Fractals and Chaos

Semester 1, 2001–2002

Contents

CHAPTER 1

Fractal Dimension

One of the main ideas that comes from the study of fractal sets is that some of our usual notions of geometry must be re-assessed. In standard geometry, we prefer to focus our attention on smooth (i.e. C^∞) objects. However it is known that, for example, the set of smooth functions on an interval is a subset of measure zero in the space of all continuous functions. In a similar way, when studying of differential equations, one tends to concentrate on systems of equations which can be solved explicitly in terms of known functions, although one can show that this is a set of measure zero in the set of all differential equations.

In fractal theory, we often study objects which would usually be looked on as pathological. So, for example, we will consider figures which are non-differentiable, often with infinitely fine structure. To study these different types of geometry we use different tools from those we would use for ordinary geometry. One of the main ideas that will be of use to us is the idea of scaling.

1. Scaling symmetry

One of the most famous figures of fractal theory is the Mandelbrot set, which we will study in Section 3 (see Figure 1). One of the properties that this figure has (which we will see in the

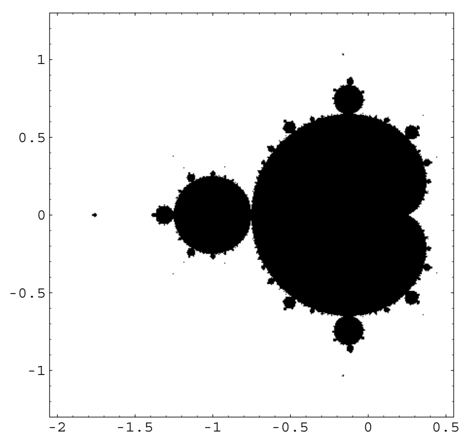


Figure 1. The Mandelbrot Set

computer labs) is that if we zoom in on particular regions of the figure, we discover approximate copies of the figure itself. These are *approximate scaling symmetries* of the Mandelbrot set. We now consider some simpler figures with exact scaling symmetries.

2. The (Middle Third) Cantor Set

We construct the Cantor set C from the unit interval in the real line by an iterative procedure. Let $E_0 = [0, 1] \subset \mathbb{R}$. We define an operation O which acts on (unions of) closed intervals:

$$O(I) := I - \{\text{open middle third of } I\}.$$

For example

$$O([0,1]) = [0,1] - \left(\frac{1}{3}, \frac{2}{3}\right)$$
$$= \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right].$$

Given this operation, we define

$$E_n := O(E_{n-1}), \qquad n = 1, 2, 3, \ldots.$$

Each E_n is a closed subset of the interval $[0,1]$ so, for example,

$$E_1 = \left[0, \frac{1}{3}\right] \cup \left[\frac{2}{3}, 1\right],$$
$$E_2 = O(E_1) = \left[0, \frac{1}{9}\right] \cup \left[\frac{2}{9}, \frac{1}{3}\right] \cup \left[\frac{2}{3}, \frac{7}{9}\right] \cup \left[\frac{8}{9}, 1\right].$$

Thus the first few steps look like:

Definition 1.1. The *Cantor set*, C, is $\bigcap_{n\,=\,0}^{\infty} E_k$.

Scaling symmetry. The Cantor set has an exact scaling symmetry. The set $C \cap \left[0, \frac{1}{3}\right]$ consists of the interval $\left[0, \frac{1}{3}\right]$ with the open middle third of every closed sub-interval removed. If we scale this set up by a factor of 3, we get the interval $[0,1]$ with the open middle third of every closed subinterval removed, which is simply the original Cantor set C. Therefore the Cantor set C is identical to the subset $C \cap \left[0, \frac{1}{3}\right]$ scaled up by a factor of 3. A similar remark applies to the set $C \cap \left[\frac{2}{3}, 1\right]$.

Zero length. Note that

$$\text{length}(E_n) = \frac{2}{3}\,\text{length}(E_{n-1}),$$

so, by induction, we see that

$$\text{length}(E_n) = \left(\frac{2}{3}\right)^n,$$

since $\text{length}(E_0) = 1$. Therefore

$$\text{length}(C) = \lim_{n\,\to\,\infty} \text{length}(E_n) = \lim_{n\,\to\,\infty} \left(\frac{2}{3}\right)^n = 0.$$

Therefore, as a line segment, the Cantor set is of zero length.

Triadic expansions and countability. Given $x \in [0,1]$, there exist integers $a_1, a_2, \ldots$ equal to $0, 1$ or 2 with the property that we can write

$$x = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \ldots.$$

We will write this as $x = 0.a_1 a_2 a_3 \ldots$ (it should be obvious from the context when such an expression is base 3 or base 10). This expression is the *triadic expansion* of x. We can give an alternative description of the Cantor set in terms of base-3 expansions of numbers (See **Question Sheets**):

$$C = \{x \in [0,1] : \text{ there exists an expansion } x = 0.a_1a_2a_3\dots, \text{ with } a_1, a_2, \dots \neq 1\}.$$

DEFINITION 1.2. A set S is *countable* if (a). it contains a finite number of elements, or (b). there exists a $1-1$, onto mapping ϕ from the natural numbers $\mathbb{N}$ to the elements of S. Otherwise S is *uncountable*.

PROPOSITION 1.1. *The Cantor set is uncountable.*

PROOF. Assume the Cantor set is countable. It is therefore possible to list its elements (i.e. put the elements into $1-1$ correspondence with the natural numbers or a finite set of natural numbers) so we can form a list of the triadic expansions of the numbers above:

$$\begin{array}{ll} n \in \mathbb{N} & x \in C \\ 1 & 0.a_1a_2a_3a_4\dots \\ 2 & 0.b_1b_2b_3b_4\dots \\ 3 & 0.c_1c_2c_3c_4\dots \\ 4 & 0.d_1d_2d_3d_4\dots \\ \dots & \dots \end{array} \tag{1.1}$$

where each of the $a_i, b_i, c_i, d_i, \dots$ are equal to with 0 or 2. By assumption this list contains every number that is a member of the Cantor set. However, suppose we consider the number X, with triadic expansion $0.\tilde{a}_1\tilde{b}_2\tilde{c}_3\tilde{d}_4\dots$ where we define

$$\tilde{a}_1 = 0 \text{ if } a_1 = 2$$
$$\tilde{a}_1 = 2 \text{ if } a_1 = 0,$$

and similarly for $\tilde{b}_2, \tilde{c}_3, \tilde{d}_4, \dots$. Then X is a member of the Cantor set, since its triadic expansion contains only 0's and 2's. However, it follows from the definition of X that it is not in the list of elements of the Cantor set given in equation (1.1). For example, X cannot be the first element on the list $(0.a_1a_2a_3a_4\dots)$ because by construction $\tilde{a}_1 \neq a_1$, so the triadic expansion of X differs from that of the first element on the list at the first digit of its triadic expansion. Similarly X cannot be the second element of the Cantor set on the list $(0.b_1b_2b_3b_4)$ because $\tilde{b}_2 \neq b_2$, so the triadic expansions disagree at the second digit. Continuing this process, X cannot be the i'th element on the list, because the i'th digit of the triadic expansion of X will differ from that of the i'th element on the list.

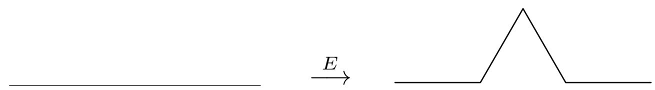
We thus have a contradiction, and our original assumption that the members of the Cantor set can be put into $1-1$ correspondence with the natural numbers must be false. Therefore the Cantor set is uncountable. □

We therefore have three important properties of the Cantor set:

- It is uncountable
- It is of zero length
- It has scaling symmetries

3. The Koch curve

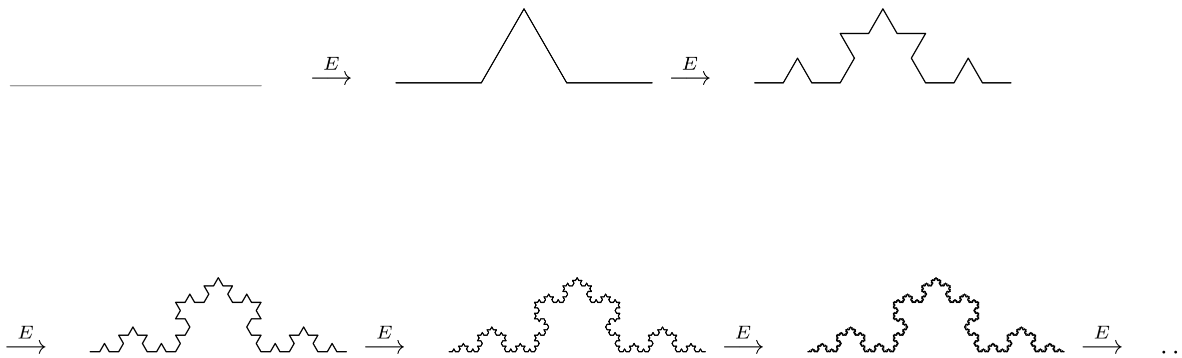
The Koch curve K is a subset of the plane again constructed iteratively. We begin with the unit interval along the x-axis: $E_0 = \{(x,y) \in \mathbb{R}^2 : x \in [0,1], y = 0\}$. We define an operation O on closed line segments which removes the middle third of the line and replaces it with two copies of the same length as follows:



As with the Cantor set, we now define the iterates

$$E_n := O(E_{n-1}), \qquad n = 1, 2, 3, \ldots.$$

The first few iterates therefore look like:



The Koch curve is defined as $\lim_{k\to\infty} E_k$. Since the length of E_n is $(4/3)^n$, the length of the Koch curve is $\lim_{n\to\infty}(4/3)^n = \infty$. Thus the Koch curve is of infinite length. However, considered as a subset of the plane, it is a line segment of zero area.

In a similar fashion to the Cantor set, we also see that the Koch curve has scaling symmetries. If we consider, for example, the left-hand part of the Koch curve $L := \{(x, y) \in K : x \in [0, 1/3]\}$, then we see that if we scale this set up by a factor of 3, we recover the full Koch curve.

Therefore the Koch curve has important properties analogous to those of the Cantor set:

- It is of infinite length
- It is of zero area
- It has scaling symmetries

4. Similarity dimension

We argued above that the Cantor set C is identical to the set $C \cap [0, 1/3]$ scaled up by a factor of 3. A similar argument shows that C is identical to the set $C \cap [2/3, 1]$ scaled up by a factor of 3 as well. Reversing this argument, we see that if we take the set C and scale it down by a factor of 3, we get the set $C \cap [0, 1/3]$, which is identical to the set $C \cap [2/3, 1]$. We therefore see that we can look on C as being composed as the union of two copies of itself scaled by a factor of 1/3. We say that the Cantor set is *self-similar*:

DEFINITION 1.3. An set is *self-similar* if it can be constructed as the union of N copies of itself, scaled by a factor of $1/r$.

We can use this idea of self-similarity to generalise the idea of dimension to fractal objects. Our usual idea of dimension comes from vector spaces, where the dimension of the vector space is defined to be the number of linearly independent vectors required to span the space. For example, consider the real line $\mathbb{R}$:

This vector space is spanned by one unit vector $\mathbf{x}$, so we say that the real line has dimension 1.

Similarly, consider the plane $\mathbb{R}^2$:

To span this space, we require two unit vectors $\mathbf{x}$ and $\mathbf{y}$, so we say that the plane has dimension 2. This is the usual geometrical definition of dimension.

One can, however, give alternative definitions of dimension which agree with the geometrical definition in the standard cases given above, but which in more general circumstances have non-integer value. The simplest alternative definition of dimension is for objects which are self-similar, and is called the *similarity dimension*.

For example, consider the square of side l in the plane:

If we consider scaling this square by a factor of $1/2$, then we end up with a square of side $l/2$. If we take 4 copies of this new square, then by translating 3 of the squares, we can reform the original square:

Therefore we can look on the original square as being the union of 4 copies of itself scaled by a factor of $1/2$. Note that this procedure is not unique, since we could equally well reform the square taking 9 copies of itself scaled by a factor $1/3$, or 16 copies scaled by $1/4$, more generally n^2 copies scaled by a factor $1/n$ for any positive integer n.

In a similar fashion, consider the interval $[0, l] \subset \mathbb{R}$. If we scale this interval by a factor of $1/2$, we get the interval $[0, l/2]$. Taking 2 copies of this interval, and translating the second copy along the real line by a distance $l/2$, we recover the original interval:

More generally, we can look on the interval $[0, l]$ as the union of n copies of scaled by a factor $1/n$ for any positive integer n.

Note that in each of the above cases, we need n^d copies of the original object scaled by a factor of $1/n$, where d is the usual geometrical dimension of the object we are considering (so $d = 2$ for the square and $d = 1$ for the interval). This is a reflection of the fact that if scale an object in d-dimensions by an overall factor of A, then we expect the d-dimensional volume of the object to scale as A^d. For example, if we take the interval $[0, l]$, which is of length l, and scale it by a factor A we get the interval $[0, Al]$, which is of length $Al = A^1 \cdot l$. If we take the square of side l, of area l^2 and scale if by a factor A, we get the square of length Al which has area $(Al)^2 = A^2 \cdot l^2$. If we take the d-dimensional cube of side l (i.e. $\{(x_1, \ldots x_d) \in \mathbb{R}^d : 0 \leq x_1, \ldots, x_d \leq l\}$) it has volume l^d, scaled by a factor A, we get the cube of side Al, which has volume $(Al)^d = A^d \cdot l^d$.

We can turn these geometrical properties around and use them to define the dimension of an object:

Definition 1.4. Given a self-similar set S which can be constructed as the union of N copies of itself scaled by a factor of $1/r$, the *similarity dimension* of the set S is defined to be:

$$d_s(S) = -\frac{\log N}{\log(1/r)}.$$

So for the interval $[0, l]$, the interval is made up of n copies of itself scaled by $1/n$, so in this case $N = r = n$, so

$$d_S = -\frac{\log n}{\log(1/n)} = 1,$$

so the similarity dimension coincides with the geometrical dimension. Similarly, the square in the plane is made up of n^2 copies of itself scaled by $1/n$ so $N = n^2, r = n$ and the similarity dimension is:

$$d_S = -\frac{\log n^2}{\log(1/n)} = 2.$$

In the case of the d-dimensional cube, we need n^d copies of the cube scaled by a factor $1/n$, so the similarity dimension is

$$d_S = -\frac{\log n^d}{\log(1/n)} = d.$$

Therefore, in each of these cases the similarity dimension is the same as the geometrical dimension. However, this is not the case for more general objects.

The Cantor set. The Cantor set is self-similar, since the sets $C \cap \left[0, \frac{1}{3}\right]$ and $C \cap \left[\frac{2}{3}, 1\right]$ are geometrically similar to C scaled by a factor of $\frac{1}{3}$. Therefore C is the union of two copies of itself, scaled by a factor of $\frac{1}{3}$. Thus we can associate with the Cantor set a similarity dimension of

$$d_S(C) = -\frac{\log 2}{\log(1/3)} = \frac{\log 2}{\log 3}.$$

The similarity dimension of the Cantor set is not an integer, and $0 < d_S(C) < 1$. This is consistent with the intuitive idea that C has zero length (so we expect its dimension to be less than 1), but it contains uncountably many points (so we expect its dimension to be greater than 0).

The Koch curve. The Koch curve is self-similar, and can be constructed as the union of 4 copies of itself scaled by a factor of 1/3. Thus we can associate with it a similarity dimension:

$$d_S(K) = -\frac{\log 4}{\log(1/3)} = \frac{\log 4}{\log 3}.$$

Again this is non-integral, and $1 < d_S(K) < 2$ consistent with the intuitive idea that the Koch curve is of infinite length (so expect dimension greater than 1) but of zero area (so expect dimension less than 2).

5. More General Definitions of Dimension

The concept of similarity dimension only really makes sense for objects that are self-similar. There are, however, more general definitions of fractal dimension which can be defined for more general sets.

DEFINITION 1.5. A *metric space* (X, d) is a set X and a function $d : X \times X \to [0, \infty)$ which obeys:

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

The elements of the set X will be called *points* and the function d will be referred to as a *metric*.

EXAMPLE 1.1. $(\mathbb{R}^n, d)$ is a metric space, where $\mathbb{R}^n$ is defined as the set of ordered n-tuples of real numbers $(x_1, \ldots, x_n)$, and the metric is defined by

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

where $\mathbf{x} = (x_1, \ldots, x_n)$ and $\mathbf{y} = (y_1, \ldots, y_n)$.

DEFINITION 1.6. Let A be a closed, bounded subset of a metric space (X, d). For each $\epsilon > 0$, let

$N(A, \epsilon) =$ the minimum number of closed balls of radius ϵ required to cover A.

Then we define the *Fractal Dimension* of A, denoted $D(A)$, to be

$$D(A) = \lim_{\epsilon \to 0} \frac{\log N(A, \epsilon)}{\log(1/\epsilon)}$$

if this limit exists.

THEOREM 1.1. *Let A be a closed, bounded subset of a metric space (X, d). Let $\epsilon_n = C\lambda^n, n = 1, 2, 3, \ldots,$ for real numbers $0 < \lambda < 1$ and $C > 0$. If*

$$D = \lim_{n \to \infty} \frac{\log N(A, \epsilon_n)}{\log \frac{1}{\epsilon_n}}$$

exists, then $D(A) = D$.

REMARK. This means that when we calculate the fractal dimension of a set, we can replace the continuous variable ϵ by a suitably chosen discrete variable in the calculations.

PROOF. Let $\epsilon > 0$. There exists an integer n with the property that

$$\epsilon_{n+1} \leq \epsilon \leq \epsilon_n,$$

which implies that

$$\frac{1}{\epsilon_n} \leq \frac{1}{\epsilon} \leq \frac{1}{\epsilon_{n+1}}. \tag{1.2}$$

From the definition of $N(A, \epsilon)$ it follows that

$$N(A, \epsilon_n) \leq N(A, \epsilon) \leq N(A, \epsilon_{n+1}). \tag{1.3}$$

We now take the log of Equations (1.2) and (1.3). Using the fact that log is an increasing function we deduce that

$$\log N(A, \epsilon_n) \leq \log N(A, \epsilon) \leq \log N(A, \epsilon_{n+1}).$$

and

$$\log \frac{1}{\epsilon_{n+1}} \geq \log \frac{1}{\epsilon} \geq \log \frac{1}{\epsilon_n}.$$

Dividing the first equation by the second, we find that

$$\frac{\log N(A, \epsilon_n)}{\log 1/\epsilon_{n+1}} \leq \frac{\log N(A, \epsilon)}{\log 1/\epsilon} \leq \frac{N(A, \epsilon_{n+1})}{\log 1/\epsilon_n}. \tag{1.4}$$

Using the fact that $\epsilon_{n+1} = C\lambda^{n+1} = \lambda\epsilon_n$, we see that the left hand expression equals

$$\text{l.h.s.} = \frac{\log N(A, \epsilon_n)}{\log 1/\lambda + \log 1/\epsilon_n},$$

and similarly the right hand side equals

$$\text{r.h.s.} = \frac{N(A, \epsilon_{n+1})}{\log \lambda + \log 1/\epsilon_{n+1}},$$

Therefore

$$\begin{aligned}
\lim_{\epsilon\to 0}(\text{l.h.s.}) &= \lim_{\epsilon\to 0} \frac{\log N(A, \epsilon_n)}{\log 1/\lambda + \log 1/\epsilon_n} \\
&= \lim_{n\to\infty} \frac{\log N(A, \epsilon_n)}{\log 1/\lambda + \log 1/\epsilon_n} \\
&= \lim_{n\to\infty} \frac{\log N(A, \epsilon_n)}{\log 1/\epsilon_n} \\
&= D,
\end{aligned}$$

and similarly

$$\begin{aligned}
\lim_{\epsilon\to 0}(\text{r.h.s.}) &= \lim_{\epsilon\to 0} \frac{\log N(A, \epsilon_{n+1})}{\log \lambda + \log 1/\epsilon_{n+1}} \\
&= \lim_{n\to\infty} \frac{\log N(A, \epsilon_{n+1})}{\log \lambda + \log 1/\epsilon_{n+1}} \\
&= \lim_{n\to\infty} \frac{\log N(A, \epsilon_{n+1})}{\log 1/\epsilon_{n+1}} \\
&= \lim_{n\to\infty} \frac{\log N(A, \epsilon_n)}{\log 1/\epsilon_n} \\
&= D.
\end{aligned}$$

Therefore, taking $\lim_{\epsilon\to 0}$ of Equation (1.4), we deduce that

$$D \leq D(A) \leq D,$$

so $D(A) = D$. □

EXAMPLE 1.2. Suppose a set S is self-similar, composed of the union of N copies of itself scaled by a factor of $1/r$. Assume that it takes $N(S, \epsilon_0)$ closed balls of radius $\epsilon_0 = C$ to cover S. Then it will require the same number of closed balls of radius ϵ_0/r in order to cover S scaled by a factor of $1/r$. Therefore, let $\lambda := 1/r$. Since S is constructed from N copies of itself scaled by $1/r$, then we must have

$$N(S, \epsilon_0/r) = N \cdot N(S, \epsilon_0).$$

In other words

$$N(S, \epsilon_1) = N \cdot N(S, \epsilon_0),$$

where $\epsilon_1 = \lambda\epsilon_0$. A similar argument implies that

$$N(S,\epsilon_n) = N \cdot N(S,\epsilon_{n-1}) = \cdots = N^n N(S,\epsilon_0),$$

where

$$\epsilon_n = \left(\frac{1}{r}\right)^n \epsilon_0.$$

Therefore, using Theorem 1.1, we see that

$$\begin{aligned} D(S) &= \lim_{n\to\infty} \frac{n\log N + \log N(S,\epsilon_0)}{-n\log(1/r) - \log\epsilon_0} \\ &= \lim_{n\to\infty} \frac{n\log N}{-n\log(1/r)} \\ &= -\frac{\log N}{\log 1/r} \\ &= d_s(S) \end{aligned}$$

Therefore, for a self-similar set S, the fractal dimension reduces to the similarity dimension.

6. Fractals

We have now encountered the idea of fractal dimension, and the idea that certain figures can have non-integer fractal dimension. Although there is no concencus on the definition of a fractal, the following (due to Mandelbrot) is the working definition we will take

Definition 1.7. A set S is a *fractal set* if its fractal dimension strictly exceeds its geometrical dimension.

For example, according to this definition the Cantor set is a fractal set since it has geometrical dimension 0 (it is a collection of points) but its fractal dimension is $d_s(C) = \log 2/\log 3 > 0$. Similarly, the Koch curve is a fractal since it is a line segment and therefore of geometrical dimension 1, but, as we have seen above, its fractal dimension is strictly greater than 1.

There are alternative definitions of fractal set. Although the technical details of these definitions may vary, they usually imply that fractal sets have the following properties:

(1) infinitely fine structure, locally looking very complicated
(2) defined by an iterative/recursive procedure
(3) exhibit approximate scaling symmetries

We will return to the ideas of fractal dimension later, especially when discussing Iterated Function Systems.

CHAPTER 2

Iteration of Functions (Dynamical Systems)

1. Fixed Points and Cycles

DEFINITION 2.1. Let X be a set and $f : X \to X$ a mapping. For each positive integer k, we define the *k'th iterate*, $f^k : X \to X$ iteratively by

$$f^0(x) = x, \qquad \forall x \in X$$

and

$$f^{(k+1)}(x) = f \circ f^k(x) = f(f^k(x)), \qquad k \geq 0.$$

If f has an inverse, f^{-1}, we may define

$$f^{-k}(x) = (f^{-1})^k(x).$$

We have the basic law:

$$f^k \circ f^l = f^l \circ f^k = f^{(k+l)}$$

which holds for all non-negative integers k and l, and for all integers k and l if f^{-1} exists.

Terminology. The pair (X, f) is sometimes referred to as a (discrete) dynamical system.

EXAMPLE 2.1. Let $X = \mathbb{R}$ and $f(x) = x^2$. Then $f^k(x) = x^{2^k}, k = 0, 1, \ldots$. Since f is not $1-1$, there does not exist an inverse f^{-1}, so we cannot extend this formula to $k < 0$. If we take $X = [0, \infty)$ and $f(x) = x^2$, then $f^k(x) = x^{2^k}, \forall k \in \mathbb{Z}$.

EXAMPLE 2.2. Let $X = \mathbb{R}$ and $f(x) = x^3$. Then $f^k(x) = x^{3^k}, k = 0, 1, \ldots$. In this case, $f^{-1}(x) = x^{1/3}$ is the inverse of f, and we generally find that $f^k(x) = x^{3^k}$ for all $k \in \mathbb{Z}$.

EXAMPLE 2.3. Let $X = [0, 1]$ and $f(x) = 2x(1-x)$. (This is a particular case of the Logistic map which we will study later.) The first few iterates of f are then

$$\begin{aligned}
f^0(x) &= x, \\
f^1(x) &= 2x(1-x), \\
f^2(x) &= f(f^1(x)) = 2f^1(x)\left(1 - f^1(x)\right) = 4x(1-x)(1-2x(1-x)),
\end{aligned}$$

and in general $f^k(x)$ is a polynomial of degree 2^k in x for each positive integer k. Since the map f is not $1-1$, there is no well-defined f^{-1}.

DEFINITION 2.2. Let X be a set and $f : X \to X$ a mapping. A point $a \in X$ is a *fixed point* of f if $f(a) = a$.

EXAMPLE 2.4. Let $X = \mathbb{R}$ and $f(x) = x^2$. Fixed points, a, are solutions of $a = f(a) = a^2$, so the only fixed points are $a = 0, 1$.

DEFINITION 2.3. Let (X, d) be a metric space and $a \in X$ be a fixed point of $f : X \to X$. The point a is an *attracting/attractive* fixed point if there exists a $c < 1$ and a $\delta > 0$ such that $d(f(x), a) \leq c d(x, a)$ for every $x \in X$ such that $d(x, a) < \delta$. It is a *repelling/repulsive* fixed point if there exists a $c > 1$ and a $\delta > 0$ such that $d(f(x), a) \geq c d(x, a)$ for every $x \in X$ such that $d(x, a) < \delta$.

EXAMPLE 2.5. Let $X = \mathbb{R}, f(x) = kx$, with k a positive real number and $d(x, y) = |x - y|$. Then if $k \neq 1$, f has a unique fixed point at 0 and

$$d(a, f(x)) = k|x| = kd(a, x), \qquad \forall x \in \mathbb{R}.$$

Letting $c = \frac{1}{2}(k + 1)$, we deduce that the fixed point at 0 is attractive if $k < 1$ and repulsive if $k > 1$.

In the case $k = 1$ every point in the real line is a fixed point of f and $d(x, f(y)) = d(x, y), \forall x, y \in \mathbb{R}$. Such fixed points, which are neither strictly attractive nor repulsive are referred to as *neutral* or *indifferent* fixed points.

EXAMPLE 2.6. Let $X = \mathbb{R}, f(x) = x^2$, with $d(x, y) = |x - y|$. If we consider the fixed point at 0 then $d(0, f(x)) = |f(x)| = x^2 = |x|d(0, x)$, so if we take any $\delta < 1$, and $c = \delta$ then $d(0, f(x)) \leq cd(0, x), \forall x$ with $d(0, x) < \delta$. Thus 0 is an attractive fixed point.

If we consider the fixed point at 1, then $d(1, f(x)) = |x^2 - 1| = |x + 1||x - 1| = |x + 1|d(1, x)$. If $d(1, x) = |x - 1| < \delta$, then we deduce that $1 - \delta < x < 1 + \delta$, so $2 - \delta < x + 1 < 2 + \delta$. Thus, $d(1, f(x)) = |x + 1|d(1, x) > (2 - \delta)d(1, x)$. Therefore if we choose any $\delta < 1$, and let $c = 2 - \delta > 1$, then $d(1, f(x)) \geq cd(1, x)$ for all x with $d(1, x) < \delta$. Therefore the fixed point at 1 is repulsive.

DEFINITION 2.4. Let X be a set and $f : X \to X$ a map. The *orbit* of a point $x \in X$ is the set $\{x, f(x), f^2(x), f^3(x) \dots\}$.

Notation. We will often start the iterations at $x = x_0$ and denote the higher iterates by $x_i := f^i(x_0)$ so, for example, $x_1 = f(x_0), x_2 = f(f(x_0))$ and so on. Equivalently, $x_n = f(x_{n-1}), n \geq 1$.

DEFINITION 2.5. Let X be a set and $f : X \to X$ a map. The point $x \in X$ is a *periodic point of period* n if $f^n(x) = x$, where n is an integer and $n \geq 1$. If n is the smallest integer for which this equality holds, then $x \in X$ is a *periodic point of prime period* n. In this case, the set $\{x, f(x), f^2(x), \dots, f^{n-1}(x)\}$ is called a *cycle of order* n.

A periodic point of f of order n is attractive if it is an attractive fixed point of f^n. A cycle of period n, C, is attractive if it contains an attractive periodic point of order n. We define repulsive periodic points and cycles in an analogous fashion.

LEMMA 2.1. *Let x be a periodic point of order n of function f. Then $f^{kn}(x) = x$ for all positive integers k.*

PROOF.

$$\begin{aligned}
f^{kn}(x) &= f^n(f^n(\dots f^n(x))) && (k \text{ times}) \\
&= f^n(f^n(\dots f^n(x))) && (k-1 \text{ times, since } x \text{ a periodic point}) \\
&= \dots \\
&= f^n(x) \\
&= x.
\end{aligned}$$

□

DEFINITION 2.6. Let X be a set and $f : X \to X$ a map. A set $Y \subseteq X$ is an *invariant set* of f if $f(Y) \subseteq Y$. (i.e. $\forall y \in Y, f(y) \in Y$.)

EXAMPLE 2.7. $X = \mathbb{C}, f(z) = z^2$. In this case, the unit circle (i.e. $\{z \in \mathbb{C} : |z| = 1\}$ is an invariant set. The non-zero periodic points are those of the form $e^{2\pi i\theta}$, where $\theta = \frac{m}{2^n - 1}$ for positive integers m and n. The points on the unit circle with finite orbits are those for which θ is rational. (**See Question Sheets**).

EXAMPLE 2.8. If we use the Newton-Raphson method to approximate the roots of the quadratic equation $x^2 - k = 0$, where we assume $k > 0$ and $x \in (0, \infty)$, this leads us to consider the map

$$f : (0, \infty) \to (0, \infty) : x \mapsto f(x) = \frac{1}{2}\left(x + \frac{k}{x}\right). \tag{2.1}$$

If we choose any $x_0 > 0$ then define

$$x_n = f(x_{n-1}), \qquad n = 1, 2, \ldots.$$

If $\{x_n\}$ converges to a limit x then

$$x = f(x) = \frac{1}{2}\left(x + \frac{k}{x}\right),$$

with solutions

$$x = \pm\sqrt{k}.$$

For example, if we take $x_0 = 2, k = 2$ we find that

$$x_1 = \frac{3}{2}$$
$$x_2 = \frac{17}{12}$$
$$x_3 = \frac{577}{408} \sim 1.4142 \qquad \text{(giving } \sqrt{2} \text{ correct to 4 decimal places).}$$

To understand why this process works, we consider iterating the function $f(x)$ given in equation (2.1). As a function of x, f is monotonically decreasing on $(0, \sqrt{k})$ and monotonically increasing on $(\sqrt{k}, \infty)$. We also have that

$$f(x) < x \Longleftrightarrow x > \sqrt{k}$$

and

$$f(x) > x \Longleftrightarrow x < \sqrt{k}.$$

If we suppose $x_0 > \sqrt{k}$ then

$$x_1 = f(x_0) < x_0,$$

and

$$x_1 = f(x_0) > f(\sqrt{k}) = \sqrt{k},$$

so we know that

$$\sqrt{k} < x_1 < x_0.$$

Repeating this process:

$$\sqrt{k} < x_n < x_{n-1} < \cdots < x_2 < x_1 < x_0.$$

The sequence $\{x_1, \ldots, x_n\}$ is decreasing and bounded below, and as such must converge to some limit x. However, since $\lim_{n\to\infty} f^n(x_0) = \lim_{n\to\infty} f^{n+1}(x_0) = x$, we deduce that x must be a fixed point of f, so we must have $x = \sqrt{k}$. (**See Question Sheets**).

2. Web diagrams

In the special case where $X \subseteq \mathbb{R}$, there is a geometrical construction for studying the orbit $\{x_n = f^n(x_0)\}_{n=1}^{\infty}$ of a given starting point $x_0 \in X$, called a *web diagram*.

Assuming $X \subseteq \mathbb{R}$, we begin by drawing the plane $\{(x, y) : x \in X, -\infty < y < \infty\}$ (in practice, we choose some sensible range of y values), and draw the graphs $y = f(x)$ and $y = x$. If we wish to consider the orbits of initial point x_0, we begin in the diagram at the point $(x_0, 0)$. We take the vertical line through this point, and continue it to the point where it cuts the curve $y = f(x)$, namely $(x_0, f(x_0)) = (x_0, x_1)$. We connect this point with a horizontal line to the point (x_1, x_1) on the line $y = x$. From there we draw the vertical line to cut $y = f(x)$ at (x_1, x_2), from which we draw the horizontal line to cut $y = x$ at (x_2, x_2), and continue this process to infinity. The orbit of the point x_0 can be deduced from the sequence of points, $(x_1, x_1), (x_2, x_2), \ldots$ which occur on the diagonal $y = x$.

Note that if the function f has a fixed point at a, then the curves $y = f(x)$ and $y = x$ will intersect at the point (a, a). Conversely, if the curves $y = f(x)$ and $y = x$ intersect at a point (a, a), then a will be a fixed point of the function f. If a fixed point is attractive then a path in the web diagram starting at $(x_0, 0)$ will converge to the point (a, a) for all x_0 on a suitable neighbourhood of a. Similarly, if a fixed point is repulsive a path starting at $(x_0, 0)$ will diverge away from the point (a, a) for all x_0 on a suitable neighbourhood of a.

If the function f has a cycle of period n, then there will exist closed paths in the web diagram. The number of distinct points on the diagonal curve $y = x$ of such a curve will equal the order n of the cycle. Similarly to above, a cycle is attractive if paths in the web diagram starting on a neighbourhood of the cycle converge towards it, and is repulsive if paths diverge away from it.

In the case where $f(x) = (x + k/x)/2$, the web diagram for an initial point $x_0 > \sqrt{k}$ will be qualitatively the same as that in Figure 1.

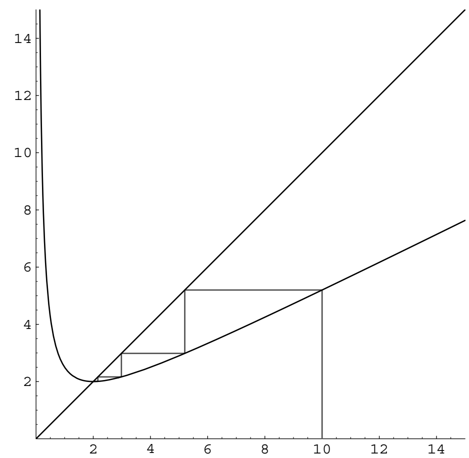


FIGURE 1. Web diagram for iteration of $f(x) = (x + k/x)/2$ with $k = 4$, $x_0 = 10$

THEOREM 2.1. *Consider $f : X \to X$, where $X \subseteq \mathbb{R}$ or $X \subseteq \mathbb{C}$, with $d(x, y) = |x - y|$. If a is a fixed point of f and f is a differentiable function, then*

a). $|f'(a)| < 1 \Longrightarrow$ *a is an attractive fixed point*
b). $|f'(a)| > 1 \Longrightarrow$ *a is a repulsive fixed point.*

PROOF. We will prove the first statement, the second follows by reversing the inequalities in the appropriate places.

From the definition of $f'(a)$, we know that for all $\epsilon > 0$, there exists a $\delta > 0$ such that for all x with $|x - a| < \delta$ we have

$$\left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| < \epsilon.$$

Since $|f'(a)| < 1$, we can choose

$$\epsilon = \frac{1}{2} \left(1 - |f'(a)|\right) > 0,$$

then there exists a $\delta > 0$ such that for all x with $|x - a| < \delta$ we have

$$\left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| < \frac{1}{2} \left(1 - |f'(a)|\right).$$

By the triangle inequality, we know that

$$\left| \frac{f(x) - f(a)}{x - a} \right| \leq \left| \frac{f(x) - f(a)}{x - a} - f'(a) \right| + |f'(a)|.$$

Therefore, for all x with $|x-a|<\delta$ we have

$$\begin{aligned}\left|\frac{f(x)-f(a)}{x-a}\right| &\le \left|\frac{f(x)-f(a)}{x-a}-f'(a)\right|+|f'(a)| \\ &< \frac{1}{2}\left(1-|f'(a)|\right)+|f'(a)| \\ &= \frac{1}{2}\left(1+|f'(a)|\right).\end{aligned}$$

Therefore we let $c := \frac{1}{2}\left(1+|f'(a)|\right)$, and note that $c<1$. Multiplying the last equation through by $|x-a|$ and using the fact that $f(a)=a$, we find that, for all x with $|x-a|<\delta$

$$|f(x)-a| < c|x-a|.$$

Therefore the fixed point is attractive. □

COROLLARY 2.1. *Let $X \subseteq \mathbb{R}$ or $X \subseteq \mathbb{C}$ with their standard metrics and let $f : X \to X$ be differentiable. Then a cycle of period n, C, is an attracting cycle if $|(f^n)'(x)| < 1$ for some $x \in C$.*

Aside. Note that if a cycle C consists of periodic points $(x_1, \dots, x_n)$ then

$$(f^n)'(x_1) = f'(x_1)f'(x_2)\dots f'(x_n).$$

As such $|(f^n)'(x_1)| = |(f^n)'(x_2)| = \cdots = |(f^n)'(x_n)|$. Therefore, the value of $|(f^n)'(x)|$ is the same whatever periodic point x we choose on the cycle C. It is therefore consistent to speak of the cycle itself as being attractive or repulsive. (**See Questions Sheets**)

EXAMPLE 2.9. $X = (0,\infty), f(x) = \frac{1}{2}(x+\frac{a}{x})$. There is a unique fixed point $\sqrt{a}$. This is an attractive fixed point. There are no other periodic points. Some invariant sets: X, $\{\sqrt{a}\}$, $[\sqrt{a},\infty)$.

THEOREM 2.2. *Let (X,d) be a metric space, with $d(x,y) = |x-y|$, $f : X \to X$ a mapping, and a a fixed point of f. Suppose that there is a set N such that $a \in N$, $f(N) \subseteq N$ and $\forall x \in N$ we have $|f(x)-a| \le c|x-a|$ for some $0 \le c < 1$. Then, for any $x_0 \in N$, $f^n(x_0) \to a$ as $n \to \infty$.*

PROOF.

$$|f^n(x_0)-a| = |f(f^{n-1}(x_0))-a| \le c|f^{n-1}(x_0)-a| = c|f(f^{n-2}(x_0))-a| \le \cdots \le c^{n-1}|x_0-a|.$$

We have used the fact that if $x_0 \in N$ then $f^n(x_0) \in N$, which means we can apply the inequality $|f(x)-a| \le c|x-a|$ with x in turn being $f^{n-1}(x_0), f^{n-2}(x_0), \dots, f(x_0)$. Since $c<1$, we therefore deduce that as $n \to \infty$, we have $|f^n(x_0)-a| \le c^{n-1}|x_0-a| \to 0$. □

COROLLARY 2.2. *If $f : X \to X$ is a differentiable function, where $X \subseteq \mathbb{R}$, or $X \subseteq \mathbb{C}$, and if a is a fixed point of f for which $|f'(a)| < 1$, then a has a neighbourhood N such that $f^n(x) \to a$ for every $x \in N$.*

3. The Logistic Map

We let $X = [0,1] \subset \mathbb{R}$, with $d(x,y) = |x-y|$. The logistic map is defined by iterating the function:

$$f(x) = \lambda x\,(1-x)\,,$$

where $\lambda \in [1,4]$ is a real parameter. We wish to study the attractive orbits of this system as the value of λ increases from $\lambda = 1$. Note that

$$f : [0,1] \to \left[0, \frac{\lambda}{4}\right],$$

so that $\lambda = 4$ is the maximal value of λ for which the unit interval is mapped into itself. The map is $2-1$, as illustrated by the graph of f (see Figure 2).

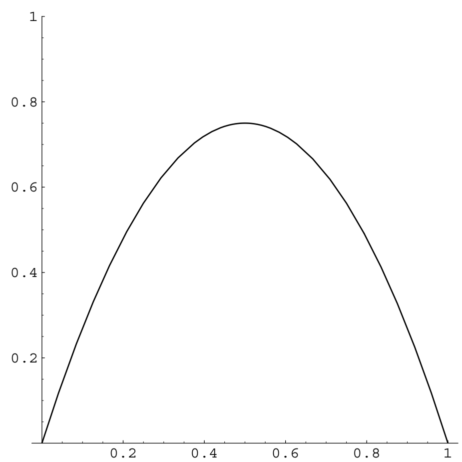


FIGURE 2. The graph of $f(x) = \lambda x\,(1-x)$

First consider fixed points of f. Therefore we look for solutions a of $f(a) = a = \lambda a\,(1-a)$. This means we have fixed points at

$$a = 0, 1 - \frac{1}{\lambda}.$$

To decide whether these fixed points are attractive or repulsive, we consider

$$|f'(a)| = \lambda |1-2a|,$$

which is < 1 for an attractive fixed point and > 1 for a repulsive fixed point. In particular, for the fixed points of the logistic map we have:

$$|f'(0)| = \lambda,$$
$$\left| f'\left(1 - \frac{1}{\lambda}\right)\right| = |2-\lambda|.$$

We now consider various ranges of λ:

$\boxed{\lambda = 1}$ In this case, there is a unique fixed point of f at $a = 0$. In this case $|f'(0)| = 1$ and the fixed point is indifferent. (It turns out that orbits are attracted to 0 but the convergence is extremely slow.)

$\boxed{1 < \lambda < 3}$ For each value of λ in this range,

$$|f'(0)| = \lambda > 1,$$
$$\left| f'\left(1 - \frac{1}{\lambda}\right)\right| = |2-\lambda| < 1.$$

Therefore $\forall \lambda \in (1,3)$, the fixed point at 0 is repulsive and the fixed point at $1 - 1/\lambda$ is attractive.

In the special case when $\lambda = 2$, we have an attractive fixed point at $1/2$ and $f'(1/2) = 0$. Such behaviour is sometimes referred to as *super-attractive*. In this case, the convergence to the fixed point is extremely rapid. If $x = 1/2 + \epsilon$, where ϵ is a small parameter, one can prove by induction that

$$f^n(x) = \frac{1}{2} - 2^{2^n - 1}\epsilon^{2^n}, \qquad n = 1, 2, \ldots,$$

so we have quadratic convergence to the fixed point (See Figure 3).

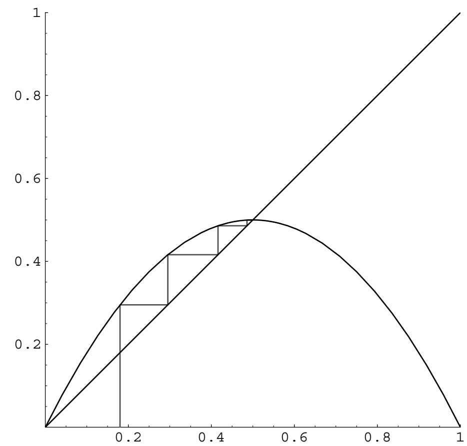


FIGURE 3. Web diagram for the logistic map with $\lambda = 2$

The special case $\lambda = 2$ is the case which separates two different types of behaviour. If $1 < \lambda < 2$ then $f'(1 - 1/\lambda) > 0$ and if we choose a starting point x_0 with $0 < x_0 < 1 - 1/\lambda$ then the sequence $\{f^k(x_0)\}$ increases monotonically towards the fixed point (see Figure 4).

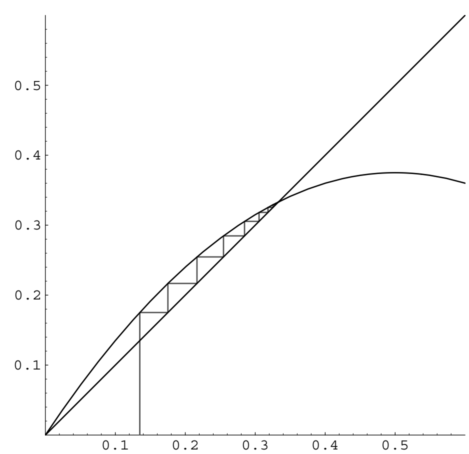


FIGURE 4. Web diagram for the logistic map with $\lambda = 1.5$

If $2 < \lambda < 3$ then $f'(1 - 1/\lambda) < 0$. If we start at x_0 with $0 < x_0 < 1 - 1/\lambda$ then the sequence $\{f^k(x_0)\}$ will increase past the value $1 - 1/\lambda$ and then spiral back in towards the fixed point with values of $f^k(x_0)$ being alternately less than then greater than $1 - 1/\lambda$ (see Figure 5).

$\boxed{\lambda = 3}$ At this value of λ, the non-zero fixed point of f ceases to be attractive, since

$$|f'(2/3)| = 1.$$

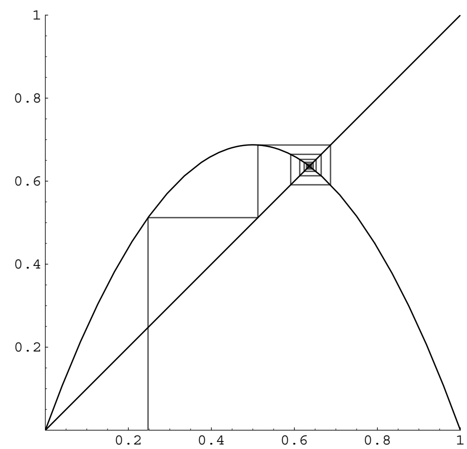


FIGURE 5. Web diagram for the logistic map with $\lambda = 2.75$

It turns out, as in the case when $\lambda = 1$, that orbits do converge to the fixed point at 2/3, but the convergence is extremely slow.

$\boxed{\lambda > 3}$ For each value of λ in this range,

$$|f'(0)| = \lambda > 1,$$
$$\left| f'\left(1 - \frac{1}{\lambda}\right)\right| = \lambda - 2 > 1,$$

so both fixed points at 0 and $1 - 1/\lambda$ are repulsive.

We look for cycles of order 2. Points on such cycles are solutions of

$$x = f^2(x) = f(\lambda x(1-x)) = \lambda^2 x(1-x)(1 - \lambda x(1-x)).$$

Although this is a quartic equation for x, we know that two roots of this equation must correspond to the fixed points at 0 and $1 - 1/\lambda$, since any solution of the equation $x = f(x)$ is automatically a solution of the equation $x = f^k(x)$ for any $k \geq 1$. Expanding out this last equation we get

$$\lambda^2 x \left[-\lambda x^3 + 2\lambda x^2 - (\lambda+1)x + 1 - \frac{1}{\lambda^2}\right] = 0. \tag{2.2}$$

We know that $x = 0$ and $x = 1 - 1/\lambda$ are solutions of this equation, therefore $x - (1 - 1/\lambda)$ must be a factor of the expression inside the square brackets. Using this knowledge, we find that we can rewrite equation (2.2) as

$$\lambda^2 x \left(x - \left(1 - \frac{1}{\lambda}\right)\right) \left(-\lambda x^2 + (\lambda+1)x - \left(1 + \frac{1}{\lambda}\right)\right) = 0.$$

For points on a period 2 cycle we therefore have to solve the quadratic equation

$$-\lambda x^2 + (\lambda+1)x - \left(1 + \frac{1}{\lambda}\right) = 0.$$

The solutions of this equation are

$$x = \frac{1}{2\lambda}\left[\lambda + 1 \pm \sqrt{\lambda^2 - 2\lambda - 3}\right],$$

which we denote by $x_\pm$.

Note that $\lambda^2 - 2\lambda - 3 = (\lambda-3)(\lambda+1)$ which is greater than zero only for $\lambda > 3$. Therefore $x_\pm$ are real for $\lambda > 3$, and we get a cycle of period 2 (see Figure 6). It can be shown (**See Question Sheets**) that this period 2 cycle is attractive for $3 < \lambda < \lambda_2$, where we have defined $\lambda_2 = 1 + \sqrt{6}$.

When we increase λ past λ_2 the period 2 cycle becomes repulsive. It can then be shown that for a region $\lambda_2 < \lambda < \lambda_3$, where λ_3 is approximately 3.56, there is an attractive cycle of period 4.

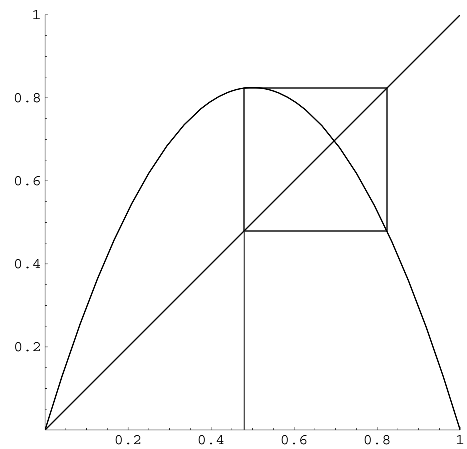


FIGURE 6. Period 2 cycle of the logistic map with $\lambda = 3.3$

As λ is increased past λ_3, this period 4 cycle becomes repulsive, and we then find an attractive cycle of period 8 for $\lambda_3 < \lambda < \lambda_4$, where λ_4 is approximately 3.567.

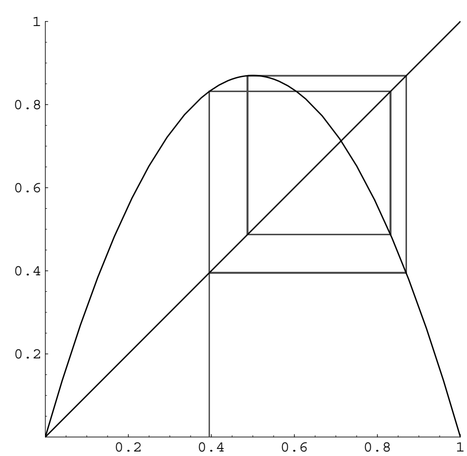


FIGURE 7. Web diagram for the logistic map with $\lambda = 3.48$

In a similar fashion, it is found that we have attractive cycles of period 2^n for each positive integer n, appearing in order, as we increase the value of λ. The period 2^n cycle is attractive for λ in the range $(\lambda_n, \lambda_{n+1})$. This phenomenon is known as *period doubling*. As we let n tend to infinity, we find that the range of values of λ for which cycles are attractive decreases. In particular,

$$\lim_{n \to \infty} \frac{\lambda_n - \lambda_{n-1}}{\lambda_{n+1} - \lambda_n} = \delta,$$

where

$$\delta = 4.6692\ldots. \tag{2.3}$$

This constant δ is called the *Feigenbaum Constant*. In particular, the λ_n form an increasing sequence, and this sequence tends to a limiting value:

$$\lim_{n \to \infty} \lambda_n \equiv \lambda_\infty = 3.57699\ldots.$$

Therefore the system goes through the infinite number of cycles of period 2^n in a finite range of λ. At $\lambda = \lambda_\infty$, the attractive orbit is a type of Cantor set, and can be shown to have dimension $0.538\ldots$.

For $\lambda_\infty < \lambda < 4$, there exist both chaotic attractive orbits, but also regions where there are attractive periodic orbits. In particular, there exists a set $K \subset (\lambda_\infty, 4)$ such that if $\lambda \in K$ then the attractive orbit is chaotic.

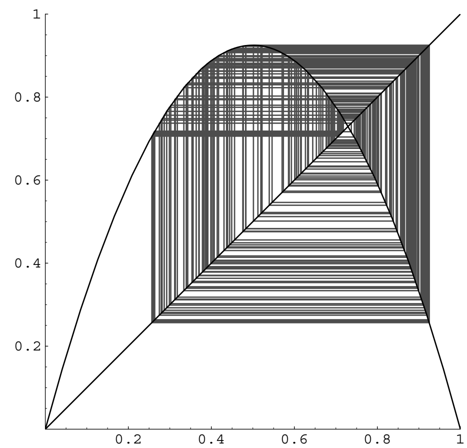


FIGURE 8. A chaotic orbit of logistic map with $\lambda = 3.7$

If $\lambda \notin K$, then there is an attractive cycle of some period. What generally happens is that if we increase the value of λ through values where the attractive orbit is chaotic, we will reach a critical value of λ at which the attractive orbit turns into a cycle of some period, p. As we increase λ past this value, the cycle of period p will become repulsive at some value of λ, and an attractive cycle of period $2.p$ appears. Increasing λ further, the cycle of period $2.p$ becomes repulsive, and we find an attractive cycle of period $4.p$. Similarly, as we increase λ we again find a period doubling region where we go through regions where the attractive cycle has period $2^n.p$ for each positive integer n, ending after some finite range of λ at which the attractive orbit is a Cantor set, and after which the attractive orbit becomes chaotic.

The question of what order the different attractive orbits appear is quite subtle. In order of *decreasing* λ they occur in the order:

$$\begin{array}{cccccc}
3 & 5 & 7 & 9 & 11 & \ldots \\
2.3 & 2.5 & 2.7 & 2.9 & 2.11 & \ldots \\
2^2.3 & 2^2.5 & 2^2.7 & 2^2.9 & 2^2.11 & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
2^n.3 & 2^n.5 & 2^n.7 & 2^n.9 & 2^n.11 & \ldots \\
\ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\
 & & 1. & & &
\end{array} \tag{2.4}$$

In each of these cases, the cycle of a particular period will become attractive at the end of a chaotic region, and after a region where the cycle of this period is attractive, there will be a region of period doubling ending with a attractor which is a Cantor set, and then another region of chaos. For example, the final cascade of period doubling as we increase λ begins with attractive orbits of period 3 at $\lambda \sim 3.83$ (shown in Figure 9) As we increase λ we find attractive cycles of period 6 (see Figure 10) followed by attractive cycles of period $12, 24, 48, \ldots$ with the period doubling terminating in a Cantor set at $\lambda = 3.855\ldots$.

Finally, at $\lambda = 4$, the attractive orbit is *ergodic*. This means that the attractive orbit passes arbitrarily close to any point in the interval $[0, 1]$.

The various attractive orbits are best demonstrated in a *bifurcation diagram*, where we plot the value of λ horizontally and vertically the points that lie on the attractive orbit for that value

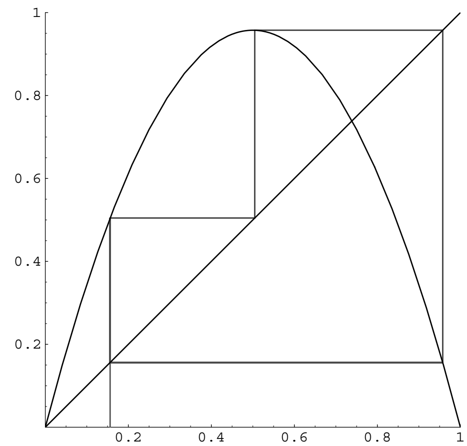


FIGURE 9. Period 3 cycle with $\lambda = 3.83$

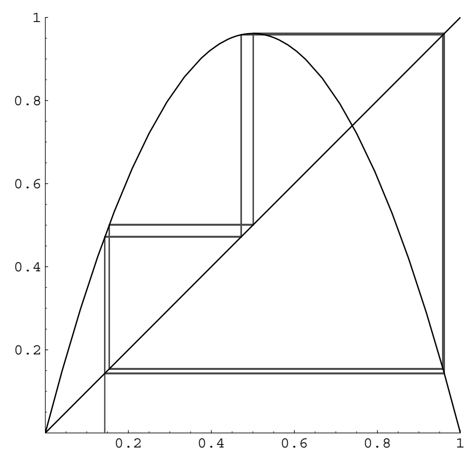


FIGURE 10. Period 6 cycle with $\lambda = 3.845$

of λ. For $1 < \lambda < 3$ we have a single line $x = 1 - 1/\lambda$. At $\lambda = 3$, this line splits up into the two line segments which characterise the attractive orbit of period 2 for $3 < \lambda < 1 + \sqrt{6}$. At $\lambda = 1 + \sqrt{6}$, these two lines each split into two more lines, corresponding to the attractive orbit of period 4, and so on. For the logistic map, the bifurcation diagram is shown in Figure 11.

Universality. Although the logistic map may seem like a rather simplistic system, it can be shown that for *any* transformation on the interval of the form $f_\lambda(x) = \lambda f(x)$, where $f(x)$ has a unique maximum on the interval, then as λ increases the behaviour of the attractive orbits will be qualitatively the same as in the logistic map. In particular, we will see the period doubling regions and chaotic regions interspersed with periodic regions (with the individual periods appearing in the same order as in the logistic map). Although the values $\lambda_2, \lambda_3, \ldots$ at which the period doubling occurs will depend on the particular function f that we choose, the ratio of the range of λ for which these periods occur is universal, and given by the Feigenbaum constant δ, as in equation (2.3). Also, the dimension of the Cantor set attractive orbit at λ_∞ is $0.538\ldots$ for any function f that is differentiable and has a single maximum on the interval.

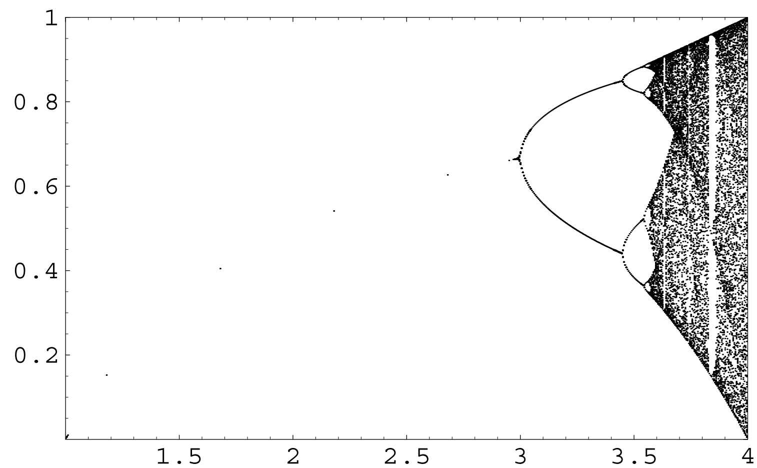


FIGURE 11. Bifurcation diagram for the logistic map

4. Chaos

One of the basic features we expect of a chaotic system is that it be sensitive to initial conditions. This means that if we have an orbit, D, of a map f and we consider iterations starting at two points on the orbit which are close together, then we eventually expect the orbits of the points to become totally uncorrelated.

DEFINITION 2.7. Let D be a subset of a metric space (X, d). A map $f : D \to D$ is *sensitive to initial conditions* if there exists a $\delta > 0$ such that $\forall x \in D$ and $\forall \epsilon > 0$, there exists an integer n and a $y \in D$ with $d(x, y) < \epsilon$, but $d(f^n(x), f^n(y)) > \delta$.

Implicit in the term "chaos", however, is also the assumption that if we start off the iterations at a generic point, then the orbit of the point will fill out a region of X densely.

DEFINITION 2.8. Let D be a subset of a metric space (X, d). A map $f : D \to D$ is *transitive* if for points $x, y \in D$ and for all $\epsilon > 0$, there exists a $z \in D$ such that $d(x, z) < \epsilon$ and $d(y, f^n(z)) < \epsilon$ for some n.

In practice, if $X = D$, what we often search for is an orbit of f which is *dense* in X. Recall

DEFINITION 2.9. If A and B are sets with $A \subseteq B$ then A *is dense in* B if B equals the union of A with the limit points. If (B, d) is a metric space this condition is equivalent to saying that for all $x \in B$ and for all $\epsilon > 0$, there exists $y \in A$ with $d(x, y) < \epsilon$.

It then follows that if there exists an orbit $\{x, f^1(x), f^2(x), \dots\}$ which is dense in X, then the map $f : X \to X$ is transitive.

Taking into account the fact that a system will generally have a large number of repulsive periodic cycles, we are lead to the following working definition of a chaotic system:

DEFINITION 2.10. Let D be a subset of a metric space (X, d). A map $f : D \to D$ is *chaotic* if:

(1) it is sensitive to initial conditions,
(2) it is transitive
(3) the set of periodic points of f is dense in X.

It is generally very difficult to show that a given system is chaotic. The following example demonstrates a system which is sensitive to initial conditions:

EXAMPLE 2.10. Consider the sequence of numbers generated by the relation

$$x_{n+1} = x_n + r x_n \left(1 - x_n\right),$$

which corresponds to iterating the function

$$f(x) = x + rx\,(1-x)\,.$$

This equation is sometimes used to model population growth. If we let the parameter r equal 3, then it can be shown that this system is sensitive to initial conditions. For example, if we consider iterating the above map on two calculators where one calculator is accurate to 10 decimal places and the other is accurate to 12 decimal places, we find that the rounding errors introduced by the calculators lead to uncorrelated orbits (i.e. different answers) after about 50 iterations:

Iterations	Casio	HP
1	0.0397	0.0397
2	0.15407173	0.15407173
3	0.5450726260	0.545072626044
4	1.288978001	1.28897800119
5	0.1715191421	0.171519142100
10	0.7229143012	0.722914301711
15	1.270261775	1.27026178116
20	0.5965292447	0.596528770927
25	1.315587846	1.31558435183
30	0.3742092321	0.374647695060
35	0.9233215064	0.908845072341
40	0.0021143643	0.143971503996
45	1.219763115	1.23060086551
50	0.0036616295	0.225758993390

TABLE 1. Iterations of $f(x) = x + rx\,(1-x)$ with $r = 3$ on two different calculators.

EXAMPLE 2.11 (Symbol/Sequence space). Let

$$X = \{\text{ Sequences } (s_0 s_1 s_2 \dots) : s_i = 0 \text{ or } 1, \forall i \geq 0\}\,.$$

So if $s = s_0 s_1 s_2 \dots \in X$ consists of a sequence of 0's and 1's, $s_0 s_1 s_2 \dots$. If $s = s_0 s_1 s_2 \dots \in X$ and $t = t_0 t_1 t_2 \dots \in X$ then we define the metric

$$d(s,t) := \sum_{k=0}^{\infty} \frac{|s_k - t_k|}{2^k}.$$

Therefore

$$\begin{aligned} d(s,t) = 0 &\Rightarrow \sum_{k=0}^{\infty} \frac{|s_k - t_k|}{2^k} = 0 \\ &\Rightarrow |s_k - t_k| = 0, \forall k \geq 0 \\ &\Rightarrow s_k = t_k, \forall k \geq 0 \\ &\Rightarrow s = t. \end{aligned}$$

Also, since $|s_k - t_k| \leq 1, \forall k \geq 0$, we have that

$$d(s,t) := \sum_{k=0}^{\infty} \frac{|s_k - t_k|}{2^k} \leq \sum_{k=0}^{\infty} \frac{1}{2^k} = 2.$$

Therefore $0 \leq d(s,t) \leq 2$ for all $s, t \in X$.

For example, if $s = 000000\dots, t = 101010\dots$ then

$$d(s,t) = \frac{1}{2^0} + \frac{0}{2^1} + \frac{1}{2^2} + \frac{0}{2^3} + \frac{1}{2^4} + \frac{0}{2^5} + \dots = 1 + \frac{1}{4} + \frac{1}{16} + \dots = \frac{1}{1-\frac{1}{4}} = \frac{4}{3}.$$

LEMMA 2.2. *Let* $s = s_0s_1s_2\dots, t = t_0t_1t_2\dots \in X$.

(1) *If* $s_0 = t_0, s_1 = t_1, \dots, s_n = t_n$ *then* $d(s,t) \le 2^{-n}$.
(2) *If* $d(s,t) \le 2^{-n}$ *then* $s_0 = t_0, s_1 = t_1, \dots, s_{n-1} = t_{n-1}$.

PROOF.

(1) Suppose $s_0 = t_0, s_1 = t_1, \dots, s_n = t_n$, then

$$\begin{aligned} d(s,t) &= \sum_{k=0}^{\infty} \frac{|s_k - t_k|}{2^k} \\ &= \sum_{k=0}^{n} \frac{|s_k - t_k|}{2^k} + \sum_{k=n+1}^{\infty} \frac{|s_k - t_k|}{2^k} \\ &\le 0 + \sum_{k=n+1}^{\infty} \frac{1}{2^k} \\ &= 2^{-(n+1)} \cdot 2 = 2^n. \end{aligned}$$

(2) We prove the equivalent result that if $s_l = t_l$ for some $l < n$ then $d(s,t) > 2^{-n}$. Therefore, suppose there exists an $l < n$ with $s_l = t_l$. Then

$$d(s,t) = \sum_{k=0}^{\infty} \frac{|s_k - t_k|}{2^k} \ge \frac{|s_l - t_l|}{2^l} = 2^{-l} > 2^{-n}.$$

□

DEFINITION 2.11. The *shift map* $\sigma : X \to X$ is defined by

$$\sigma(s_0s_1s_2s_3\dots) = s_1s_2s_3\dots.$$

To find the fixed points of σ, we have to solve the equation $\sigma(a) = a$ for $a = a_0a_1a_2\dots \in X$. This implies that

$$\sigma(a_0a_1a_2\dots) = a_1a_2\dots = a_0a_1a_2\dots.$$

Hence we require $a_0 = a_1 = a_2 = \dots$. There are therefore two fixed points of σ of the form

$$a_0 := 000000\dots, \qquad a_1 := 111111\dots.$$

To find periodic points of period k, we must solve $\sigma^k(s) = s$. If $s = s_0s_1\dots$, this implies that

$$\sigma^k(s_0s_1\dots) = s_ks_{k+1}s_{k+2}\dots = s_0s_1s_2\dots.$$

Therefore we must have

$$s_{k+i} = s_i, \qquad \forall i \ge 0$$

for points of period k. Therefore

$$s = s_0s_1s_2\dots s_{k-1}s_0s_1s_2\dots s_{k-1}s_0s_1s_2\dots s_{k-1}s_0s_1\dots,$$

where the length k segment $s_0\dots s_{k-1}$ repeats ad infinitum.

Our aim is to show that the map σ is chaotic on X. We first show that it is sensitive to initial conditions.

PROPOSITION 2.1. *Let* $s \in X$ *and* $\epsilon > 0$. *Then there exists a* $t \in X$ *and an integer* N *such that* $d(s,t) < \epsilon$ *and* $d(\sigma^n(s), \sigma^n(t)) = 2$ *for all* $n > N$.

PROOF. Let $s = s_0s_1s_2\dots \in X$ and $\epsilon > 0$. Pick $N > 0$ with the property that $2^{-N} < \epsilon$. Then let $t = t_0t_1t_2\dots$ where we let

$$\begin{aligned} t_i &= s_i && \forall i \le N \\ t_i &\ne s_i && \forall i > N. \end{aligned}$$

Then Lemma 2.2 implies that

$$d(s,t) \le 2^{-N} < \epsilon.$$

Also we have

$$d(\sigma^N(s), \sigma^N(t)) = d(s_N s_{N+1} s_{N+2} \dots, t_N t_{N+1} t_{N+2} \dots) = \sum_{k=0}^{\infty} \frac{|s_{N+k} - t_{N+k}|}{2^k} = \sum_{k=0}^{\infty} \frac{1}{2^k} = 2.$$

□

LEMMA 2.3. *There exists an $s \in X$ the orbit of which is dense in X.*

PROOF. We define the Morse sequence

$$s = 0\,1\,00\,01\,10\,11\,000\,001\,010\,011\,100\,101\,110\,111\,0000\,0001\dots$$

where we first write down all the blocks of length 1 (i.e. 0 then 1) then all the blocks of length 2 (i.e. 00, 01, 10 then 11), then all the blocks of length 3 (i.e. 000, 001, etc), then all the blocks of length 4 (i.e. 0000, 0001, etc), and so on. We need to show that for any $t \in X$ and for all $\epsilon > 0$ there exists an integer N such that $d(\sigma^N(s), t) < \epsilon$. Therefore, given any $t = t_0 t_1 t_2 \dots \in X$ and any $\epsilon > 0$ find an integer n such that $2^{-n} < \epsilon$. Then we look through s for the first occurrence of the sequence of $(n+1)$ numbers $t_0 t_1 \dots t_n$. Since s contains all sequences of length $n+1$, this will always happen at some point in s. Therefore, assume that this sequence occurs for the first time at $s_N s_{N+1} s_{N+2} \dots s_{N+n}$ (i.e. $s_N = t_0, s_{N+1} = t_1, \dots, s_{N+n} = t_n$). Then we have

$$\sigma^N(s) = s_N s_{N+1} \dots s_{N+n} s_{N+n+1} \dots = t_0 t_1 \dots t_n s_{N+n+1} \dots.$$

Therefore

$$d(\sigma^N(s), t) = d(t_0 t_1 \dots t_n s_{N+n+1} \dots, t_0 t_1 \dots t_n t_{n+1} \dots) \le 2^{-N} < \epsilon,$$

where we have used Lemma 2.2 for the second-last inequality. □

LEMMA 2.4. *Periodic orbits of σ are dense in X.*

PROOF. Given any $t = t_0 t_1 \dots \in X$ and for any $\epsilon > 0$, we need to show that there exists a periodic point s of σ with $d(s, t) < \epsilon$. Given $\epsilon > 0$, find an integer n with $2^{-n} < \epsilon$. Then we construct the number

$$s := t_0 t_1 \dots t_n t_0 t_1 \dots t_n t_0 t_1 \dots t_n t_0 t_1 \dots,$$

where the $n+1$ term block $t_0 t_1 \dots t_n$ is repeated. From the discussion of periodic points of σ above, we deduce that s is a periodic point of σ with period $n+1$. We also deduce, from Lemma 2.2 that

$$d(s, t) \le \frac{1}{2^n} < \epsilon$$

as required. □

We therefore have

PROPOSITION 2.2. *The map $\sigma : X \to X$ is chaotic.*

CHAPTER 3

The Mandelbrot Set

For each $c \in \mathbb{C}$, define a map $P_c : \mathbb{C} \to \mathbb{C}$ by

$$P_c(z) = z^2 + c, \qquad \forall z \in \mathbb{C}.$$

DEFINITION 3.1. The *Mandelbrot set* M is the set of complex numbers c for which $P_c^n(0)$ remains bounded as $n \to \infty$.

To see what $P_c^n(0)$ looks like, define $Q_n(c) = P_c^n(0)$, then

$$\begin{aligned} Q_1(c) &= c, \\ Q_2(c) &= c^2 + c, \\ Q_3(c) &= c^4 + 2c^3 + c^2 + c, \end{aligned}$$

and in general $Q_{n+1}(c) = (Q_n(c))^2 + c$. $Q_n(c)$ is a polynomial in c of degree 2^{n-1}, which has non-negative coefficients.

THEOREM 3.1.

1. *If $c \in M$, then $|c| \leq 2$. i.e. M is contained in the ball of radius 2 centred at the origin,*
2. *$-2 \in M$, so M is not contained in any smaller ball centred at 0,*
3. *$\{c \in \mathbb{C} : |c| \leq \frac{1}{4}\} \subset M$,*
4. *$M \cap \mathbb{R} = [-2, \frac{1}{4}]$.*

PROOF. 1. We prove the converse, that if $c \in \mathbb{C}$ and $|c| > 2$ then $c \notin M$. Let $c \in \mathbb{C}$, with $|c| > 2$, and define $\alpha = |c| - 1 > 1$. Then, if we consider any $z \in \mathbb{C}$ with $|z| \geq |c|$ we have:

$$\begin{aligned} |P_c(z)| &= |z||z + \frac{c}{z}| \\ &\geq |z|||z| - |\frac{c}{z}|| \\ &\geq |z|\,(|c| - 1) \\ &= \alpha|z|, \end{aligned}$$

where the second inequality follows from fact that $|z| \geq c$, which means that $|\frac{c}{z}| \leq 1$. It therefore follows by induction that

$$|P_c^n(z)| \geq \alpha^n|z| \to \infty \text{ as } n \to \infty,$$

since $\alpha > 1$. In particular if we take $z = c$, then we deduce that $|P_c^{n+1}(0)| \to \infty$ as $n \to \infty$. Therefore the iterations are unbounded, so $c \notin M$.

2.

$$\begin{aligned} P_{-2}(0) &= -2, \\ P_{-2}^2(0) &= P_{-2}(P_{-2}(0)) = P_{-2}(-2) = (-2)^2 - 2 = 2, \\ P_{-2}^3(0) &= P_{-2}(P_{-2}^2(0)) = P_{-2}(2) = (2)^2 - 2 = 2, \end{aligned}$$

and we deduce that $P_{-2}^n(0) = 2, \forall n \geq 2$. Thus $P_{-2}^n(0)$ is bounded for all n, so $-2 \in M$.

3. If $|c| \leq \frac{1}{4}$, consider any $z \in \mathbb{C}$ with $|z| \leq \frac{1}{2}$. Then

$$|P_c(z)| = |z^2 + c| \leq |z|^2 + |c| \leq \frac{1}{4} + \frac{1}{4} = \frac{1}{2}.$$

Thus, by induction, for all z with $|z| \leq \frac{1}{2}$ and $|c| \leq \frac{1}{4}$ we have $|P_c^n(z)| \leq \frac{1}{2}$ for all n. In particular, taking $z = 0$, we deduce that $P_c^n(0)$ are bounded for all c with $|c| \leq \frac{1}{4}$. Therefore $\{c \in \mathbb{C} : |c| \leq \frac{1}{4}\} \subset M$.

4. From 3. we know that $[-1/4, 1/4] \subset M$ and from 2. that $-2 \in M$. Also, from 1. that if $c \in (-\infty, -2)$ or $c \in (2, \infty)$ then $c \notin M$. We therefore need to consider real c in the regions $(-2, -1/4)$ and $(1/4, 2]$. It turns out to be easier to consider the regions $[-2, 0]$ and $(1/4, \infty)$:

$\underline{(1/4, \infty)}$.

Suppose that $c \in \mathbb{R}$ and $c \in (\frac{1}{4}, \infty)$. For any $x \geq 0$, we know that $x^2 + c = P_c(x) > x$, because the quadratic $x^2 - x + c$ has no real roots. So $P_c^n(0)$ is an increasing sequence of real numbers. If $P_c^n(0)$ is bounded above, then it must converge to a limit a. Since $P_c(x)$ is a continuous function of x, we deduce that a must be a (real) fixed point of P_c. However, P_c has no real fixed points for $c > 1/4$. Therefore the sequence $P_c^n(0)$ cannot be bounded above, and is therefore unbounded. Therefore for any c in this range, $c \notin M$.

$\underline{[-2, 0]}$.

Assume now that $-2 \leq c \leq 0$, and that x is any real number with $x \in [c, |c|]$. Since $x^2 \geq 0$, we have

$$P_c(x) = x^2 + c \geq c.$$

But since $|x| \leq |c|$ and $|x| \leq 2$ we also have

$$P_c(x) = x^2 + c = |x| \cdot |x| + c \leq 2|c| + c = |c|,$$

since $c \leq 0$. Therefore $\forall x \in [c, |c|]$, we have $P_c(x) \in [c, |c|]$. Repeating the argument, we deduce that $\forall x \in [c, |c|]$, we have $P_c^n(x) \in [c, |c|]$, for all n. Letting $x = 0$, we deduce that the iterations $P_c^n(0)$ are bounded in the region $[c, |c|]$ for all $c \in [-2, 0]$, and therefore that all such c are in the Mandelbrot set. $\square$

Some Properties of M (without proof):

- M is symmetric about the real axis. (This follows because $|P_c^n(0)| = |P_{\bar{c}}^n(0)|$, where the bar denotes complex conjugate.)
- $M = \bigcap_{n=0}^{\infty} Q_n^{-1}(\overline{B(0;2)})$. Therefore M is closed.
- If a closed path is contained in M, all the points inside the path will be in M. (This follows from the Maximum Modulus Theorem for analytic functions.)
- M is connected.

1. Fixed Points of P_c and the Main Cardioid

We now wish to study the attractive fixed points of the map $P_c(z)$. It turns out that such points only exist for certain values of c, and we want to find the boundary of this set of values.

Let z be an attractive fixed point of P_c. Therefore $P_c(z) = z$, so $c = -z^2 + z$, and we require that $|P_c'(z)| = 2|z| \leq 1$ for z to be attractive. The boundary of the region where P_c has an attractive fixed point is therefore described by setting $|z| = \frac{1}{2}$, so we put $z = \frac{1}{2}e^{i\theta}$ with $c = -z^2 + z$.

Then

$$\begin{aligned} c - \frac{1}{4} &= -\left(z - \frac{1}{2}\right)^2 \\ &= -\frac{1}{4}\left(\mathrm{e}^{i\theta} - 1\right)^2 \\ &= -\frac{1}{4}\left((\cos\theta - 1) + i\sin\theta\right)^2 \\ &= -\frac{1}{4}\left(-2\sin^2\frac{\theta}{2} + 2i\sin\frac{\theta}{2}\cos\frac{\theta}{2}\right)^2 \\ &= -\sin^2\frac{\theta}{2}\left(-\sin\frac{\theta}{2} + i\cos\frac{\theta}{2}\right)^2 \\ &= -\sin^2\frac{\theta}{2}\left(i\mathrm{e}^{i\theta/2}\right)^2 \\ &= \sin^2\frac{\theta}{2}\cdot\mathrm{e}^{i\theta}. \end{aligned}$$

Therefore

$$c - \frac{1}{4} = r\mathrm{e}^{i\theta}$$

where

$$r = \sin^2\frac{\theta}{2}.$$

This curve in the complex c plane is the called the *Main Cardioid* of the Mandelbrot set (see Figure 1).

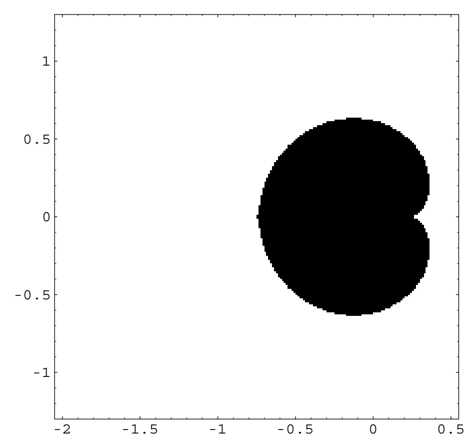


FIGURE 1. The main cardioid of the Mandelbrot set

2. Attractive cycles of period 2

$$P_c^2(z) = (z^2 + c)^2 + c.$$

Therefore

$$\begin{aligned} P_c^2(z) - z &= z^4 + 2cz^2 - z + c^2 + c \\ &= (z^2 - z + c)(z^2 + z + c + 1). \end{aligned}$$

For points z of period 2,

$$z^2 + z + c + 1 = 0 \tag{3.1}$$

For an *attractive* cycle of period 2, $|(P_c^2)'(z)| < 1$, so $|4z(z^2+c)| < 1$. On the boundary, $|4z(z^2+c)| = 1$. Now

$$\begin{aligned} z(z^2+c) &= -z(z+1) \qquad \text{(from (3.1))} \\ &= -(z^2+z) \\ &= c+1 \end{aligned}$$

therefore $|c+1| = \frac{1}{4}$. This is the circle of radius $\frac{1}{4}$, centre -1.

Taking the main cardioid and the region with attractive cycles of period 2 therefore gives us the following regions of the Mandelbrot set (see Figure 2).

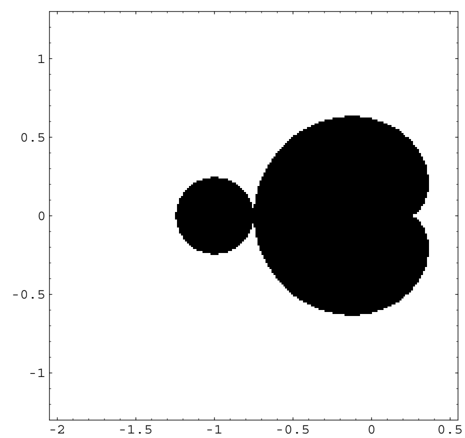


FIGURE 2. The region where P_c has attractive fixed points or attractive cycles of period 2

Note that we have only proved that on the interior of the main cardioid and on the interior of the circle of radius $\frac{1}{4}$, centre -1 there are attractive fixed points and attractive period-2 cycles of the map P_c. It is not obvious that the iterations starting at 0 necessarily converge to these fixed points and cycles. The fact that they do is a consequence of the following result.

THEOREM 3.2 (Fatou). *If the complex polynomial $f(z) = az^2+bz+c$ has an attractive periodic cycle, then the orbit of the critical point of f (i.e. $-b/(2a)$) converges to this periodic cycle.*

In the case of the Mandelbrot set, $a = 1, b = 0$, so the critical point of f is at 0, so we deduce that if the map P_c has an attractive periodic cycle, then the sequence $\{P_c^n(0)\}$ will converge to that cycle. This implies that the points on the interior of the main cardioid and on the interior of the circle centre -1 radius $1/4$ are indeed elements of the Mandelbrot set. The fact mentioned before, that the Mandelbrot set is closed, then implies that the points on the main cardioid and on this circle are also elements of the Mandelbrot set.

As we move along the real axis to more negative values, we find that we get that there are regions of the Mandelbrot set corresponding to higher order attractive cycles. In particular, after the main cardioid and the region corresponding to attractive cycles of period 2, there are regions corresponding to attractive cycles of period $4, 8, 16, \ldots, 2^n, \ldots$. There are an infinite number of these regions corresponding to the powers of 2 in order, but they finish by the time we reach roughly $c = -1.4$ (see Figure 3).

This behaviour is the same as the period doubling phenomenon found in the logistic map earlier. If we continue to move to more negative, real values of c, we encounter regions which look similar to the full Mandelbrot set, but much smaller. These regions correspond to the higher order attractive cycles of the map P_c. In particular, in the same way that we encountered periodic cycles in the Logistic map which started as some fundamental period p and then underwent period

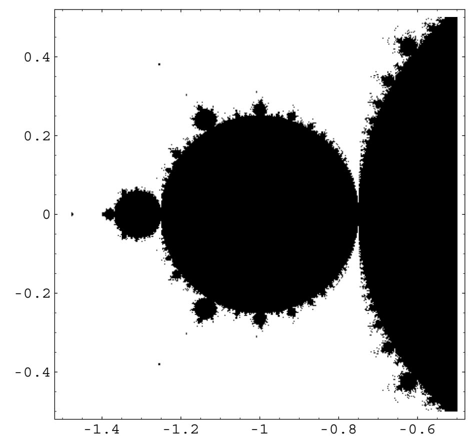


FIGURE 3. The regions where P_c has attractive cycles of period $2, 4, 8, 16 \ldots$

doubling, through all the periods $s^n \cdot p$, in the Mandelbrot set we encounter smaller copies of the main cardioid, corresponding to attractive cycles of period p for the map P_c, and then regions at more negative values of c corresponding to the attractive cycles of periods $2^n \cdot p$. To this end it is useful to consider the bifurcation diagram for the map $f(x) = x^2 + c$ for real values of c and compare it with the regions along the real axis of the Mandelbrot set (See Figure 4)

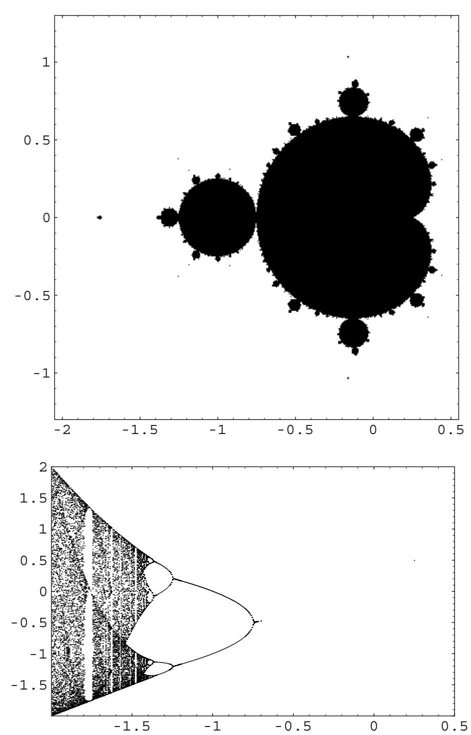


FIGURE 4. The Mandelbrot Set and the bifurcation diagram for $f(x) = x^2 + c$

There are also regions in the complex c plane that correspond to attractive cycles of some order for P_c. Although it is difficult to carry out any explicit calculations, one finds that there are nodules located attached to the boundary of the main cardioid and the areas corresponding to periodic cycles on the real axis, and that each of these nodules corresponds to a region of complex values of c for which there is a cycle of some period. The periods for some regions around the main cardioid are as shown in Figure 5.

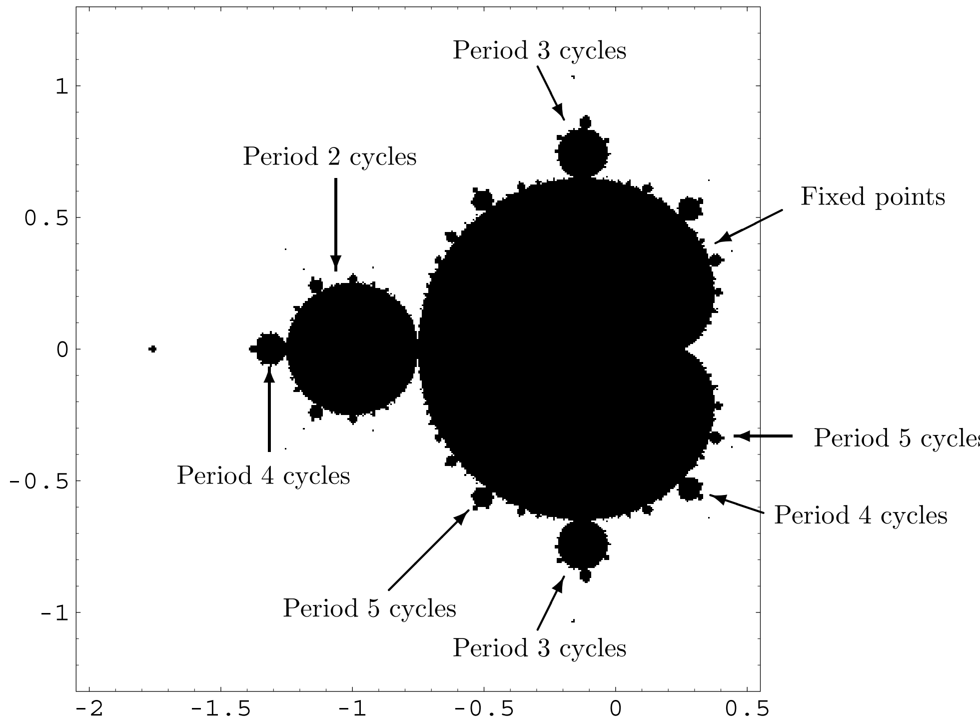


FIGURE 5. Regions of the Mandelbrot corresponding to attractive cycles of P_c

3. Julia sets

DEFINITION 3.2. Given a metric space (X, d) and a map $f : X \to X$, the *prisoner set* of f is the set of points in X whose orbits remain bounded. The *escape set* of f is the complement of the prisoner set. The *Julia set* is the boundary of the prisoner set.

EXAMPLE 3.1. Let $X = \mathbb{C}$ with $d(w, z) = |z - w|$ and $f(z) = z^2$. Then the escape set of f is $\{z \in \mathbb{C} : |z| > 1\}$, and the prisoner set is $\{z \in \mathbb{C} : |z| \leq 1\}$. The Julia set is the unit circle in the complex plane, $\{z \in \mathbb{C} : |z| = 1\}$.

If we perturb this example and consider $X = \mathbb{C}$ with $d(w, z) = |z - w|$ and $f(z) = z^2 + c$, where $c \in \mathbb{C}$ is a complex number, we find that the Julia set for a general value of c is actually a fractal set. Such a set can be either connected (i.e. roughly speaking, any two points in the set can be connected by a path in the set) or it may not be connected. This gives us an alternative definition of the Mandelbrot set:

DEFINITION 3.3. The Mandelbrot set M is the set of complex numbers c for which the corresponding Julia set is connected.

Aside. Recall that a space X is *connected* if it cannot be expressed as the union of two disjoint open sets. Roughly speaking, this means that the set X is in one part. In many cases (but not always) this is equivalent to the notion of a set being *path connected*, which means that any two points in the set can be joined by a path in the set.

We will not prove the equivalence of our two definitions of M, but simply state some simple properties of Julia sets.

PROPOSITION 3.1. *Julia sets are closed and bounded.*

Terminology. The prisoner set, defined above as the set of complex numbers z for which $P_c^n(z)$ remains bounded as $n \to \infty$, is sometimes also called the *filled in Julia set* and denoted F_c for a fixed complex number c.

THEOREM 3.3. *Given $c \in \mathbb{C}$. Suppose that $|z| > 2$ and that $|z| \geq |c|$. Then z is in the escape set of P_c.*

PROOF. See **Question Sheets**. Let $c, z \in \mathbb{C}$ with $|z| > 2$ and that $|z| \geq |c|$. Defining $\alpha = |z| - 1 > 1$, then

$$\begin{aligned} |P_c(z)| &= |z^2 + c| \\ &= |z||z + \frac{c}{z}| \\ &\geq |z| \left(|z| - |\frac{c}{z}|\right) \\ &> |z| \left(|z| - 1\right) \\ &> \alpha|z|. \end{aligned}$$

Repeating this argument, we deduce that $|P_c^n(Z)| \geq \alpha^n |z| \to \infty$ as $n \to \infty$. □

COROLLARY 3.1. *For a given $c \in \mathbb{C}$ then $z \in F_c$ if and only if $|P_c^n(z)| \leq max\{2, |c|\}$, for all $n \geq 1$.*

CHAPTER 4

Contraction Mappings and Iterated Function Systems

1. An alternative construction of the Cantor set

We saw earlier that we could describe the Cantor set, C, in terms of a base 3 expansion of fractions and that

$$C = \left\{ x \in [0,1] : x = \frac{a_1}{3} + \frac{a_2}{3^2} + \frac{a_3}{3^3} + \dots, \text{with } a_1, a_2, \dots \neq 1 \right\}.$$

We now wish to investigate an alternative, but equivalent, construction of the Cantor set.

Consider the maps w_1, w_2 from the interval $[0,1]$ to itself defined by

$$w_1(x) = \frac{x}{3},$$
$$w_2(x) = \frac{x}{3} + \frac{2}{3}.$$

Then we found earlier that the Cantor set is the set which satisfies the relation $C = w_1(C) \cup w_2(C)$.

Similarly, if we iterate the map $w = w_1 \cup w_2$ on the unit interval $[0,1]$, we generate the Cantor set in the same way that it was constructed earlier:

2. Contraction Mappings

We should note that in the maps w_1 and w_2 defined above, the coefficient of the linear terms in x is $1/3$ in both maps, meaning that the image of the unit interval under each map is of length $1/3$, which is strictly less than 1. Therefore, the images are shrinking. If the coefficient was greater than 1, the images of sets would be growing, and under iteration of such maps, the interval $[0,1]$ would grow, and become unbounded. Maps with the property that the image of a set is smaller than the set itself may be formalised:

Definition 4.1. Let (X, d) be a metric space. A mapping $f : X \to X$ is a *contraction mapping* if there exists a real number s with $0 \leq s < 1$ such that $d(f(x), f(y)) \leq s d(x, y)$ for all $x, y \in X$. In this case, the number s is called the *contraction factor* for f.

Example 4.1. Let $X = \mathbb{R}, d(x,y) = |x - y|$, and $f(x) = ax + b$. Then for any $x, y \in \mathbb{R}$ we have

$$d(f(x), f(y)) = |ax + b - (ay + b)| = |a(x - y)| = |a||x - y| = |a| d(x, y).$$

Therefore the map f is a contraction mapping if and only if $|a| < 1$ and in this case the contraction factor is $|a|$. Note that therefore the two maps w_1 and w_2 used in the definition of the Cantor set are contraction mappings.

Of particular interest to us will be affine maps:

DEFINITION 4.2. A map from $\mathbb{R}^m$ to $\mathbb{R}^n$ is *affine* if it is the sum of a linear map and a translation.

What this means is that if we represent vectors $x \in \mathbb{R}^m$ by m component column vectors, and vectors in $\mathbb{R}^n$ by n component column vectors, then we can write $f(x) \in \mathbb{R}^n$ in the form

$$f(x) = Ax + B,$$

where A is an $n \times m$ constant real matrix, and B is a constant $n \times 1$ matrix.

In the current context, we define a continuous function in the following way:

DEFINITION 4.3. On a metric space (X, d), a function $f : X \to X$ is *continuous* at $x \in X$ if $\forall \epsilon > 0, \exists \delta > 0$ with the property that $d(f(x), f(y)) < \epsilon$ for all $y \in X$ such that $d(x, y) < \delta$.

One basic property of a contraction is:

PROPOSITION 4.1. *Let $f : X \to X$ be a contraction mapping on metric space (X, d). Then f is continuous on X.*

PROOF. Let $\epsilon > 0$ and s with $0 \le s < 1$ be the contraction factor for f. Then $d(f(x), f(y)) \le s \cdot d(x, y)$ since f is a contraction mapping. Thus $d(f(x), f(y)) \le s \cdot d(x, y) < \epsilon$ if we take $d(x, y) < \delta$ with $\delta = \epsilon/s$. □

We also recall the following:

DEFINITION 4.4. A sequence $\{x_n\}$ in a metric space (X, d) is a *Cauchy sequence* if, for all $\epsilon > 0$, there exists an integer N such that $d(x_m, x_n) < \epsilon$ whenever $m, n > N$.

DEFINITION 4.5. A metric space is *complete* if every Cauchy sequence in the space converges to a limit in the space.

We then have the following fundamental result:

THEOREM 4.1 (Contraction Mapping Theorem). *Let $f : X \to X$ be a contraction mapping on a complete metric space (X, d). Then f possesses exactly one fixed point $a \in X$, and for any point $x \in X$, the sequence $\{f^n(x)\}$ converges to a, so*

$$\lim_{n\to\infty} f^n(x) = a, \qquad \forall x \in X.$$

PROOF. Let $x \in X$ and s with $0 \le s < 1$ be the contraction factor for f. Let m, n be non-negative integers and, without loss of generality, assume that $m \le n$. Then

$$d(f^m(x), f^n(x)) \le s^m d(x, f^{n-m}(x)), \forall x \in X, \tag{4.1}$$

where we have used the fact that f is a contraction mapping. If we consider a general point $x \in X$, then

$$\begin{aligned}
d(x, f^k(x)) &\le d(x, f(x)) + d(f(x), f^k(x)) \\
&\le d(x, f(x)) + d(f(x), f^2(x)) + d(f^2(x), f^k(x)) \\
&\le \ldots \\
&\le d(x, f(x)) + d(f(x), f^2(x)) + \cdots + d(f^{k-1}(x), f^k(x)) \\
&\le d(x, f(x)) + s \cdot d(x, f(x)) + \cdots + s^{k-1} \cdot d(x, f(x)) \\
&= d(x, f(x)) \left[1 + s + s^2 + \cdots + s^{k-1}\right] \\
&= d(x, f(x)) \frac{1 - s^k}{1 - s} \\
&\le \frac{d(x, f(x))}{1 - s},
\end{aligned} \tag{4.2}$$

where, in the first four lines, we have repeatedly used the fact that $d(x, y) \le d(x, z) + d(z, y), \forall x, y, z \in X$, the fact that f is a contraction mapping with contraction factor s in the fifth line, the fact that $0 \le s < 1$ in the seventh and eighth line.

If we define $x_0 = x$ and $x_{n+1} = f(x_n), n \geq 0$, then we deduce from equations (4.1) and (4.2) that

$$d(x_m, x_n) \leq \frac{s^m}{1-s} d(x, f(x)).$$

Thus, since $s < 1$, choosing $\epsilon > 0$, we may choose positive integer N such with the property that

$$\frac{s^N}{(1-s)} d(x, f(x)) < \epsilon,$$

in which case

$$d(x_m, x_n) < \epsilon, \qquad \forall m, n > N.$$

Thus the sequence $\{f^n(x)\}_{n=0}^{\infty}$ is a Cauchy sequence. Since X is complete, this sequence converges to a limit $a \in X$ with

$$\lim_{n \to \infty} f^n(x) = a.$$

Since f is a contraction mapping, it is continuous. Thus:

$$f(a) = f(\lim_{n \to \infty} f^n(x)) = \lim_{n \to \infty} f^{n+1}(x) = a,$$

where we have used the fact that f is continuous in the second equality, and the last equality arises since the sequence $\{f^{n+1}(x)\}_{n=0}^{\infty}$ is the same sequence as $\{f^n(x)\}_{n=0}^{\infty}$ with the first term deleted, so it must also have limit a. Thus we have $f(a) = a$, so the limit point a is a fixed point of f.

To show that the limit point is unique, suppose that the map f has two fixed points a_1 and a_2. Using the facts that a_1 and a_2 are fixed points of f, an that f is a contraction mapping in turn we have:

$$d(a_1, a_2) = d(f(a_1), f(a_2)) \leq s \cdot d(a_1, a_2),$$

from which we deduce that

$$(1-s) \cdot d(a_1, a_2) \leq 0.$$

Since $s < 1$, this implies that $d(a_1, a_2) \leq 0$. From the properties of d we must have $d(a_1, a_2) \geq 0$, so the only possibility is $d(a_1, a_2) = 0$, which therefore means that $a_1 = a_2$. Thus the fixed point a of f is unique. □

EXAMPLE 4.2. Consider the Newton-Raphson example that we saw earlier. Take $X = [\sqrt{k}, \infty) \subset \mathbb{R}$, $d(x, y) = |x - y|$ and

$$f(x) = \frac{1}{2}\left(x + \frac{k}{x}\right).$$

We then have, $\forall x, y \in X$:

$$\begin{aligned} d(f(x), f(y)) &= \frac{1}{2}\left|x - y + k\left(\frac{1}{x} - \frac{1}{y}\right)\right| \\ &= \frac{1}{2}|x-y|\left|1 - \frac{k}{xy}\right| \\ &= \frac{1}{2}|x-y|\left(1 - \frac{k}{xy}\right) \\ &\leq \frac{1}{2}|x-y| \\ &= \frac{1}{2}d(x, y), \end{aligned}$$

where in the third and fourth lines we have used the fact that x and y are $\geq \sqrt{k}$. Therefore the mapping f is a contraction mapping with contraction factor $1/2$. Assuming that (X, d) is a complete metric space (which it is), the Contraction Mapping Theorem implies that f will have a unique fixed point on X, and that $f^n(x)$ will tend to this fixed point for any $x \in X$ as we let n tend to ∞. Since we know that the function f has a fixed point at $\sqrt{k}$, we therefore deduce that $f^n(x) \to \sqrt{k}$ as $n \to \infty$, $\forall x \in X$.

Note. It is important that the contraction factor be strictly less than 1 in order to apply the Contraction Mapping Theorem. For example, consider the map $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = x+1$. Taking $d(x, y) = |x - y|$, we have that $d(f(x), f(y) = d(x, y)$, however the map f has no fixed points.

3. The space of fractals and the Hausdorff Metric

We now wish to introduce the idea of a *space of fractals*. The main idea is that we will define fractals in terms of a contraction mapping on this space, and then use the Contraction Mapping Theorem to deduce the existence of a limit point, which will be our fractal.

Note. From now on we will specialise to the case where X is a subset of $\mathbb{R}^n$, with $n \geq 1$, and where the metric takes the standard form $d(x, y) = |x - y|, \forall x, y \in X$.

DEFINITION 4.6. $\mathscr{H}(\mathbb{R}^n)$ is the space of non-empty closed bounded subsets of $\mathbb{R}^n$. ($\mathscr{H}(X)$ is sometimes referred to as the *space of fractals*.)

A point $x \in \mathscr{H}(\mathbb{R}^n)$ corresponds to a non-empty, closed, bounded subset of $\mathbb{R}^n$. Therefore if $x, y \in \mathscr{H}(\mathbb{R}^n)$ then $x \cup y$ is the union of two non-empty, closed, bounded subsets of $\mathbb{R}^n$ and is therefore itself a non-empty, closed, bounded subset of $\mathbb{R}^n$, and thus an element of $\mathscr{H}(\mathbb{R}^n)$. If we consider $x \cap y$, however, this could be the empty set if the sets corresponding to x and y do not intersect, therefore $x \cap y$ is not necessarily an element of $\mathscr{H}(\mathbb{R}^n)$.

The space $\mathscr{H}(\mathbb{R}^n)$ carries a natural metric:

DEFINITION 4.7. The *Hausdorff Metric*, h, on $\mathscr{H}(\mathbb{R}^n)$ is defined by stating that $h(A, B)$ is the smallest real number δ such that $d(a, B) \leq \delta, \forall a \in A$, and $d(A, b) \leq \delta, \forall b \in B$.

Aside. Note that if p is a point in $\mathbb{R}^n$ and X is a closed, bounded subset of $\mathbb{R}^n$, then

$$d(p, X) = \min_{x \in X} d(a, x).$$

Intuitively, $d(p, X)$ is the least distance from p to any point of the set X. Such a number necessarily exists, since we have taken the set X to be closed and bounded. For example, if we consider $\mathbb{R}^2$, with X a straight line in the plane, then the distance $d(p, X)$ is the distance along the straight line through p which intersects the line X at right angles:

THEOREM 4.2. *$(\mathscr{H}, h)$ is a complete metric space.*

We will not prove this result, although it is fairly straightforward. The interested reader can find the proof in Chapter 2 of Barnsley's book.

EXAMPLE 4.3. Consider $\mathscr{H}(\mathbb{R}) = \{\text{intervals } [a, b] \text{ with } -\infty < a < b < \infty\} \cup \{\text{points } \{\text{a}\} \text{ with } a \in \mathbb{R}\}$.

(1) Suppose $A = \{\frac{1}{3}\}, B = \{0, 1\}$, what is $h(A, B)$? The situation is shown in the following picture

By definition $h(A, B)$ is the smallest real number δ with $d(a, B) \leq \delta, \forall a \in A$ and $d(b, A) \leq \delta, \forall b \in B$. Therefore we require:

$$d\left(\frac{1}{3}, 0\right) = \frac{1}{3} \leq \delta, \qquad d\left(\frac{1}{3}, 1\right) = \frac{2}{3} \leq \delta,$$
$$d\left(0, \frac{1}{3}\right) = \frac{1}{3} \leq \delta, \qquad d\left(1, \frac{1}{3}\right) = \frac{2}{3} \leq \delta.$$

Therefore $h(A, B)$ is the smallest δ that satisfies $\delta \geq 2/3$, so $h(A, B) = 2/3$.

(2) Let $A = [0, 1]$ and $B = [0, \frac{1}{2}]$.

To find $h(A, B)$ we require the smallest δ with $d(a, B) \leq \delta, \forall a \in A$ and $d(b, A) \leq \delta, \forall b \in B$. If we consider the second condition first, given any point $b \in B$, then since $B \subset A$, we know that $b \in A$, therefore $d(b, A) = \min_{a \in A} d(b, a) = d(b, b) = 0$, so $\delta \geq 0$. However, we also require $d(a, B) \leq \delta, \forall a \in A$, and thus we need to find $\max_{a \in A} d(a, B)$. The point in A that is furthest away from the set B is the point $1 \in A$, and

$$d(1, B) = d\left(1, \frac{1}{2}\right) = \frac{1}{2} \leq \delta.$$

$h(A, B)$ is the smallest δ with this property, so we conclude that $h(A, B) = 1/2$ in this case.

4. Iterated Function Systems

DEFINITION 4.8. An *Iterated Function System* (IFS) consists of a complete metric space (X, d) and a finite number of contraction mappings $w_1, \ldots, w_k : X \to X$ with contraction factors $s_1, \ldots, s_k$. Such an IFS will be denoted $\{X; w_1, \ldots, w_k\}$. The contraction factor of the IFS is defined to be $s = \max\{s_1, \ldots s_k\}$.

Note. Since $s_i \leq s$ for each i, the contraction factor, s of the IFS is a contraction factor for each of the w_i, since $d(w_i(x), w_i(y)) \leq s_i \cdot d(x, y) \leq s \cdot d(x, y), \forall x, y \in X$.

THEOREM 4.3. *Let* $\{\mathbb{R}^n : w_1, \ldots, w_k\}$ *be an IFS with contraction factor* s*. Then the mapping* $w : \mathscr{H}(\mathbb{R}^n) \to \mathscr{H}(\mathbb{R}^n)$ *defined by* $w(A) = w_1(A) \cup \cdots \cup w_k(A), \forall A \in \mathscr{H}(\mathbb{R}^n)$ *is a contraction mapping with contraction factor* s*.*

PROOF. Suppose that $h(A, B) = \delta$. Let $x \in w(A)$. Then $x \in w_i(A)$ for some i, and so $x = w_i(a)$ for some $a \in A$. There will be a point $b \in B$ for which $d(a, b) \leq \delta$. This implies that $d(w_i(a), w_i(b)) \leq s\delta$. Since $w_i(b) \in w(B)$, then $d(x, w(B)) \leq s\delta$. Similarly, for any $y \in w(B)$, we have $d(y, w(A)) \leq s\delta$. Thus $h(w(A), w(B)) \leq s\delta$. □

The idea is that, given an IFS $\{X : w_1, \ldots, w_n\}$ we construct a contraction mapping w on $(\mathscr{H}(X), h)$. Since $(\mathscr{H}(X), h)$ is a complete metric space, we can then invoke the Contraction Mapping Theorem, which will tell us that the map w will have a unique fixed point in $\mathscr{H}(X)$, which will be our fractal.

DEFINITION 4.9. Let $(\mathbb{R}^n : w_1, \ldots w_k)$ be an IFS. Let $w : \mathscr{H}(\mathbb{R}^n) \to \mathscr{H}(\mathbb{R}^n)$ be defined by $w(X) = w_1(X) \cup \cdots \cup w_n(X)$. The *attractor* of the IFS is the unique set $A \in \mathscr{H}(\mathbb{R}^n)$ for which $w(A) = A$. It has the property that for all $X \in \mathscr{H}(\mathbb{R}^n)$, $w^r(X) \to A$ as $r \to \infty$, in the sense that $h(w^r(X), A) \to 0$.

Note. Such an attractor of an IFS is sometimes referred to as a *deterministic fractal*.

So far, we have considered IFS's with general contraction mappings. One special type of IFS of particular interest are those where all the mappings are affine:

DEFINITION 4.10. An Iterated Function System is affine if each of the mappings $w_1, \ldots, w_k : \mathbb{R}^n \to \mathbb{R}^n$ is affine. Viewing the vectors in $\mathbb{R}^n$ as column vectors,

$$\begin{aligned} w_1(x) &= A_1 x + B_1, \\ w_2(x) &= A_2 x + B_2, \\ &\cdots \end{aligned}$$

where $A_1, \ldots A_k$ are $n \times n$ matrices and $B_1, \ldots, B_k$ are $n \times 1$ matrices.

In the special case when $n = 2$, each map has the form:

$$\begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix},$$

for some constants a, b, c, d, e, f. The *code* of the IFS specifies these constants for each of the maps.

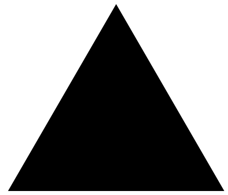
EXAMPLE 4.4 (The Cantor Set). Consider $I = [0, 1] \subset \mathbb{R}$, with

$$\begin{aligned} w_1(x) &= \frac{x}{3}, \\ w_2(x) &= \frac{x}{3} + \frac{2}{3}. \end{aligned}$$

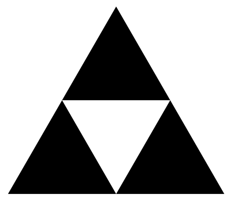
As we have seen above, both w_1 and w_2 are contraction mappings with contraction factor $s_1 = s_2 = 1/3$. Therefore $\{I; w_1, w_2\}$ is an IFS (affine) with $s = 1/3$. Therefore the map $w : \mathscr{H}(I) \to \mathscr{H}(I)$ defined by $X \mapsto w_1(X) \cup w_2(X), \forall X \in \mathscr{H}(I)$ is a contraction mapping.

Therefore w has a unique fixed point in $\mathscr{H}(I)$, the attractor of the IFS, which corresponds to the Cantor set $C \subset I$.

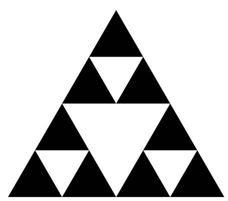
Example 4.5 (The Sierpinski Gasket). To construct the Sierpinski Gasket, we begin with the triangle T in the plane:



In a similar fashion to the way that we remove the middle third of a line segment to get a Cantor set, in this case we remove an inverted triangle of $1/2$ the size of triangle T from the centre of T. We then iterate this procedure of removing an inverted triangle from each solid triangle to form the fractal. After the first iteration we have:



and after two iterations:

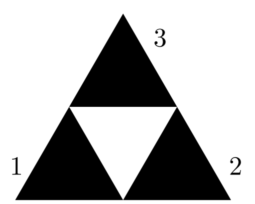


Note that after 1 iteration, the shape we have is made up of 3 triangles. Each of these 3 triangles is identical to T scaled by a factor of $1/2$, and in two cases, we then translate the triangles to the points $(1/2, 0)$ and $(1/4, \sqrt{3}/4)$. Therefore, we could equally well construct this shape as the set $w(T)$ where

$$w(T) = w_1(T) \cup w_2(T) \cup w_3(T),$$

where the maps w_1, w_2 and w_3 are maps of the plane which map the original triangle T into the three separate triangles which make up the first iteration.

For example, if we label the three triangles as follows:



then w_1 is the map which scales the triangle T by a factor of $1/2$, so

$$w_1 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}.$$

The map w_2 should scale the triangle T by a factor of $1/2$ and then translate it the point $(1/2, 0)$ so

$$w_2 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{1}{2} \\ 0 \end{pmatrix}.$$

Finally, the map w_3 should scale the triangle T by a factor of $1/2$ and translate it the point $(1/4, \sqrt{3}/4)$ so:

$$w_3 \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} \frac{1}{4} \\ \frac{\sqrt{3}}{4} \end{pmatrix}.$$

The maps w_1, w_2, w_3 are all contraction maps on $\mathbb{R}^2$ with contraction factors all equal to $1/2$. Thus $\{\mathbb{R}^2; w_1, w_2, w_3\}$ is an IFS with contraction factor $1/2$. The code of the IFS is written:

	a	b	c	d	e	f
w_1	$\frac{1}{2}$	0	0	$\frac{1}{2}$	0	0
w_2	$\frac{1}{2}$	0	0	$\frac{1}{2}$	$\frac{1}{2}$	0
w_3	$\frac{1}{2}$	0	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{\sqrt{3}}{4}$

From the fact that $(\mathscr{H}(\mathbb{R}^2), h)$ is a complete metric space, we deduce that the map $w : \mathscr{H}(\mathbb{R}^2) \to \mathscr{H}(\mathbb{R}^2)$ defined by $w(X) = w_1(X) \cup w_2(X) \cup w_3(X)$ for any $X \in \mathscr{H}(\mathbb{R}^2)$ has a unique fixed point, and that $w^n(Y)$ will tend to this fixed point as $n \to \infty$, for any $Y \in \mathscr{H}(\mathbb{R}^2)$. As a point in $\mathscr{H}(\mathbb{R}^2)$, this fixed point will correspond to a non-empty, closed and bounded subset of $\mathbb{R}^2$ which is the Sierpinski gasket: