WINTER 2023: INTRODUCTION TO METRIC SPACES AND ITERATED FUNCTION SYSTEMS

ALEXANDER MUNDEY

Abstract. These notes are a brief introduction to metric spaces and iterated function systems compiled for a 2023 Winter Project at the University of Wollongong. They are probably full of typos and should not be distributed. If you find typos, record them and let me know.

Throughout these notes are a bunch of exercises that I encourage you to attempt. Many of them build on previous exercises. I also encourage you to work together, discuss the exercises, and explain them to each other.

1. Metric Spaces

Metrics allow us to mathematically formalise what we mean by "distance" between points. They are briefly mentioned in MATH222 but to introduce iterated function systems we will need to spend a bit of time further developing the theory.

1.1. **Metric spaces.**

Definition 1.1. A *metric* on a set X is a function $d\colon X \times X \to [0, \infty)$ satisfying the following properties:

(i) $d(x, y) = d(y, x)$ for all $x, y \in X$, (Symmetry)
(ii) $d(x, z) \leq d(x, y) + d(y, z)$ for all $x, y, z \in X$, and (Triangle Inequality)
(iii) $d(x, y) = 0$ if and only if $x = y$. (Identity of Indiscernibles)

The pair (X, d) is called a *metric space*. A metric is sometimes also called a *distance*.

There are many examples of metric spaces that you are probably already familiar with, but there are many more unfamiliar ones. The first, rather hefty, exercise below introduces some examples that we'll refer back to.

Exercise 1.2. For each of the following sets X and functions $d\colon X \times X \to [0, \infty)$ prove that (X, d) is a metric space.

(i) (Euclidean metric) Let $X = \mathbb{R}^n$ and define

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^{n} |x_i - y_i|^2}$$

for all $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ in $\mathbb{R}^n$. This is the "usual" distance on $\mathbb{R}^n$. When $n = 1$ the Euclidean metric is simply $d(x, y) = |x - y|$.

(ii) (Discrete metric) Let X be any set and for $x, y \in X$ define

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

(iii) (Uniform metric) Let $X = \mathbb{R}^n$. For $x, y \in X$ define

$$d(x, y) = \max\{|x_i - y_i|\colon 1 \leq i \leq n\}.$$

(iv) (Taxicab/Manhattan/L^1-distance) Let $X = \mathbb{R}^n$ and for $x, y, \in X$ define

$$d(x, y) = \sum_{i=1}^{n} |x_i - y_i|.$$

In the case where $n = 2$, work out why this is called the Taxicab/Manhattan metric.

(v) (Post office metric) Let $X = \mathbb{R}^2$ and for $x, y \in \mathbb{R}^2$ define

$$d(x, y) = \begin{cases} \|x\| + \|y\| & \text{if } x \neq y \\ 0 & \text{if } x = y. \end{cases}$$

Why is this called the post office metric?

(vi) (River crossing metric) Let $X = \mathbb{R}^2$ and for $x = (x_1, x_2)$ and $y = (y_1, y_2)$ define

$$d(x,y) = \begin{cases} |x_2 - y_2| & \text{if } x_1 = y_1 \\ |x_2| + |x_1 - y_1| + |y_2| & \text{if } x_1 \neq y_1. \end{cases}$$

Why is this called the river crossing metric?

(vii) (Uniform metric on functions) Let $X = C([0,1])$ be the collection of all continuous functions from $[0,1]$ to $\mathbb{R}$. For $f, g \in C([0,1])$ define

$$d(f,g) = \sup_{x \in [0,1]} |f(x) - g(x)|.$$

Why is $d(f,g)$ finite? How is it related to the uniform metric on $\mathbb{R}^n$?

(viii) (L^1-distance on functions) Let $X = C([0,1])$. For $f, g \in C([0,1])$ define

$$d(f,g) = \int_0^1 |f(x) - g(x)|\, dx$$

How is this metric related to the L^1-distance on $\mathbb{R}^n$?

(ix) (Hamming distance) This is an example which shows up in information theory. Let X be a finite set of symbols, which we call an *alphabet*. For example we could have $X = \{a, b, c\}$. For $k \in \mathbb{N}$ a *word of length* k is just an element of X^k which we typically write without parenthesis, so for example $abbaca \in \{a, b, c\}^6$. For $w \in X^k$ we write $w = w_1 w_2 \cdots w_k$. Given words $w, u \in X^k$ we define

$$d(w,u) = |\{1 \leq i \leq k \mid w_i \neq u_i\}|.$$

If w was thought of as a signal to be transmitted (say a binary string) then what is this metric measuring?

(x) (The Cantor space) Let $X = \{0,1\}^{\mathbb{N}}$. That is X is the collection of all sequences of numbers $x = (x_1, x_2, x_3, \cdots)$ where x_i is either 0 or 1 for each $i \in \mathbb{N}$. Often we get lazy and just write $x = x_1 x_2 x_3 \cdots$, with the parentheses and commas implied. Given sequences $x, y \in X$ define

$$d(x,y) = \begin{cases} 2^{1 - \min\{k \in \mathbb{N} :\, x_k \neq y_k\}} & \text{if } x \neq y \\ 0 & \text{if } x = y. \end{cases}$$

(xi) (The p-adic metric) This is an example from number theory. Let $X = \mathbb{Q}$ and let $p \in \mathbb{N}$ be a fixed prime number. Given $x \in \mathbb{Q}$ we can uniquely write $x = p^n(\frac{a}{b})$ where $n \in \mathbb{Z}$ and a and b have no common factors. We define the p-adic absolute value of x to be

$$|x|_p := p^{-n}.$$

For example if $p = 2$, then $|20|_2 = |2^2 \times 5|_2 = 2^{-2}$, while $|\frac{3}{10}|_2 = |2^{-1} \times \frac{3}{5}|_2 = 2$. The p-adic metric on $\mathbb{Q}$ is defined by

$$d(x,y) = |x - y|_p.$$

When is a number close to 0 in the p-adic metric?

(xii) (The Poincaré half-plane model of hyperbolic space) Let $X = \{z \in \mathbb{C} \mid \operatorname{Im}(z) > 0\}$ and for $z_1, z_2 \in X$ define

$$d(z_1, z_2) = 2 \operatorname{arcsinh}\Big(\frac{|z_1 - z_2|}{2\sqrt{\operatorname{Im}(z_1)\operatorname{Im}(z_2)}}\Big).$$

Compare the distance between points with the same real component as their imaginary component gets larger.

Exercise 1.3. Come up with your own examples of a metric spaces. You can let X be $\mathbb{R}^2$; or if you're feeling adventurous you could use something more exotic like words in the English language, or some kind of geometric object like a sphere. You can even build metrics on groups or graphs! (if you've seen them before).

Exercise 1.4. Prove that all metrics satisfy the *reverse triangle inequality*: that is for all $x, y, z \in X$,

$$|d(x,z) - d(z,y)| \leq d(x,y).$$

Exercise 1.5. An *ultrametric* is a metric which satisfies a stronger version of the triangle inequality. Namely that

$$d(x,z) \leq \max\{d(x,y), d(y,z)\}.$$

Which of the examples from Exercise 1.2 are ultrametrics?

Exercise 1.6. Let (X, d_1) be a metric space and fix $x_0 \in X$. Define $d_2 \colon X \times X \to [0, \infty)$ by

$$d_2(x,y) = \begin{cases} d(x,x_0) + d(x_0,y) & \text{if } x \neq y \\ 0 & \text{if } x = y \end{cases}$$

for all $x, y \in X$. Prove that d_2 is a metric.

Exercise 1.7. Let (X, d_1) and (Y, d_2) be metric spaces. Find a metric d on $X \times Y$ such that for all $x_1, x_2 \in X$ and $y_1, y_2 \in Y$,

$$d((x_1,y_1),(x_2,y_1)) = d_1(x_1,x_2) \quad \text{and} \quad d((x_1,y_1),(x_1,y_2)) = d_2(y_1,y_2).$$

Hopefully you're now convinced that there's a lot of metrics out there[1], and they each capture their own unique notion of "distance".

A fundamental idea in the theory of metric spaces is that of open and closed balls.

Definition 1.8. Let (X, d) be a metric space and let $x \in X$ and $\varepsilon > 0$. The *open ball centred at x of radius ε* is the set

$$B_\varepsilon(x) := \{y \in X \mid d(x,y) < \varepsilon\}.$$

If we want to refer to the metric we sometimes write $B^d_\varepsilon(x)$ instead. Similarly, the *closed ball centred at x of radius $\varepsilon \geq 0$* is the set

$$\overline{B_\varepsilon(x)} = \{y \in X \mid d(x,y) \leq \varepsilon\}.$$

Open balls around a point x give us a way to describe what it means to be "near" x in the sense that if $y \in B_\epsilon(x)$ for some small $\varepsilon > 0$ then $d(x,y)$ is small.

Exercise 1.9. For each of the examples in Exercise 1.2 work out $B_\varepsilon(x)$ looks like (in the case where $X = \mathbb{R}^n$ just let $n = 2$). Note that the shape of open balls can vary depending on both ε and the center of the ball x. Also try to find what open balls look like in your own examples of metric spaces.

Definition 1.10. Let (X, d) be a metric space. We say that a subset $U \subseteq X$ is *open* if for every $x \in U$ there exists $\varepsilon > 0$ such that

$$B_\varepsilon(x) \subseteq U.$$

We say that a set $C \subseteq X$ is *closed* if its compliment $X \setminus C$ is open.

[1] There's even a Canberra distance! Look it up.

Unfortunately the terminology of open and closed can be confusing as, unlike doors, "not open" does not mean "closed". There are subsets of a metric space that are *both* open and closed and also subsets that are *neither* open nor closed.

Exercise 1.11. In $\mathbb{R}^2$ with the Euclidean metric, find an example of a set with is both open and closed, and a set which is neither open nor closed.

Exercise 1.12. Prove that in a metric space the closed ball $\overline{B_\varepsilon(x)}$ is actually closed.

Exercise 1.13. Is the empty set open? Is the empty set closed?

Exercise 1.14. Let (X, d_1) be a metric space. Define $d_2 \colon X \times X \to \mathbb{R}$ by

$$d_2(x, y) = \frac{d_1(x, y)}{1 + d_1(x, y)}.$$

Show that d_2 is also a metric satisfying the following:

(i) $d_2(x, y) \leq 1$ for all $x, y \in X$,
(ii) for every $\varepsilon_1 > 0$ and $x \in X$ there exists $\varepsilon_2 > 0$ such that $B^{d_2}_{\varepsilon_2}(x) \subseteq B^{d_1}_{\varepsilon_1}(x)$, and
(iii) for every $\varepsilon_2 > 0$ and $x \in X$ there exists $\varepsilon_1 > 0$ such that $B^{d_1}_{\varepsilon_1}(x) \subseteq B^{d_2}_{\varepsilon_2}(x)$.

The last two conditions can be interpreted as saying that points are close with respect to d_1 if and only if they're close with respect to d_2.

1.2. **Sequences.** Metric spaces provide a suitable setting to talk about sequences. We define what it means for a sequence to converge in a metric space analogously to how we define convergence in $\mathbb{R}$.

Definition 1.15. Let (X, d) be a metric space. A *sequence* in a metric space is just a function $x \colon \mathbb{N} \to X$. Instead of writing $x(n)$ we write x_n and denote the sequence by $(x_n)_{n=1}^\infty$. We say that $(x_n)_{n=1}^\infty$ *converges* to a point $x \in X$, if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_n, x) < \varepsilon$ for all $n \geq N$. In this case we say $x_n \to x$ as $n \to \infty$, or write $\lim_{n\to\infty} x_n = x$.

Remark 1.16. Convergence can also be rephrased in terms of open balls. A sequence $(x_n)_{n=1}^\infty$ converges to x if for every open ball of radius $\varepsilon > 0$ about x, there is some $N \in \mathbb{N}$ such that $x_n \in B_\varepsilon(x)$ for all $n \geq N$.[2]

Exercise 1.17. Convince yourself that the definition of convergence in metric spaces agrees with what you know about convergence of sequences in $\mathbb{R}$ with the Euclidean distance.

What it means for a sequence to converge very much depends on the choice of metric. Two different metrics on $\mathbb{R}^2$ for example, can have different convergent sequences. In a typical metric space, convergence behaviour can be quite unintuitive.

Exercise 1.18. Let $X = \mathbb{R}^2$ and let d be the river crossing metric. Prove that the sequence $x_n = (\frac{1}{n}, 0)$ converges but the sequence $y_n = (\frac{1}{n}, 1)$ does not converge.

[2]This version of convergence generalises to topological spaces: an abstraction of metric spaces.

Exercise 1.19. Let $X = \{0,1\}^{\mathbb{N}}$ be the Cantor space with its associated metric d. Consider the sequence

$$\begin{aligned} x_1 &= 1111111\cdots \\ x_2 &= 0111111\cdots \\ x_3 &= 0011111\cdots \\ x_4 &= 0001111\cdots \\ x_5 &= 0000111\cdots \\ &\vdots \end{aligned}$$

Prove that $(x_n)_{n=1}^{\infty}$ converges and find its limit.

Exercise 1.20. Let d_2 be the Euclidean metric on $\mathbb{R}^2$. Find a metric d_1 on $\mathbb{R}^2$ and a sequence $(x_n)_{n=1}^{\infty}$ such that x_n converges with respect to d_1, but does not converge with respect to d_2.

Exercise 1.21. Let $X = \mathbb{R}^2$. What are some non-trivial examples of convergent sequences with respect to the discrete metric?

Exercise 1.22. Let (X, d) be a metric space. Prove that $C \subseteq X$ is closed if and only if for any convergent sequence $x_n \to x$ with $x_n \in C$ we have $x \in C$.

We can also make sense of Cauchy sequences in metric spaces.

Definition 1.23. Let (X, d) be a metric space. A sequence $(x_n)_{n=1}^{\infty}$ is said to be *Cauchy* if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(x_m, x_n) < \varepsilon$ for all $m, n \geq N$.

Exercise 1.24. Prove that every convergent sequence in a metric space (X, d) is also Cauchy.

In general, it is not true that every Cauchy sequence in a metric space converges, for example this is not true of the Euclidean distance on $\mathbb{Q}$. Many important theorems of analysis rely on the fact that Cauchy sequences converge, so we give metric spaces with this property a name.

Definition 1.25. A metric space (X, d) is said to be *complete* if every Cauchy sequence converges.

Exercise 1.26. Which of the metric spaces from Exercise 1.2 are complete?[3]

1.3. **Functions between metric spaces.** Since metric spaces are just sets with a distance, we can consider functions between them. Continuity can be defined for functions between metric spaces, and we can also introduce some slightly more refined metric-dependent concepts.

Definition 1.27. Let (X_1, d_1) and (X_2, d_2) be metric spaces. A function from (X_1, d_1) to (X_2, d_2) is denoted by $f\colon X_1 \to X_2$ or $f\colon (X_1, d_1) \to (X_2, d_2)$ if we want to specify the metrics. We say that $f\colon X_1 \to X_2$ is

(i) *continuous at* $x \in X_1$ if for every $\varepsilon > 0$ there exists $\delta > 0$ such that whenever $y \in X_1$ satisfies
$$d_1(y, x) < \delta \quad \text{then} \quad d_2(f(y), f(x)) < \varepsilon,$$
(ii) *continuous* if f is continuous at each $x \in X_1$,
(iii) *isometric* or *an isometry* if $d_1(x, y) = d_2(f(x), f(y))$ for all $x, y \in X_1$,

[3]Some of them are very non-trivially incomplete, don't waste too much time on it.

(iv) *contractive* or *a contraction* if there exists $0 \leq C < 1$ such that

$$d_2(f(x), f(y)) \leq C d_1(x, y)$$

for all $x, y \in X_1$. The supremum of all such C is sometimes called the *contractivity ratio* of the contraction.

Often we care about the case where $(X_1, d_1) = (X_2, d_2)$ in which case we would say that f is continuous/isometric/contractive on X_1.

Intuitively, a function f is continuous at x if for any point z in the range of f near $f(x)$ we can find a y near x such that $z = f(y)$. Or more crudely put, a function is continuous if "nearby stuff in the range of f comes from nearby stuff in the domain".

A function is isometric if it preserves the distance between points, and a function is contractive if points get closer together after applying the function. These are both special types of continuous functions. Contractions will be important when we come to talk about iterated function systems.

Exercise 1.28. Prove that if f is either isometric or contractive, then f is continuous.

Exercise 1.29. Prove that if f is isometric, then f is one-to-one.

Exercise 1.30. Consider the function $\mathrm{id}\colon \mathbb{R}^2 \to \mathbb{R}^2$ defined by $\mathrm{id}(x) = x$ for all $x \in \mathbb{R}^2$. This function is often called the *identity function*. Equip $\mathbb{R}^2$ with two different metrics so that $\mathrm{id}\colon (\mathbb{R}^2, d_1) \to (\mathbb{R}^2, d_2)$. For different choices of d_1 and d_2, examine whether id is isometric, continuous, or a contraction.

Exercise 1.31. Write down some examples of contractions on $\mathbb{R}^2$ with the Euclidean metric. Are your examples also contractions with respect to the taxicab metric or the river crossing metric?

Exercise 1.32. Prove the "sequential characterisation of continuity" for metric spaces. That is prove that a function $f\colon (X_1, d_1) \to (X_2, d_2)$ is continuous at $x \in X_1$ if and only if for every sequence $x_n \to x$ we have $f(x_n) \to f(x)$.

Exercise 1.33. Let $X = \{0,1\}^{\mathbb{N}}$ be the Cantor space with its metric d. Define a map $\sigma_0\colon (X, d) \to (X, d)$ by

$$\sigma_0(x_1 x_2 x_3 \cdots) = 0 x_1 x_2 x_3 \cdots$$

Prove that σ_0 is contractive. What is the contractivity ratio of σ_0?

Exercise 1.34. Let $X = \mathbb{Q}$ with the 2-adic metric. Define $f_2\colon (\mathbb{Q}, d) \to (\mathbb{Q}, d)$ by $f_2(x) = 2x$ for $x \in \mathbb{Q}$. Prove that f_2 is contractive. What is its contractivity ratio? Is f_2 contractive with respect to the Euclidean metric?

1.4. **The Banach Fixed-Point Theorem.** The first, and most fundamental, result we'll need when we get to iterated function systems is the Banach Fixed-Point Theorem. It guarantees that contraction mappings on complete metric spaces always have a unique fixed-point.

Theorem 1.35 (The Banach Fixed-Point Theorem)**.** *Let (X, d) be a non-empty complete metric space and let $f\colon X \to X$ be a contraction mapping. Then f has a unique fixed point x^*: that is there exists a unique point $x^* \in X$ such that*

$$f(x^*) = x^*.$$

Moreover, for any element $x_1 \in X$, the sequence defined by $x_n = f(x_{n-1})$ converges to x^*.

Remark 1.36. Although its easy to overlook, the second part of the Banach Fixed-Point Theorem is fairly incredible from a practical standpoint. It says that if you want to actually find the fixed-point of a contraction, you can start with **any** point $x_1 \in X$ and just repeatedly apply f to it in order to approximate the fixed point x^*.

Exercise 1.37. Prove the Banach Fixed-Point Theorem. (Hint: completeness is your friend.)

The Banach Fixed-Point Theorem has uses all over the world of numerical analysis, but also plays huge a role in existence theorems like the Picard-Lindelöf Theorem for first order ODE. More importantly for us, its one of the key ingredients in proving Hutchinson's Theorem, which we aim to prove in Section 2.

Exercise 1.38. Equip $\mathbb{R}^2$ with the Euclidean metric. Find the fixed point of the function $f\colon \mathbb{R}^2 \to \mathbb{R}^2$ given by $f(x,y) = \frac{1}{2}(x+y, x-y)$.

Exercise 1.39. Find the fixed-point of the function σ_0 of Exercise 1.33.

1.5. **Compactness.** The last metric space concept we'll need is the idea of compactness. Compactness is a fairly subtle concept. Roughly speaking, compactness is a measure of "finiteness" of a set, but not in the sense that the set is actually has finite cardinality. Think of it more along the lines of when people down at the pub say that a sphere is "finite" whereas a plane is "infinite". Those pub-goers would be more correct to say "compact" and "not compact".[4]

Another analogy is that compactness can be thought of as an infinite version of the Pidgeonhole principle (you'll see why in the exercises below).

Definition 1.40. Let (X,d) be a metric space. A subset $K \subseteq X$ is said to be *compact* if for every sequence $(x_n)_{n=1}^{\infty}$ with $x_n \in K$ there is a subsequence $(x_{n_k})_{k=1}^{\infty}$ which converges in K. If X itself is compact, then we say that (X,d) is a *compact metric space.*

In the definition of compactness, the phrase "converges in K" is important. For example in $\mathbb{R}$ with the Euclidean distance, the set $(0,1)$ is not compact. To see why, note that the sequence $x_n = \frac{1}{n}$ converges in $\mathbb{R}$ but it does not converge to an element of $(0,1)$.

Example 1.41. In $\mathbb{R}$ with the Euclidean distance, closed intervals $[a,b]$ are examples of compact sets. This is a statement of the Bolzano-Weierstrass Theorem.

Many of the results from MATH222, like the Extreme-Value Theorem, actually rely on compactness going on behind the scene.

Exercise 1.42. Let $X = \{1,2,3,4\}$ which we equip with the discrete metric. Prove that any subset $K \subseteq X$ is compact. (Hint: Coo! coo!)

Exercise 1.43. Find a subset of $\mathbb{R}^2$ with the Euclidean distance that is closed but not compact?

Exercise 1.44. Characterise compact subsets of $\mathbb{R}$ with the discrete metric.

Exercise 1.45. Let $K_1, \dots, K_N$ be a collection of compact sets in a metric space (X,d). Prove that $\bigcup_{i=1}^{N} K_i$ is compact. Is the same result true for an infinite collection of compact sets?

In $\mathbb{R}^n$ with the Euclidean distance compactness is characterised completely by the Heine-Borel Theorem (don't try to prove it now).

[4]I don't recommend correcting them!

Theorem 1.46 (The Heine-Borel Theorem)**.** *A subset $K \subseteq \mathbb{R}^n$ is compact with respect to the Euclidean distance if and only if K is closed and bounded (that is there exists $M \geq 0$ such that $d(0,x) < M$ for all $x \in K$.)*

Unfortunately, the Heine-Borel Theorem does not apply in general. For example, in $C([0,1])$ with the uniform metric the closed ball $B_1(0)$ is closed and bounded but not compact (this is proved in MATH222). The Arzelà–Ascoli Theorem instead characterises compact sets in $C([0,1])$. In general, it can be difficult characterise compact sets.

Continuity plays nicely with compactness.

Exercise 1.47. Prove that if $K \subset X_1$ is compact and $f\colon (X_1,d_1) \to (X_2,d_2)$ is continuous, then $f(K)$ is also compact.

Although the Heine-Borel Theorem does not hold in general, compact sets in metric spaces are always closed.

Exercise 1.48. Prove that if $K \subseteq X$ is compact and $(x_n)_{n=1}^{\infty}$ is a convergent sequence with $x_n \in K$ and $x_n \to x$, then $x \in K$. Deduce that K is closed.

Exercise 1.49. Prove that $X = \{0,1\}^{\mathbb{N}}$ with its associated metric is a compact metric space. Using previous exercises, deduce that the set

$$X_0 := \{0x_1x_2x_3\cdots \mid x_1x_2x_2\cdots \in X\}$$

is also compact.

Exercise 1.50. Is the empty set compact?

The final compactness result we will need is the Extreme-Value Theorem.

Theorem 1.51 (The Extreme-Value Theorem)**.** *Suppose that (X,d) is a metric space and that $K \subset X$ is compact. Let $f\colon K \to \mathbb{R}$ be a continuous function. Then there exists x_1 and x_2 in K such that*

$$f(x_1) = \sup_{x\in K} f(x) \qquad \textit{and} \qquad f(x_2) = \inf_{x\in K} f(x).$$

Exercise 1.52. Prove the Extreme-Value Theorem using previous exercises.

2. Fractals and Iterated Function Systems

The term *fractal* was coined by Benoit Mandelbrot in 1975, who gave the now-famous example of the *Mandelbrot set*. Mandelbrot noticed that fractals appear everywhere from coastlines to stock prices, and instigated the study of fractal geometry. The precise definition of what makes a fractal a fractal is hotly contended. Roughly speaking a fractal is subset of a metric space that exhibits two features:

- It can be assigned a non-integer notion of dimension (this is where the *frac-* in fractal comes from).
- A degree of *self-similarity*. That is, the set is made up of parts that look approximately like the set itself.

A few years later, in 1981, Australian mathematician John Hutchinson formulated a systematic way of producing fractals using what have come to be known as *iterated function systems* (a term coined by fellow fractal pioneer and pseudo-Aussie Michael Barnsley[5]).

2.1. **Iterated function systems.**

Definition 2.1. An *iterated function system* or just *IFS* on a metric space (X, d) is a finite collection of functions $\Gamma = \{\gamma_1, \dots, \gamma_N\}$ on X such that $\gamma_i \colon X \to X$ is a contraction for each $1 \leq i \leq N$.

Given an iterated function system $\Gamma = \{\gamma_1, \dots, \gamma_N\}$ on (X, d), Hutchinson introduced the following function on subsets $S \subseteq X$ now called the *Hutchinson operator*,

$$H_\Gamma(S) = \bigcup_{i=1}^{N} \gamma_i(S),$$

where $\gamma_i(S) = \{\gamma_i(x) \colon x \in S\}$. If $\mathcal{P}(X) = \{S \colon S \subseteq X\}$ denotes the power set of X, then we may think of the Hutchinson operator as a function $H_\Gamma \colon \mathcal{P}(X) \to \mathcal{P}(X)$. The Hutchinson operator applies each function in our iterated function system to S and unions together the results to yield a new subset of X.

Exercise 2.2. Prove that if K is compact then $H_\Gamma(K)$ is also compact.

The fundamental theorem of iterated function systems was proved by Hutchinson in his 1981 paper [Hut81]. He proved that to an iterated function system is a unique associated set that is invariant under the Hutchinson operator.

Theorem 2.3 (Hutchinson's Theorem)**.** *Let* $\Gamma = \{\gamma_1, \dots, \gamma_N\}$ *is an iterated function system on a* complete *metric space* (X, d)*. Then there is a unique non-empty compact subset* $\mathbb{A} \subseteq X$ *such that*

$$\mathbb{A} = \bigcup_{i=1}^{N} \gamma_i(\mathbb{A}).$$

Moreover, for any non-empty compact subset $K_0 \subseteq X$ *we can define a sequence of sets by* $K_n := H_\Gamma(K_{n-1})$ *for all* $n \geq 1$*. Then* $(K_n)_{n=1}^{\infty}$ *converges to* $\mathbb{A}$ *in an appropriate sense.*

Definition 2.4. The set $\mathbb{A}$ of Hutchinson's Theorem is called the *attractor* of the iterated function system. The name is due to the second part of the theorem.

---

[5]The fern on the front cover is named after him!

As you will see in the exercises below, the attractor of an iterated function system is often a fractal. Consequently, iterated function systems provide a fairly easy way of generating interesting fractals on all sorts of metric spaces.

You may have noticed in the statement of Hutchinson's Theorem that the notion of convergence of sets was left intentionally unclear. We're going to expend a bit of effort trying to fix that. However, before we do, lets look at some examples.

Example 2.5 (The Middle-Thirds Cantor Set). Let $X = \mathbb{R}$ with the Euclidean metric, which we know is complete. Define $\Gamma = \{\gamma_1, \gamma_2\}$, where $\gamma_1(x) = \frac{x}{3}$ and $\gamma_2(x) = \frac{x}{3} + \frac{2}{3}$ for $x \in \mathbb{R}$. Both γ_1 and γ_2 are contractions (check), so we have an iterated function system.

Let $K_0 = [0, 1]$. Applying the Hutchinson operator H_Γ to K_0 we see that

$$K_1 = H_\Gamma(K_0) = [0, 1/3] \sqcup [2/3, 1],$$

and

$$K_2 = H_\Gamma(K_1) = [0, 1/9] \sqcup [2/9, 1/3] \sqcup [2/3, 7/9] \sqcup [8/9, 1].$$

Further iterations are pictured in Figure 1.

By Hutchinson's Theorem the sequence of set $(K_n)_{n=1}^\infty$ converges to an attractor $\mathbb{A}$. The attractor of this particular iterated function system is called the *middle-thirds Cantor set*.[6] The middle-thirds Cantor set is an example of a compact subset of $\mathbb{R}$ that is *totally-disconnected*.

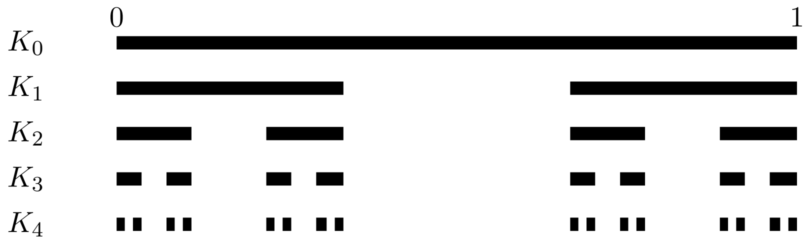


FIGURE 1. The result of applying the Hutchinson operator from Example 2.5 to $K_0 = [0, 1]$.

Exercise 2.6. Hutchinson's Theorem says that it doesn't matter which compact set $K_0 \subseteq \mathbb{R}$ we start with. Let $K_0 = \{0\}$ and repeatedly apply the Hutchinson operator from Example 2.5 to see what happens. Also, try starting with other choices of K_0, like maybe $[-1, 0]$.

Exercise 2.7. Let $X = \mathbb{R}$ with the Euclidean metric and define $\Gamma = \{\gamma_1, \gamma_2\}$ by $\gamma_1(x) = \frac{x}{2}$ and $\gamma_2(x) = \frac{x}{2} + \frac{1}{2}$. Show that Γ defines an iterated function system and find the attractor. How does it compare to Example 2.5?

Exercise 2.8 (Sierpinski gasket)**.** Let $X = \mathbb{R}^2$ with the Euclidean metric and let $\Gamma = \{\gamma_1, \gamma_2, \gamma_3\}$ where

$$\gamma_1(x_1, x_2) = \left(\frac{x_1}{2}, \frac{x_2}{2}\right)$$
$$\gamma_1(x_1, x_2) = \left(\frac{x_1}{2}, \frac{x_2 + 1}{2}\right)$$
$$\gamma_3(x_1, x_2) = \left(\frac{2x_1 + 1}{4}, \frac{2x_2 + \sqrt{3}}{4}\right).$$

[6]You may wonder what this "Cantor set" has to do with the "Cantor space" defined earlier. It turns out there's a sense in which they are actually the same. Feel free to ask me about it.

Show that Γ defines an iterated function system and determine what $\mathbb{A}$ looks like. In the process you should work you what each of the maps are doing "geometrically". Also remember that you can choose any K_0 you like. You might find better choices of K_0 as you experiment.

Exercise 2.9 (Koch curve)**.** Let $X = \mathbb{R}^2$ with the Euclidean metric, and let $\Gamma = \{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ be given by

$$\gamma_1(x_1, x_2) = \frac{1}{3}(x_1, x_2) \qquad \gamma_2(x_1, x_2) = \frac{1}{6}(x_1 - \sqrt{3}x_2 + 2, \sqrt{3}x_1 + x_2)$$
$$\gamma_3(x_1, x_2) = \frac{1}{6}(x_1 + \sqrt{3}x_2 + 3, -\sqrt{3}x_1 + x_2 + \sqrt{3}) \quad \gamma_4(x_1, x_2) = \frac{1}{3}(x_1 + 2, x_2).$$

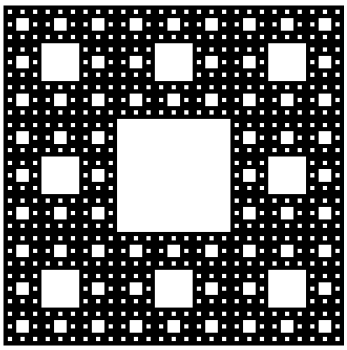
Work out what each of the maps does "geometrically" and find the attractor $\mathbb{A}$.

Exercise 2.10. Let $X = \{0, 1\}^{\mathbb{N}}$ be the Cantor space with its associated metric. Define $\Gamma = \{\sigma_0, \sigma_1\}$ by

$$\sigma_0(x_1x_2x_3\cdots) = 0x_1x_2x_3\cdots \quad \text{and} \quad \sigma_1(x_1x_2x_3\cdots) = 1x_1x_2x_3\cdots$$

Show that Γ is defines an iterated function system and find its attractor.

Exercise 2.11. Find an iterated function system such that the attractor is the following:



Exercise 2.12. Find an example of an iterated function system on $\mathbb{R}^2$ with the river crossing metric.

Exercise 2.13. Experiment with building iterated functions systems. The weirder the metric space, the better.

Exercise 2.14 (If you have time)**.** Devise a way in which you could use Hutchinson's Theorem to plot fractals on a computer in a programming language/maths package of your choice.

2.2. **Proving Hutchinson's Theorem.** You may have noticed that the statement of Hutchinson's Theorem is vaguely reminiscent of the Banach Fixed-Point Theorem, and indeed we're going to use it. The question is, on which metric space are we applying the Banach Fixed-Point Theorem? We start by introducing the space.

Definition 2.15. Let (X, d) be a metric space. The *hyperspace*[7] of (X, d) is the set

$$\mathbb{H}(X, d) := \{K \subseteq X \mid K \text{ is non-empty and compact}\}.$$

Note that the points in $\mathbb{H}(X, d)$ are precisely the non-empty compact subsets of X. It's a bit weird thinking of sets as points and can take some getting used to.

Exercise 2.16. Let $X = \mathbb{Z}$ with the discrete metric d. What is the hyperspace $\mathbb{H}(\mathbb{Z}, d)$?

[7] Queue sci-fi jokes.

The strategy for proving Hutchinson's Theorem is the following:

(1) Find a metric d_H on the set $\mathbb{H}(X,d)$ which makes $(\mathbb{H}(X,d),d_H)$ into a complete metric space.
(2) Show that the Hutchinson operator H_Γ defines a contraction on $(\mathbb{H}(X,d),d_H)$.
(3) Apply the Banach Fixed-Point Theorem to H_Γ to prove Hutchinson's Theorem.

We're going to expend a bit of effort trying to construct an appropriate metric d_H on $\mathbb{H}(X,d)$. Ideally, our metric d_H should respect the original metric d on X in the sense that for all $x,y \in X$,

$$d_H(\{x\},\{y\}) = d(x,y). \qquad (\dagger)$$

The following exercises are designed to get you to the definition of d_H on your own.[8]

Exercise 2.17. Let (X,d) be a metric space. We'll start by defining d_H between a point in X (thought of as a singleton set) and a set. Fix $x \in X$ and let K be a compact subset of X. Find a sensible definition of $d_H(\{x\},K)$. Ensure that if $x,y \in X$, then $(\dagger)$ holds and also

$$d_H(\{x\},K) \leq d_H(\{x\},\{y\}) + d_H(\{y\},K).$$

Draw pictures and think about the Euclidean metric on $\mathbb{R}^2$ when trying to come up with a definition.

Exercise 2.18. Now suppose that $\{x_1,\dots,x_N\}$ is a finite set of points in X. Find a sensible definition of $d_H(\{x_1,\dots,x_N\},K)$ that extends your previous definition. Ensure that if $\{y_1,\dots,y_M\}$ is another finite set of points in X, then

$$d_H(\{x_1,\dots,x_N\},K) \leq d_H(\{x_1,\dots,x_N\},\{y_1,\dots,y_M\}) + d_H(\{y_1,\dots,y_M\},K).$$

Remember that you also want d_H to be symmetric and satisfy the Identity of Indiscernibles on finite sets.

Exercise 2.19. Now suppose that K_1 and K_2 are two compact subsets of X. Come up with a sensible definition of $d_H(K_1,K_2)$ and check that d_H defines a metric on $\mathbb{H}(X,d)$.

Since d_H was designed to make $(\dagger)$ hold, the map $\iota\colon (X,d) \to (\mathbb{H}(X,d),d_H)$ defined by $\iota(x) = \{x\}$ is an isometry. In other words we can think of the original metric space (X,d) as embedded inside its hyperspace $\mathbb{H}(X,d)$. The hyperspace can be BIG!

We're now on the final stretch of proving Hutchinson's Theorem. We have a metric space $(\mathbb{H}(X,d),d_H)$ on which we can think about the Hutchinson operator, but there's still two crucial hypotheses of the Banach Fixed-Point Theorem remaining. Completeness of $(\mathbb{H}(X,d),d_H)$ and contractivity of the Hutchinson operator $H_\Gamma\colon \mathbb{H}(X,d) \to \mathbb{H}(X,d)$.

Exercise 2.20. Prove that if (X,d) is a complete metric space then $(\mathbb{H}(X,d),d_H)$ is also a complete metric space.

Exercise 2.21. Prove that the Hutchinson operator $H_\Gamma\colon \mathbb{H}(X,d) \to \mathbb{H}(X,d)$ is contractive with respect to d_H.

Proving Hutchinson's Theorem turns out to be a bit of a task, but you're now (hopefully) pretty much there.

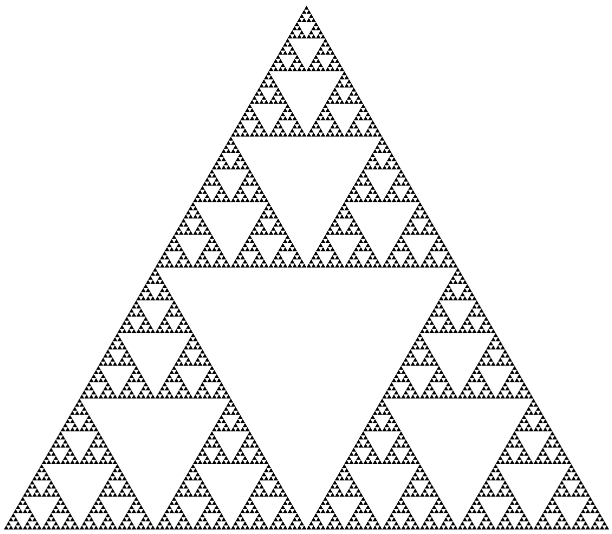
Exercise 2.22. Assemble the previous exercises into a proof of Hutchinson's Theorem. Once complete, give yourself a pat on the back.

[8]In mathematical research the correct definition is often the last thing you actually figure out!

3. Directions to go from here

If you've gotten this far and want to go further there are many directions in which to go. Ask me about any of the following:

- The code map. A map from the cantor space onto the attractor any iterated function system which "encodes" the iterated function system. The code map generalises the idea of decimal expansions for real numbers.
- Dimensions. Fractals have fractional dimension. What does this mean?
- Completions. Given an incomplete metric space, we can turn it into a complete one by adding in all the limits of Cauchy sequences. This is how you build $\mathbb{R}$ from $\mathbb{Q}$.
- The artistic route. Continue to generate pretty pictures and coming up with interesting iterated function systems. Work out how to make the Barnsley fern on the cover of these notes, or ask me about how to make iterated function systems in PowerPoint.
- Anything else! I've included some references at the end for further reading.



References

[Bar93] M. Barnsley, "Fractals Everywhere", 2nd ed., *Academic Press Professional, Boston, MA* (1993) ISBN: 0-12-079061-0

[Edg08] G. Edgar, "Measure, topology, and fractal geometry", 2nd ed., Undergraduate Texts in Mathematics, *Springer, New York* (2008) ISBN: 978-0-387-74748-4 28-01

[Fal14] K. Falconer, "Fractal geometry. Mathematical foundations and applications.", 3rd ed., *John Wiley & Sons Ltd., Chichester* (2014) ISBN: 978-1-119-94239-9

[Hut81] J.E. Hutchinson, *Fractals and Self Similarity*, Indiana University Mathematics Journal, **30**, p713-747 (1981)

[Kig01] J. Kigami, "Analysis on Fractals", *Cambridge University Press, Cambridge* (2001) ISBN: 0-521-79321-1