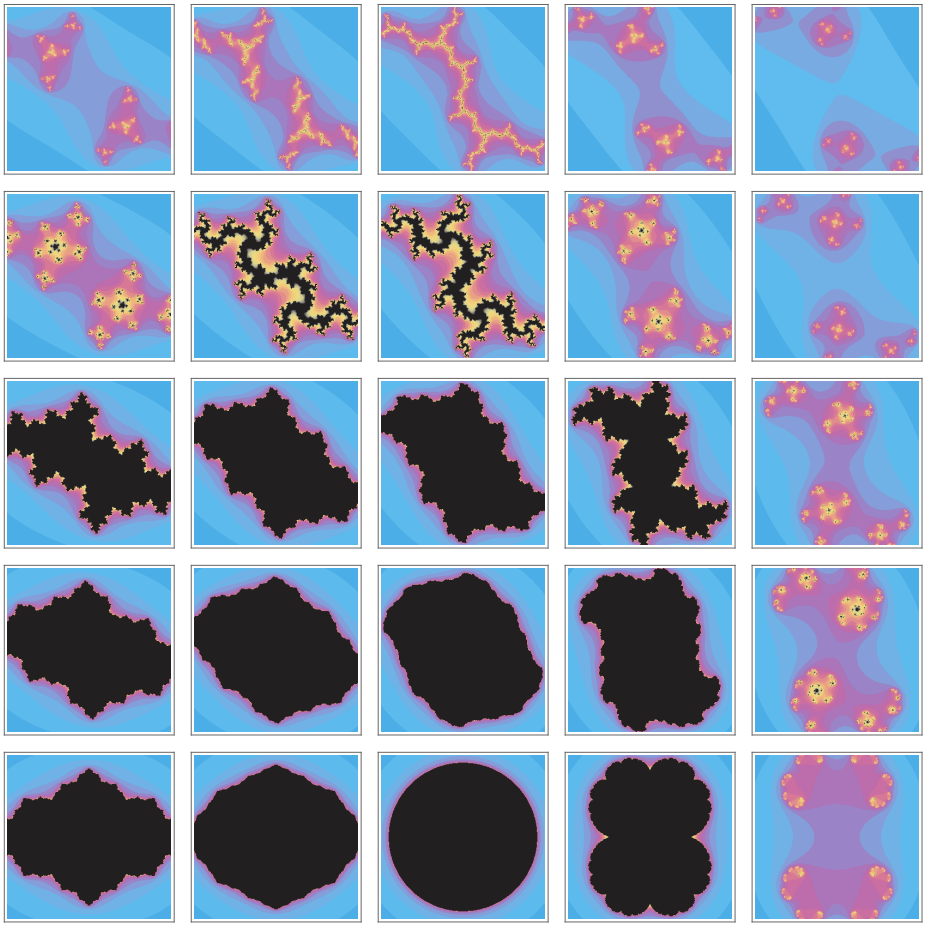


FIGURE 1. Filled Julia sets with $\lambda = -.5$ in the lower left and $\lambda = .5 + i$ in the upper right.

to eliminate the false positives and upping *numgrid* improves the resolution of the plot. We will experiment further with our Mathematica code using this algorithm.

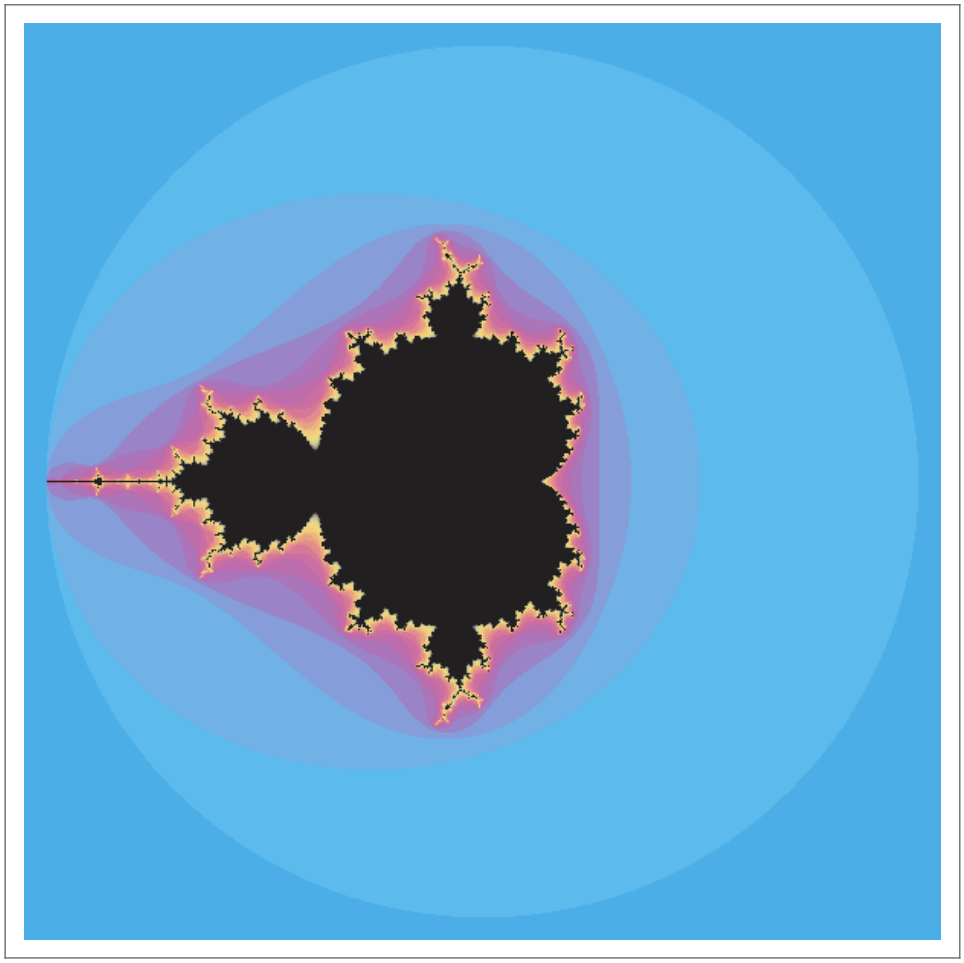


FIGURE 2. A rendering of the Mandelbrot set.

Bibliography

[Bar08] Michael F. Barnsley, *Superfractals*, Cambridge University Press, 2008. 15

[Bar12] ———, *Fractals everywhere: New edition*, Dover Publications, 2012. 15, 26, 36, 53, 55, 58

[Edg90] Gerald E. Edgar, *Measure, topology, and fractal geometry*, Undergraduate Texts in Mathematics, Springer-Verlag New York, 1990. 16, 49

[Fal06] Kenneth Falconer, *Fractal geometry: Mathematical fpoundations and applications*, John Wiley & Sons, 2006. 15, 16, 53, 58, 65

[FN] Michael Frame and Nial Neger, *Fractal and multifractal geometry: A gateway to advanced mathematics*, Unpublished lecture notes for Yale's Math 290. 46, 53, 58, 61

[Man77] B.B. Mandelbrot, *Fractals: Form, chance, and dimension*, W. H. Freeman and Co., 1977. 15

[PC09] Yakov Pesin and Vaughn Climenhaga, *Lectures on fractal geometry and dynamical systems*, Student mathematical library, vol. 52, Americal Mathematical Society, 2009. 15, 53